Optimal Rates for Community Estimation in the Weighted Stochastic Block Model

Min Xu[†]
mx76@stat.rutgers.edu

Varun Jog[‡]
vjog@wisc.edu

Po-Ling Loh[‡*]
loh@ece.wisc.edu

Department of Statistics[†]
Rutgers University
Piscataway, NJ 08854

Departments of ECE[‡] & Statistics[*]
University of Wisconsin - Madison
Madison, WI 53706

August 2018

Abstract

Community identification in a network is an important problem in fields such as social science, neuroscience, and genetics. Over the past decade, stochastic block models (SBMs) have emerged as a popular statistical framework for this problem. However, SBMs have an important limitation in that they are suited only for networks with unweighted edges; in various scientific applications, disregarding the edge weights may result in a loss of valuable information. We study a weighted generalization of the SBM, in which observations are collected in the form of a weighted adjacency matrix and the weight of each edge is generated independently from an unknown probability density determined by the community membership of its endpoints. We characterize the optimal rate of misclustering error of the weighted SBM in terms of the Renyi divergence of order 1/2 between the weight distributions of within-community and between-community edges, substantially generalizing existing results for unweighted SBMs. Furthermore, we present a computationally tractable algorithm based on discretization that achieves the optimal error rate. Our method is adaptive in the sense that the algorithm, without assuming knowledge of the weight densities, performs as well as the best algorithm that knows the weight densities.

1 Introduction

The recent explosion of network datasets has created a need for new statistical methodology [35, 14, 26, 18]. One active area of research with diverse scientific applications pertains to community detection and estimation, where observations take the form of edges between nodes in a graph, and the goal is to partition the nodes into disjoint groups based on their relative connectivity [15, 23, 38, 41, 31, 37].

A standard model assumption in community recovery problems is that—conditioned on the community labels of the nodes of the graph—each edge is generated independently according to a distribution governed solely by the community labels of its endpoints. This is the setting of the stochastic block model (SBM) [25]. Community recovery may also be viewed as estimating the latent cluster memberships of the nodes a random graph generated by an SBM. The last decade has seen great progress on this problem, beginning with the seminal conjecture of Decelle et al. [13] (see, e.g., the excellent survey paper by Abbe [1]). Various algorithms for community recovery have been devised with guaranteed optimality properties, measured in terms of correlated recovery [32, 34, 30], exact recovery [3, 5, 4], and minimum misclustering error rate [17, 43].

However, an important shortcoming of SBMs is that all edges are assumed to be binary. In contrast, the edges appearing in many real-world networks possess weights reflecting a diversity of strengths or characteristics [36, 11]: Edges in social or cellular networks may quantify the frequency of interactions between pairs of individuals [40, 10]. Similarly, edges in gene co-expression networks are assigned weights corresponding to the correlation between expression levels of pairs of genes [44];

and in brain networks, edge weights may indicate the level of neuronal activity between corresponding regions in the brain [39]. Although an unweighted adjacency matrix could be constructed by disregarding the edge weight data, this might result in a loss of valuable information that could be used to recover hidden communities.

This motivates the *weighted* stochastic block model, which we study in this paper. Each edge is generated from a Bernoulli(p) or Bernoulli(q) distribution, depending on whether its endpoints lie in the same community, and then each edge is assigned an edge weight generated from one of two arbitrary densities, $p(\cdot)$ or $q(\cdot)$. We study the problem of community estimation based on observations of the edge weights in the network, *without* assuming knowledge of p, q, $p(\cdot)$ or $q(\cdot)$. Since $p(\cdot)$ and $q(\cdot)$ are allowed to be continuous, our model strictly generalizes the discrete labeled SBMs considered in previous literature [24, 29, 27], as well as the censored SBM [2, 19, 20].

We emphasize key differences between the weighted SBM framework and the setting of other clustering problems involving continuous edge weights [8, 21]. First, we do not assume that between-cluster edges tend to have heavier weights than within-cluster edges (e.g., in mean-separation models). Such an assumption is critical to many algorithms for weighted networks, since it allows existing algorithms for unweighted SBMs, such as spectral clustering, to be applied in relatively straightforward ways. In contrast, the algorithms in this paper allow us to exploit other potential differences in $p(\cdot)$ and $q(\cdot)$, such as differences in variance or shape. This is crucial to achieve optimal performance. Second, our setting is *nonparametric* in the sense that the densities $p(\cdot)$ and $q(\cdot)$ may be arbitrary and are only required to satisfy mild regularity conditions, whereas previous approaches generally assume that $p(\cdot)$ and $q(\cdot)$ belong to a specific parametric family. Nonparametric density estimation is itself a difficult problem, made even more difficult in the case of weighted SBMs, since we do not know a priori which edge weights have been drawn from which densities.

Our main theoretical contribution is to characterize the *optimal rate of misclustering error* in the weighted SBM. On one side, we derive an information-theoretic lower bound for the performance of any community recovery algorithm for the weighted SBM. Our lower bound applies to all parameters in the parameter space (thus is not minimax) and all algorithms that produce the same output on isomorphic networks—a property that we call *permutation equivariance*. On the other side, we present a computationally tractable algorithm with a rate of convergence that matches the lower bound. Our results show that the optimal rate for community estimation in a weighted SBM is governed by the Renyi divergence of order $\frac{1}{2}$ between two mixed distributions, capturing the discrepancy between the edge probabilities and edge weight densities for between-community and within-community connections. This provides a natural but highly nontrivial generalization of the results in Zhang and Zhou [43] and Gao et al. [17], which show that the optimal rate of the unweighted SBM is characterized by the Renyi divergence of order $\frac{1}{2}$ between two Bernoulli distributions corresponding only to edge probabilities.

Remarkably, our rate-optimal algorithm is fully *adaptive* and does not require prior knowledge of $p(\cdot)$ and $q(\cdot)$. Thus, even in cases where the densities belong to a parametric family, it is possible— *without making any parametric assumptions*—to obtain the same optimal rate as if one imposes the true parametric form. This is in sharp contrast to most nonparametric estimation problems in statistics, where nonparametric methods usually lead to a slower rate of convergence than parametric methods if a specific parametric form is known. The apparent discrepancy is explained by the simply stated observation that in weighted SBMs, one does *not* need to estimate edge densities well in order to recover communities to desirable accuracy. This intuition is also reflected in the work of Abbe and Sandon [4] for the exact recovery problem and Gao et al. [17] for the unweighted SBM. Our proposed recovery algorithm hinges on a careful discretization technique: When the edge weights are bounded, we discretize the distribution via a uniformly spaced binning to convert the weighted SBM into an instance of a *labeled* SBM, where each edge possesses a label from a discrete set with finite (but divergent) cardinality; we then perform community recovery in the labeled SBM by extending a coarse-to-fine clustering algorithm that computes an initialization through spectral clustering [12, 28] and then performs refinement through nodewise likelihood maximization [17].

When the edge weights are unbounded, we reduce the problem to the bounded case by first applying an appropriate transformation to the edge weight distributions.

The remainder of our paper is organized as follows: Section 2 introduces the mathematical framework of the weighted SBM, defines the community recovery problem, and formalizes the notion of permutation equivariance. Section 3 provides an informal summary of our results, later formalized in Section 5. Section 4 outlines our proposed community estimation algorithm. The key technical components of our proofs are highlighted in Section 6, and Section 7 reports the results of various simulations. Section 8 concludes the paper with further implications and open questions.

Notation: For a positive integer n, we write $[n]$ to denote the set $\{1, \ldots, n\}$ and S_n to denote the set of permutations of $[n]$. We write $o(1)$ to denote a sequence indexed by n that tends to 0 as $n \to \infty$, and write $\Theta(1)$ to denote a sequence indexed by n that is bounded away from 0 and ∞ as $n \to \infty$. For two real numbers a and b, we write $a \vee b$ to denote $\max(a, b)$ and write $a \wedge b$ to denote $\min(a, b)$.

2 Model and problem formulation

We begin with a formal definition of the homogeneous weighted SBM and a description of the community recovery problem.

2.1 Weighted stochastic block model

Let n denote the number of nodes in the network and let $K \geq 2$ denote the number of communities. A *clustering* σ is a function $[n] \to [K]$. For each node $u \in [n]$, we refer to $\sigma(u)$ as the cluster of node u.

Definition 2.1. *For a positive number $\beta \geq 1$, we define $\mathcal{C}(\beta, K)$ as the set of clusterings with minimum cluster size is at least $\frac{n}{\beta K}$, i.e., $\sigma \in \mathcal{C}(\beta, K)$ if and only if $|\sigma^{-1}(k)| \geq \frac{n}{\beta K}$ for all $k \in [K]$. We refer to β as the* cluster-imbalance constant.

We first define the homogeneous unweighted SBM, which is characterized by the following probability distribution over adjacency matrices $A \in \{0, 1\}^{n \times n}$:

Definition 2.2 (Homogeneous unweighted SBM). *Let $\sigma_0 \in \mathcal{C}(\beta, K)$ and $p, q \in [0, 1]$. We say that a random binary-valued matrix A has the distribution $SBM(\sigma_0, p, q)$ if for all $u < v$, the entries of A are generated independently according to*

$$A_{uv} \sim \begin{cases} Ber(p) & \text{if } \sigma_0(u) = \sigma_0(v), \\ Ber(q) & \text{if } \sigma_0(u) \neq \sigma_0(v). \end{cases}$$

Thus, the parameters p and q correspond to the within-cluster and between-cluster edge probabilities. The more general *heterogenous* unweighted SBM is characterized by a matrix $P \in \mathbb{R}^{K \times K}$ of probabilities instead of two scalars p and q, and edges are generated independently according to $A_{uv} \sim Ber(P_{\sigma_0(u), \sigma_0(v)})$.

A homogeneous weighted SBM is parametrized by $\sigma_0 \in \mathcal{C}(\beta, K)$, the edge *absence* probabilities P_0 and Q_0, and the edge weight probability densities $p(\cdot)$ and $q(\cdot)$ supported on $S \subset \mathbb{R}$, where S may be $[0, 1]$, $[0, \infty)$, or $\mathbb{R}$. The weighted SBM is then characterized by a distribution over symmetric matrices $A \in S^{n \times n}$ in the following manner:

Definition 2.3 (Homogeneous weighted SBM). *Let $\sigma_0 \in \mathcal{C}(\beta, K)$. We say that a random real-valued matrix A has the distribution $WSBM(\sigma_0, (P_0, p), (Q_0, q))$ if for all $u < v$,*

$$A_{uv} \sim \begin{cases} P_0 \delta_0(\cdot) + (1 - P_0)p(\cdot) & \text{if } \sigma_0(u) = \sigma_0(v), \\ Q_0 \delta_0(\cdot) + (1 - Q_0)q(\cdot) & \text{if } \sigma_0(u) \neq \sigma_0(v). \end{cases} \tag{1}$$

where $P_0\delta_0(\cdot) + (1 - P_0)p(\cdot)$ denotes a probability distribution whose singular part (with respect to the Lebesgue measure) is a point mass at 0 with probability P_0 and whose continuous part has $(1 - P_0)p(\cdot)$ as its Radon-Nikodym derivative with respect to the Lebesgue measure; and $Q_0\delta_0(\cdot) + (1 - Q_0)q(\cdot)$ is defined analogously.

Note that if $p(\cdot)$ and $q(\cdot)$ are Dirac delta masses at 1, the weighted SBM reduces to the unweighted version. We make a few additional remarks about the definition of the weighted SBM. First, we observe that $\mathbb{E}(A)$ may not exhibit the familiar block structure found in unweighted SBMs, since our model includes the case where $(P_0, p(\cdot))$ and $(Q_0, q(\cdot))$ have the same mean. Second, our definition treats an edge with weight 0 as a missing edge, but it is straightforward to distinguish the two notions by defining P and Q as probability measures over $S \cap \{*\}$, where the symbol $*$ denotes a missing edge. Lastly, it is possible to generalize the weighted SBM to a *weighted and labeled* SBM with the model

$$A_{uv} \sim \begin{cases} P, & \text{if } \sigma_0(u) = \sigma_0(v) \\ Q, & \text{if } \sigma_0(u) \neq \sigma_0(v). \end{cases}$$

where P and Q are general probability distributions over S (and the labels correspond to a discrete part). The theory derived in this paper extends in a straightforward fashion to the cases where the discrete portion of P and Q has finite support.

2.2 Community estimation

Given an observation $A \in S^{n \times n}$ generated from a weighted SBM, the goal of community estimation is to recover the true cluster membership structure σ_0. We assume throughout our paper that the number of clusters K is known.

We evaluate the performance of a community recovery algorithm in terms of its misclustering error. For a clustering algorithm $\hat{\sigma}$, let $\hat{\sigma}(A) : [n] \to [K]$ denote the clustering produced by $\hat{\sigma}$ when provided with the input A. We have the following definition:

Definition 2.4. *We define the* misclustering error *to be*

$$l(\hat{\sigma}(A), \sigma_0) := \min_{\pi \in S_K} \frac{1}{n} d_H(\pi \circ \hat{\sigma}(A), \sigma_0),$$

where $d_H(\cdot, \cdot)$ denotes the Hamming distance. The risk *of $\hat{\sigma}$ is defined as $R(\hat{\sigma}, \sigma_0) := \mathbb{E}l(\hat{\sigma}(A), \sigma_0)$, where the expectation is taken with respect to both the random network A and any potential randomness in the algorithm $\hat{\sigma}$.*

The goal of this paper is to characterize the minimal achievable risk for community recovery on the weighted SBM in terms of the parameters $(n, \beta, K, (P_0, p), (Q_0, q))$.

2.3 Permutation equivariance

Since the cluster structure in a network does not depend on how the nodes are labeled, it is natural to focus on estimation algorithms that output equivalent clusterings when provided with isomorphic inputs. We formalize this property in the following definition:

Definition 2.5. *For an $n \times n$ matrix A and a permutation $\pi \in S_n$, let πA denote the $n \times n$ matrix such that $A_{uv} = [\pi A]_{\pi(u), \pi(v)}$. Let $\hat{\sigma}$ be a deterministic clustering algorithm. Then $\hat{\sigma}$ is* permutation equivariant *if, for any A and any $\pi \in S_n$,*

$$\tau \circ \hat{\sigma}(\pi A) \circ \pi = \hat{\sigma}(A) \text{ for some } \tau \in S_K. \tag{2}$$

Note that $\hat{\sigma}(\pi A)$ by itself is not equivalent to $\hat{\sigma}(A)$, since the nodes in πA are labeled with respect to the permutation π. It is straightforward to extend Definition 2.5 to randomized algorithms by requiring condition (2) to hold almost everywhere in the probability space that underlies the algorithmic randomness. Permutation equivariance is a natural property satisfied by all the clustering algorithms studied in literature except algorithms that leverage extra side information in addition to the given network. In Section 5.2, we study permutation equivariance in detail and provide some properties of permutation equivariant estimators.

3   Overview of main results

The difficulty of community recovery depends on the extent to which (P_0, p) and (Q_0, q) are different; it is clearly impossible to have a consistent clustering algorithm if (P_0, p) and (Q_0, q) are equal. We show in this paper that natural measure of discrepancy between (P_0, p) and (Q_0, q) which governs the optimal rate of convergence is the Renyi divergence of order $\frac{1}{2}$.

Given any probability distributions P and Q that are absolutely continuous with respect to each other, the Renyi divergence of order $\frac{1}{2}$ is defined as $I(P, Q) := -2 \log \int \left(\frac{dP}{dQ}\right)^{1/2} dQ$. For our setting, the Renyi divergence takes on the special form

$$I((P_0, p), (Q_0, q)) = -2 \log \left(\sqrt{P_0 Q_0} + \int \sqrt{(1 - P_0)(1 - Q_0)p(x)q(x)} dx \right).$$

If $I((P_0, p), (Q_0, q))$ is bounded above by a universal constant, the Renyi divergence is of the same order as the Hellinger distance (cf. Lemma H.2):

$$I((P_0, p), (Q_0, q)) \asymp (\sqrt{P_0} - \sqrt{Q_0})^2 + \int_S (\sqrt{(1 - P_0)p(x)} - \sqrt{(1 - Q_0)q(x)})^2 dx$$

$$= (\sqrt{P_0} - \sqrt{Q_0})^2 + (\sqrt{1 - P_0} - \sqrt{1 - Q_0})^2$$

$$+ \sqrt{(1 - P_0)(1 - Q_0)} \int_S (\sqrt{p(x)} - \sqrt{q(x)})^2 dx.$$

Thus, we can think of $I((P_0, p), (Q_0, q))$ as having two components, the first of which captures the divergence between the edge presence probabilities (and also appears in the analysis of unweighted SBM), and the second of which captures the divergence between the edge weight densities.

The presence of the second term illustrates how the weighted SBM behaves quite differently from its unweighted counterpart—in particular, *dense* networks may be interesting in a weighted setting. For example, even if the weighted network is completely dense in the sense that $1 - P_0 = 1 - Q_0 = 1$, a nonzero signal I may still exist if $p(\cdot)$ and $q(\cdot)$ are sufficiently different. Our results apply simultaneously to dense and sparse settings; it is important to note that dense weighted networks arise in real-world settings, such as gene co-expression data.

We now provide an informal overview of our main results.

Theorem. (Informal statement) *Let A be generated from a weighted SBM. Under regularity conditions on $((P_0, p), (Q_0, q))$, any permutation equivariant estimator $\hat{\sigma}$ satisfies the lower bound*

$$\mathbb{E}l(\hat{\sigma}(A), \sigma_0) \geq \exp\left(-(1 + o(1))\frac{n}{\beta K} I((P_0, p), (Q_0, q)) \right).$$

Theorem. (Informal statement) *Under regularity conditions on $((P_0, p), (Q_0, q))$, there exists a permutation equivariant algorithm $\hat{\sigma}$ achieving the following misclustering error rate:*

$$\lim_{n \to \infty} P\left(l(\hat{\sigma}(A), \sigma_0) \leq \exp\left(-(1 + o(1))\frac{n}{\beta K} I((P_0, p), (Q_0, q)) \right) \right) \to 1.$$

Furthermore, if $\frac{nI}{\beta K \log n} \leq 1$, we have

$$\mathbb{E}l(\hat{\sigma}(A), \sigma_0) \leq \exp\left(-(1 + o(1))\frac{nI}{\beta K}\right).$$

Taken together, the theorems imply that in the regime where $\frac{nI}{\beta K \log n} \leq 1$, the optimal risk is tightly characterized by the quantity $\exp\left(-(1 + o(1))\frac{nI}{\beta K}\right)$. On the other hand, if $\frac{nI}{\beta K \log n} > 1$, we have $\exp\left(-(1 + o(1))\frac{nI}{\beta K}\right) < \frac{1}{n}$ for large enough n, so $\lim_{n \to \infty} P\left(l(\hat{\sigma}(A), \sigma_0) = 0\right) \to 1$ (since $l(\hat{\sigma}(A), \sigma_0) < \frac{1}{n}$ implies $l(\hat{\sigma}(A), \sigma_0) = 0$). Thus, the regime where $\frac{nI}{\beta K \log n} > 1$ is in some sense an easier problem, since we can guarantee perfect recovery with high probability.

3.1 Relation to previous work

Our result generalizes the work of Zhang and Zhou [43], which establishes the minimax rate of $\exp\left(-(1 + o(1))\frac{n}{\beta K} I\left(Ber(p), Ber(q)\right)\right)$ for the unweighted SBM, where

$$I(Ber(p), Ber(q)) = -2\log\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right).$$

The optimal algorithm proposed in Zhang and Zhou [43] is intractable, but a computationally feasible version was developed by Gao et al. [17]; the latter algorithm is a building block for the estimation algorithm proposed in this paper.

Our result should also be viewed in comparison to Yun and Proutiere [42], who studied the optimal risk for the heterogenous labeled SBM with finitely many labels, with respect to a prior on the cluster assignment σ_0. They characterize the optimal rate under a notion of divergence that reduces to the Renyi divergence of order $\frac{1}{2}$ between two discrete distributions over a fixed finite number of labels in the homogeneous setting (cf. Lemma G.2). Since the discussion is somewhat technical, we provide a more detailed comparison of our work to the results of Yun and Proutiere in Section 6.1.

Jog and Loh [27] proposed a similar weighted block model and show the exact recovery threshold to be dependent on the Renyi divergence. They focus on the setting where the distributions are discrete and known, whereas we consider continuous densities that are unknown. Aicher et al. [6] introduced a version of a weighted SBM that is a special case of the setting discussed in this paper, where the densities P and Q in equation (1) are drawn from a known exponential family. Notably, the definition of Aicher et al. [6] cannot incorporate sparsity. The weighted SBM model considered in Hajek et al. [22] is also similar to the one we propose in our paper, except it only involves a single hidden community and assumes knowledge of the distributions P and Q. Weighted networks have also received some attention in the physics community [36, 9], and various ad-hoc methods have been proposed; since theoretical properties are generally unknown, we do not explore these connections in our paper.

Other notions of recovery: A closely related problem is that of finding the exact recovery threshold. We say that the unweighted SBM has an *exact recovery threshold* if a function $\theta(p, q, n, K, \beta, \sigma_0)$ exists such that exact recovery is asymptotically almost always impossible if $\theta < 1$, and almost always possible if $\theta > 1$. For the homogeneous unweighted SBM, Abbe et al. [3] show that when $\beta = 1, K = 2, 1 - P_0 = \frac{a \log n}{n}$, and $1 - Q_0 = \frac{b \log n}{n}$, for some constants a and b, the exact recovery threshold is $\sqrt{a} - \sqrt{b}$. This result was later generalized to multiple communities with heterogenous edge probabilities in Abbe and Sandon [5], where a notion of CH-divergence was shown to characterize the threshold for exact recovery. A notion of weak recovery, corresponding to a detection threshold, has also been considered [30, 33].

4 Estimation algorithm

A natural approach to community estimation is to first estimate the edge weight densities $p(\cdot)$ and $q(\cdot)$, but this is hindered by the fact that we do not know whether an edge weight observation originates from $p(\cdot)$ or $q(\cdot)$. An alternative approach of applying spectral clustering directly to the weighted adjacency matrix A will also be ineffective if (P_0, p) and (Q_0, q) have the same mean, so $\mathbb{E}(A)$ does not exhibit any cluster structure. A third idea is to output the clustering that maximizes the Kolmogorov-Smirnov distance (or another nonparametric two-sample test statistic) between the empirical CDFs of within-cluster edge weights and the between-cluster edge weights. This idea, though feasible, is computationally intractable, since it involves searching over all possible clusterings. Our approach is appreciably different from the methods suggested above, and consists of combining the idea of discretization from nonparametric density estimation with clustering techniques for unweighted SBMs.

4.1 Outline of algorithm

We begin by describing the main components of our algorithm. The key ideas are to convert the edge weights into a finite set of labels by discretization, and then cluster nodes on the labeled network. Our algorithm is summarized pictorially in Figure 1.

1. **Transformation & discretization.** We take as input a weighted adjacency matrix A and apply an invertible transformation function $\Phi : S \to [0, 1]$ (recall S is the support of the edge weights and can be $[0, 1]$, $[0, \infty)$, or $\mathbb{R}$) on the nonzero edges to obtain a matrix $\Phi(A)$ with weights between 0 and 1. Next, we divide the interval $[0, 1]$ into L equally-spaced subintervals. We replace the real-valued entries of $\Phi(A)$ with categorical labels in $[L]$. We denote the labeled adjacency matrix by A_L.

2. **Add noise.** We perform the following process on every edge of the labeled graph, independently of other edges: With probability $1 - \delta$ where $\delta = \frac{2(L+1)}{n}$, keep an edge as it is, and with probability δ, erase the edge and replace it with an edge with label uniformly drawn from the set of labels. We continue to denote the modified adjacency matrix as A_L.

3. **Initialization parts 1 & 2.** For each label l, we create a sub-network by including only edges of label l. We then perform spectral clustering on all sub-networks, and output the label l^* that induces the maximally separated spectral clustering. Let A_{l^*} be the adjacency matrix for label l^*. For each $u \in \{1, \ldots, n\}$, we perform spectral clustering on $A_{l^*} \setminus \{u\}$, which denotes the adjacency matrix with vertex u removed. We output n clusterings $\tilde{\sigma}_1, \ldots, \tilde{\sigma}_n$.

4. **Refinement & consensus.** From each $\tilde{\sigma}_u$, we generate a clustering $\hat{\sigma}_u$ on $\{1, 2, \ldots, n\}$ that retains the assignments specified by $\tilde{\sigma}_u$ for $\{1, 2, \ldots, n\} \setminus \{u\}$, and assigns $\hat{\sigma}_u(u)$ by maximizing the likelihood taking into account only the neighborhood of u. We then align the cluster assignments made in the previous step.
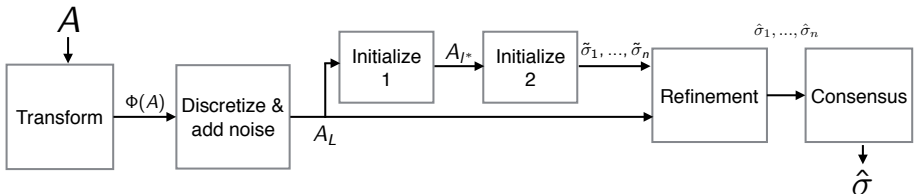

Figure 1: Pipeline for our proposed algorithm

4.2 Transformation and discretization

In the transformation step, we apply an invertible CDF $\Phi : S \to [0,1]$ as the transformation function on all the edge weights, so that each entry of $\Phi(A)$ lies in $[0,1]$. In the discretization step, we divide the interval $[0,1]$ into L equally-spaced bins of the form $[a_l, b_l]$, where $a_1 = 0, b_L = 1$, and $b_l - a_l = \frac{1}{L}$. An edge is assigned the label l if the weight of that edge lies in bin l.

Algorithm 4.1 Transformation and Discretization

Input: A weighted network A, a positive integer L, and an invertible function $\Phi : S \to [0,1]$
Output: A labeled network A_L with L labels

> Divide $[0,1]$ into L bins, labeled $\mathrm{bin}_1, \ldots, \mathrm{bin}_L$
> **for** every edge (u,v) such that $A_{uv} \neq 0$ **do**
> > Let l be the bin in which $\Phi(A_{uv})$ falls
> > Assign the edge (u,v) the label l in the labeled network A_L
> **end for**

4.3 Add noise

For technical reasons, we inject noise into the network as a form of regularization. As detailed in the proof of Proposition 6.1 in Appendix A, deliberately forming a noisy version of the graph barely affects the separation between the distributions of the within-community and between-community edge labels, but has the desirable effect of ensuring that all edge labels occur with probability at least $\frac{2}{n}$. This property is crucial to our analysis in subsequent steps of the algorithm. In the description of the algorithm below, we treat the label 0 (i.e., an empty edge) as a separate label, so we have a network with $L + 1$ labels.

Algorithm 4.2 Add noise

Input: A labeled network A_L with $L + 1$ labels
Output: A labeled network A_L with $L + 1$ labels

> **for** every edge (u,v) **do**
> > With probability $1 - \frac{2(L+1)}{n}$, do nothing
> > With probability $\frac{2(L+1)}{n}$, replace the edge label with a label drawn uniformly at random from $\{0, 1, 2, \ldots, L\}$
> **end for**

4.4 Initialization

The initialization procedure takes as input a network with edges labeled $\{0, 1, \ldots, L\}$. The goal of the initialization procedure is to create a rough clustering $\tilde{\sigma}$ that is consistent but not necessarily optimal. As outlined in Algorithm 4.3, the rough clustering is based on a single label l^*, selected based on the maximum value of the estimated Renyi divergence between within-community and between-community distributions for the unweighted SBMs based on individual labels.

For technical reasons, we actually create n separate rough clusterings $\{\tilde{\sigma}_u\}_{u=1,\ldots,n}$, where each $\tilde{\sigma}_u : [n-1] \to [K]$ is a clustering of a network of $n-1$ nodes with u removed. The clusterings $\{\tilde{\sigma}_u\}$ will later be combined into a single clustering algorithm. In practice, it is sufficient to create a single rough clustering (see Remark 4.2 below).

Remark 4.1. *The initialization procedure that we propose is based on choosing a single best label l^* and deriving an initial clustering from the unweighted network associated with l^*. This is sufficient in theory, but a better initial clustering may be gained in practice by aggregating information from all labels. Such an aggregation must, however, be performed with care, so that uninformative labels do not dilute the information content of the informative labels.*

Algorithm 4.3 Initialization

Input: A labeled network A_L with L labels
Output: A set of clusterings $\{\tilde{\sigma}_u\}_{u=1,\ldots,n}$, where $\tilde{\sigma}_u$ is a clustering on $\{1, 2, \ldots, n\} \setminus \{u\}$

1: Separate A_L into L networks $\{A_l\}_{l=1,\ldots,L}$ ▷ Stage 1
2: **for** each label l **do**
3:      Perform SPECTRAL CLUSTERING (Algorithm 4.4) with $\tau = 40K\bar{d}$, where $\bar{d} = \frac{1}{n}\sum_{u=1}^{n} d_u$ is the average degree, and $\mu = 4\beta$ to obtain $\tilde{\sigma}_l$
4:      Estimate $\hat{P}_l = \frac{\sum_{u \neq v\,:\,\tilde{\sigma}_l(u)=\tilde{\sigma}_l(v)}(A_l)_{uv}}{|\{u \neq v\,:\,\tilde{\sigma}_l(u)=\tilde{\sigma}_l(v)\}|}$ and $\hat{Q}_l = \frac{\sum_{u \neq v\,:\,\tilde{\sigma}_l(u)\neq\tilde{\sigma}_l(v)}(A_l)_{uv}}{|\{u \neq v\,:\,\tilde{\sigma}_l(u)\neq\tilde{\sigma}_l(v)\}|}$
5:      Compute $\hat{I}_l \leftarrow \frac{(\hat{P}_l - \hat{Q}_l)^2}{\hat{P}_l \vee \hat{Q}_l}$
6: **end for**
7: Choose $l^* = \arg\max_l \hat{I}_l$
8: **for** each node u **do**   ▷ Stage 2
9:      Create network $A_{l^*} \setminus \{u\}$ by removing node u from A_{l^*}
10:      Perform SPECTRAL CLUSTERING (with the same parameter setting as stage 1) on $A_{l^*} \setminus \{u\}$ to obtain $\tilde{\sigma}_u$
11: **end for**
12: Output the set of clusterings $\{\tilde{\sigma}_u\}_{u=1,\ldots,n}$

Spectral clustering: Note that Algorithm 4.3 involves several applications of SPECTRAL CLUSTERING. We describe the spectral clustering algorithm used as a subroutine in Algorithm 4.4 below. Importantly, note that we may always choose the parameter μ sufficiently large such that Algorithm 4.4 generates a set S with $|S| = K$.

4.5 Refinement and consensus

This step parallels Gao et al. [17]. In the refinement step, we use the set of initial clusterings $\{\tilde{\sigma}_u\}_{u=1,\ldots,n}$ to generate a more accurate clustering for the labeled network by locally maximizing an approximate log-likelihood for each node u. The consensus step resolves any cluster label inconsistencies present after the refinement stage.

Remark 4.2. *In our simulation studies, we find that it is sufficient to output a single clustering $\tilde{\sigma}$ on the whole of A_{l^*} in the initialization stage. In the refinement stage, we simply estimate $\{\hat{P}_l, \hat{Q}_l\}_{l\in\{0,\ldots,L\}}$ based on $\tilde{\sigma}$, assign $\hat{\sigma}(u) = \arg\max_{k\in[K]}\sum_{v\,:\,\tilde{\sigma}(v)=k,\,v\neq u}\sum_{l=0}^{L}\log\frac{\hat{P}_l}{\hat{Q}_l}\mathbf{1}(A_{uv} = l)$, and then output $\hat{\sigma}$ directly. We also note that one could in practice use a discretization level for the refinement stage that is different from that of the initialization stage (see discussions in Section 6).*

5 Optimal misclustering error

We analyze the rate of convergence of the estimation algorithm from Section 4 in Section 5.1. In Section 5.2, we provide a matching information-theoretic lower bound. In both sections, we let $\mathcal{P}$ denote the set of probability distributions on S whose singular part is a point mass at 0.

Algorithm 4.4 SPECTRAL CLUSTERING

Input: An unweighted network A with columns $\{A_u\}$, trim threshold τ, number of communities K, and tuning parameter μ
Output: A clustering σ

1: For each node u with degree $d_u \geq \tau$, set $A_u = 0$ and $(A^\top)_u = 0$ to obtain $T_\tau(A)$
2: Compute $\hat{A} := \arg\min_{\tilde{A}\,:\,\mathrm{rank}(\tilde{A})\leq K} \|\tilde{A} - T_\tau(A)\|_2$ by SVD
3: For each node u, index the other nodes by $v_{(1)}, \ldots, v_{(n-1)}$ such that

$$\|\hat{A}_u - \hat{A}_{v_{(1)}}\|_2 \leq \|\hat{A}_u - \hat{A}_{v_{(2)}}\|_2 \leq \ldots \leq \|\hat{A}_u - \hat{A}_{v_{(n-1)}}\|_2,$$

and define

$$D(u) := \|\hat{A}_u - \hat{A}_{v_{(\lceil n/\mu K\rceil)}}\|_2$$

4: Initialize $S \leftarrow 0$
5: Select node $u_1 := \arg\min_u D(u)$ and add u_1 to S as $S[1]$
6: **for** $i = 2, \ldots, K$ **do**
7:     Among all u such that $|D(u)| \leq (1 - \frac{1}{\mu K})$-quantile$\{D(v) : v \in [n]\}$, select

$$u_i = \arg\max_u \min_{v\in\{S[1],\ldots,S[i-1]\}} \|\hat{A}_u - \hat{A}_v\|_2$$

8:     Add u_i to S as $S[i]$
9: **end for**
10: **for** $u = 1, \ldots, n$ **do**
11:     Assign $\sigma(u) = \arg\min_i \|\hat{A}_u - \hat{A}_{S[i]}\|$
12: **end for**

Algorithm 4.5 Refinement

Input: A labeled network A_L and a set of clusterings $\{\tilde{\sigma}_u\}_{u=1,\ldots,n}$, where $\tilde{\sigma}_u$ is a clustering on the set $\{1, 2, \ldots, n\} \setminus \{u\}$ for each u
Output: A clustering $\hat{\sigma}$ over the whole network

1: **for** each node u **do**
2:     Estimate $\{\hat{P}_l, \hat{Q}_l\}_{l=0,\ldots,L}$ from $\tilde{\sigma}_u$
3:     Let $\hat{\sigma}_u : [n] \to [K]$, where $\hat{\sigma}_u(v) = \tilde{\sigma}_u(v)$ for all $v \neq u$ and

$$\hat{\sigma}_u(u) = \arg\max_{k\in[K]} \sum_{v\,:\,\tilde{\sigma}_u(v)=k,\,v\neq u} \sum_{l=0}^{L} \log\frac{\hat{P}_l}{\hat{Q}_l}\mathbf{1}(A_{uv} = l)$$

4: **end for**
5: Let $\hat{\sigma}(1) = \hat{\sigma}_1(1)$ ▷ Consensus Stage
6: **for** each node $u \neq 1$ **do**

$$\hat{\sigma}(u) = \arg\max_{k\in[K]} |\{v : \hat{\sigma}_1(v) = k\} \cap \{v : \hat{\sigma}_u(v) = \hat{\sigma}_u(u)\}|$$

7: **end for**
8: Output $\hat{\sigma}$

5.1 Upper bound

We begin by stating a condition on the function Φ.

Definition 5.1. *Let S be $[0,1]$, $\mathbb{R}$, or $\mathbb{R}^+$. We say that $\Phi : S \to [0,1]$ is a transformation function if it is a differentiable bijection and $\phi := \Phi'$ satisfies $\left|\frac{\phi'(x)}{\phi(x)}\right| < \infty$.*

For $S = [0,1]$, we always take Φ to be the identity. For $S = \mathbb{R}$ or $[0,\infty)$, we choose the function Φ so that all moments exist and ϕ has a subexponential tail. The specific choice of Φ is not crucial, and we will use the following definitions:

$$\phi(x) = \frac{e^{1-\sqrt{x+1}}}{4}, \text{ if } S = [0,\infty), \qquad \phi(x) = \frac{e^{1-\sqrt{|x|+1}}}{8}, \text{ if } S = \mathbb{R}. \tag{3}$$

These expressions are similar to a generalized normal density, modified so that $\left|\frac{\phi'(x)}{\phi(x)}\right|$ is bounded. It is easy to verify that $\Phi(x) = \int_0^x \phi(t)dt$ (respectively, $\Phi(x) = \int_{-\infty}^x \phi(t)dt$) is a valid transformation function. The function Φ induces a probability measure on S, and we let $\Phi\{\cdot\}$ denote the Φ-measure of a set.

We describe our regularity conditions by defining an appropriate subset of $\mathcal{P}^2$. For $C \in [1,\infty)$, $c_1, c_2 \in \text{int}(S)$, $r > 2$, and $t \in (2/r, 1)$, we define $\mathcal{G}_{\Phi,C,c_1,c_2,r,t} \subset \mathcal{P}^2$ such that $((P_0,p),(Q_0,q)) \in \mathcal{G}_{\Phi,C,c_1,c_2,r,t}$ if and only if

A0 We have $\frac{1}{C} \leq \frac{1-P_0}{1-Q_0} \leq C$ and $\frac{1}{C} \leq \frac{P_0}{Q_0} \leq C$.

A1 For all x in the interior of S, $0 < p(x), q(x) \leq C\phi(x)$.

A2 There exists a quasi-convex $g : S \to [0,\infty)$ such that $g(x) \geq \left|\log\frac{p(x)}{q(x)}\right|$ and $\int_S g(x)^r \phi(x)\,dx \leq C$.

A3 Denoting $\alpha^2 := \int_S (\sqrt{p(x)} - \sqrt{q(x)})^2\,dx$ and $\gamma(x) := \frac{p(x)-q(x)}{\alpha}$, we have

$$\int_S \left(\frac{\gamma(x)}{p(x)+q(x)}\right)^r (p(x)+q(x))\,dx \leq C.$$

A4 There exists a quasi-convex function $h : S \to [0,\infty)$ such that

$$h(x) \geq \frac{1}{\phi(x)} \max\left\{\left|\frac{\gamma(x)}{p(x)+q(x)}\right|, \left|\frac{\gamma'(x)}{p(x)+q(x)}\right|, \left|\frac{q'(x)}{q(x)}\right|, \left|\frac{p'(x)}{p(x)}\right|\right\}$$

and $\int_S |h(x)|^t \phi(x)dx \leq C$.

A5 We have $(\log p)'(x), (\log q)'(x) \geq (\log \phi)'(x)$ for all $x < c_1$, and $(\log p)'(x), (\log q)'(x) \leq (\log \phi)'(x)$ for all $x > c_2$.[1]

The above conditions depend on the choice of Φ, but it generally suffices to choose Φ such that its derivative ϕ is a heavy-tailed density where all moments exist. In particular, we show in Section 5.1.3 that choosing Φ according to equation (3) allows $\mathcal{G}_\Phi$ to encompass Gaussian, Laplace, and other broad classes of densities. We also provide an intuitive discussion of the regularity conditions in Section 5.1.1 below.

We now state our upper bound. For a given clustering σ_0 and $((P_0,p),(Q_0,q)) \in \mathcal{P}^2$, let the random network A be distributed according to $WSBM(\sigma_0,(P_0,p),(Q_0,q))$.

---

[1]If $S = [0,\infty)$ and g is non-decreasing, we only need $(\log p)'(x), (\log q)'(x) \leq (\log \phi)'(x)$ for all $x > c_2$.

Theorem 5.1. *Let $\sigma_0 \in \mathcal{C}(\beta, K)$. Let $C \geq 1$, $c_1, c_2 \in \text{int}(S)$, $r > 2$, and $t \in (2/r, 1)$, and let Φ be a transformation function. Define $\mathcal{G}_\Phi := \mathcal{G}_{\Phi, C, c_1, c_2, r, t}$. Let $\{I_n, I'_n\}_{n \in \mathbb{N}}$ be arbitrary sequences such that $I_n \to 0$ and $nI'_n \to \infty$. Let L_n be a sequence such that $\frac{nI'_n}{L_n \exp(L_n^{2/r})} \to \infty$. Let $\hat{\sigma}_{\Phi, L_n}$ be the algorithm described in Section 4 with transformation function Φ and discretization level L_n. Then there exists a sequence of real numbers $\zeta_n \to 0$ such that*

$$\lim_{n \to \infty} \sup_{\substack{((P_0, p), (Q_0, q)) \in \mathcal{G}_\Phi : \\ I'_n \leq I((P_0, p), (Q_0, q)) \leq I_n}} \mathbb{P}_{\substack{(P_0, p), \\ (Q_0, q)}} \left\{ l(\hat{\sigma}_{\Phi, L_n}(A), \sigma_0) \leq \exp\left(-(1 - \zeta_n) \frac{n}{\beta K} I((P_0, p), (Q_0, q)) \right) \right\} = 1.$$

Furthermore, if $\frac{nI_n}{\beta K \log n} \leq 1$, we have

$$\sup_{\substack{((P_0, p), (Q_0, q)) \in \mathcal{G}_\Phi : \\ I'_n \leq I((P_0, p), (Q_0, q)) \leq I_n}} \mathbb{E}_{(P_0, p), (Q_0, q)} \left[l(\hat{\sigma}_{\Phi, L_n}(A), \sigma_0) \right] \exp\left((1 - \zeta_n) \frac{n}{\beta K} I((P_0, p), (Q_0, q)) \right) \leq 1.$$

We relegate the full proof of Theorem 5.1 to Appendix D.1 but we provide a proof overview in Section 6. Since Theorem 5.1 involves many technical details, we first make a few high-level remarks to illustrate its implications.

Remark 5.1. *It is important to observe that the supremum over $\mathcal{G}_\Phi$ appears after the limit. Thus, an equivalent way to understand the theorem is to think of a sequence $((P_{0,n}, p_n), (Q_{0,n}, q_n))$, each term of which is a member of $\mathcal{G}_\Phi$. If $I((P_{0,n}, p_n), (Q_{0,n}, q_n))$ is $o(1)$ but $\omega(L_n \exp(L_n^{2/r})n^{-1})$, Theorem 5.1 states that $\mathbb{P}\left\{ l(\hat{\sigma}(A), \sigma_0) \leq \exp\left(-(1 + o(1)) \frac{n}{\beta K} I((P_{0,n}, p_n), (Q_{0,n}, q_n)) \right) \right\} \to 1$. Theorem 5.1 thus applies to the so-called sparse setting where $P_0, Q_0 \to 1$. In particular, suppose there are constants $a, b > 0$ such that $P_{0,n} = 1 - \frac{a \log n}{n}$ and $Q_0 = 1 - \frac{b \log n}{n}$. Then Theorem 5.1 states that perfect recovery is achievable if $(\sqrt{a} - \sqrt{b})^2 + \sqrt{ab} \int_S (\sqrt{p_n(x)} - \sqrt{q_n(x)})^2 \, dx > \beta K$; this generalizes the previously known result that perfect recovery for unweighted SBMs when $p = 1 - \frac{a \log n}{n}$ and $q = 1 - \frac{b \log n}{n}$ is possible if $(\sqrt{a} - \sqrt{b})^2 > \beta K$.*

Remark 5.2. *The assumption that there exist sequences $I_n \to 0$ and $I'_n = \omega(1/n)$ such that $I'_n \leq I((P_0, p), (Q_0, q)) \leq I_n$ is a very mild one. As our information-theoretic lower bound (cf. Section 5.2) shows, estimation consistency is impossible if a sequence $I'_n = \omega(1/n)$ such that $I((P_0, p), (Q_0, q)) \geq I'_n$ does not exist. Moreover, we observe that if $I((P_0, p), (Q_0, q)) > \beta K \frac{\log n}{n}$, then $\mathbb{P}(l(\hat{\sigma}(A), \sigma_0) = 0) \to 1$, and we are able to perfectly recover the clustering with high probability. Since the estimation problem is intrinsically easier if as $I((P_0, p), (Q_0, q))$ becomes larger, we expect the same perfect recovery guarantee to hold in the case where $I((P_0, p), (Q_0, q))$ is positively bounded away from 0.*

Remark 5.3. *Since $nI'_n \to \infty$, it is always possible to choose a sequence $L_n \to \infty$ satisfying the conditions of the theorem. Note that L_n must grow very slowly to satisfy the condition that $\frac{nI'_n}{L_n \exp(L_n^{2/r})} \to \infty$; indeed, our simulation studies (cf. Section 7) confirm that we should choose the discretization level to be very small in order to achieve good performance. We note that L_n has a second-order effect on the rate and appears in the ζ_n term.*

5.1.1 Additional discussion of the conditions

It is crucial to note that our algorithm does *not* require prior knowledge of the form of $p(\cdot)$ and $q(\cdot)$; the same algorithm and guarantees apply so long as $((P_0, p), (Q_0, q)) \in \mathcal{G}_{\Phi, C, c_1, c_2, r, t}$ for some universal constants C, c_1, c_2, r, and t. To aid the reader, we now provide a brief, non-technical interpretation of the regularity conditions described above.

Condition A1 is simple; the last part states that ϕ must have a tail at least as heavy as that of $p(\cdot)$ and $q(\cdot)$. Condition A2 requires that the likelihood ratio be integrable. It is analogous to

12

a bounded likelihood ratio condition, but much weaker; we add a mild quasi-convexity constraint for technical reasons related to the analysis of binning. In condition A3, the function $\gamma(\cdot)$ is of constant order in the sense that $\int_S \left(\frac{\gamma(x)}{p(x)+q(x)}\right)^r (p(x)+q(x))dx \leq C$. Requirements on $\gamma(\cdot)$ translate into convergence statements on $|p-q|$: For instance, an L_∞-bound on γ implies almost uniform convergence (with respect to Φ) of $|p-q|$ to 0. The integrability condition we impose on $\gamma(\cdot)$ in condition A3 is analogous to an L_∞-bound, but much weaker.

Condition A4 controls the smoothness of the derivatives of $\log p(\cdot)$ and $\log q(\cdot)$. Condition A5 is a mild shape constraint on $p(\cdot)$ and $q(\cdot)$. When $S = \mathbb{R}$, this condition essentially requires $p(\cdot)$ and $q(\cdot)$ to be monotonically increasing in x for $x \to -\infty$, and decreasing in x for $x \to \infty$.

5.1.2 Examples for $S = [0, 1]$

When $S = [0, 1]$, we can always take Φ to be the identity—we do not need a transformation, but we keep the same notation in order to present our results in a unified manner. The simplest example of $\mathcal{G}_\Phi$ that satisfy conditions A1–A5 is when, for all $((P_0, p), (Q_0, q)) \in \mathcal{G}_\Phi$, the densities $p(\cdot)$ and $q(\cdot)$ are bounded above and below by strictly positive universal constants, and when the function $x \mapsto \frac{p(x)-q(x)}{\alpha}$ and its derivative are bounded by universal constants.

5.1.3 Examples for $S = \mathbb{R}$ or $[0, \infty)$

We begin with a proposition that characterizes conditions A1–A5 in the setting where $p(\cdot) = e^{f_{\theta_1}(\cdot)}$ and $q(\cdot) = e^{f_{\theta_0}(\cdot)}$, for some parametrized family $\{f_\theta\}_{\theta \in \Theta}$. This result allows us to generate several large classes of examples.

Proposition 5.1. *Let $C^{**} \in [1, \infty)$, $c_1, c_2 \in S$, $r > 2$, and $t \in (2/r, 1/2)$. Let $\Theta \subset \mathbb{R}^d$ be compact and suppose $diam(\Theta) < 1 \wedge \frac{1}{2C^{**2}}$. Let $\{f_\theta\}_{\theta \in \Theta}$ be a collection of functions such that $e^{f_\theta(\cdot)}$ is a density and:*

B1 *For all $\theta \in \Theta$ and all $x \in S$, we have $0 < e^{f_\theta(x)} \leq C^* \phi(x)$.*

B2 *We have $\inf_{\theta \in \Theta} \lambda_{\min} \left(\int_S 2\nabla f_\theta(x)(\nabla f_\theta(x))^\top \phi(x)\, dx\right) \geq C^{*-1}$ and $\sup_{\theta \in \Theta} \int_S \lambda_{\max}\left(H(f_\theta)(x)\right)^2 \phi(x)\, dx \leq C^*$.*

B3 *There exists a quasi-convex function $g^* : S \to [0, \infty)$ such that $g^*(x) \geq \sup_\theta \|\nabla f_\theta(x)\|_2$ and $\int_S g^*(x)^r \phi(x)\, dx \leq C^*$.*

B4 *There exists a quasi-convex function $h^* : S \to [0, \infty)$ such that*

$$h^*(x) \geq \frac{1}{\phi(x)} \max\left\{\sup_{\theta \in \Theta} \|\nabla f_\theta(x)\|_2,\ \sup_{\theta \in \Theta} \|\nabla f_\theta'(x)\|_2,\ \sup_{\theta \in \Theta} |f_\theta'(x)|\right\}$$

and $\int_S h^*(x)^{2t} \phi(x)\, dx \leq C^*$.

B5 *For all $x \leq c_1$, we have $\inf_{\theta \in \Theta} f_\theta'(x) \geq (\log \phi)'(x)$, and for all $x \geq c_2$, we have $\sup_{\theta \in \Theta} f_\theta'(x) \leq (\log \phi)'(x)$.[2]*

Then there exists $C \in [1, \infty)$ such that for any $\theta_1, \theta_2 \in \Theta$ and any $P_0, Q_0 \in [0, 1]$ such that $\frac{1}{C} \leq \frac{P_0}{Q_0}, \frac{1-P_0}{1-Q_0} \leq C$, we have $((P_0, e^{f_{\theta_1}}), (Q_0, e^{f_{\theta_2}})) \in \mathcal{G}_{\Phi, C, c_1, c_2, r, t}$.

In all the examples below, we take Φ to be the transformation function defined in equation (3). The proofs of all statements in the examples are provided in Section E.2

Example 5.1 (Location-scale family over $\mathbb{R}$). *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function such that $\int_{-\infty}^{\infty} e^{f(x)}\, dx = 1$. Suppose*

---

[2]If $S = [0, \infty)$ and g^* is non-decreasing, we only need $\sup_{\theta \in \Theta} f_\theta'(x) \leq (\log \phi)'(x)$ for all $x \geq c_2$.

(a) $|f^{(k)}(x)|$ is bounded for some $k \geq 2$, and

(b) there exist $c, M > 0$ such that $f'(x) > M$ for $x < -c$ and $f'(x) < -M$ for $x > c$.

For any $\mu \in \mathbb{R}$ and $\sigma > 0$, define $f_{\mu,\sigma}(x) := f\left(\frac{x-\mu}{\sigma}\right) - \log \sigma$.

Then there exists $C_\mu > 0$ and $c_\sigma > 1$ such that, with $\Theta := [-C_\mu, C_\mu] \times [\frac{1}{c_\sigma}, c_\sigma]$, the family $\{f_{\mu,\sigma}\}_{(\mu,\sigma) \in \Theta}$ satisfies conditions B1–B5 in Proposition 5.1 with respect to ϕ defined in equation (3), and some universal constants C^{**}, c_1, c_2, r, and t. As a direct consequence of Proposition 5.1, for some universal constant $C > 0$, if we fix any $((\mu_1, \sigma_1), (\mu_0, \sigma_0)) \in \Theta^2$ and define

$$p(x) = \frac{1}{\sigma_1} \exp\left(f\left(\frac{x - \mu_1}{\sigma_1}\right)\right), \quad \text{and} \quad q(x) = \frac{1}{\sigma_0} \exp\left(f\left(\frac{x - \mu_0}{\sigma_0}\right)\right), \tag{4}$$

then $((P_0, p), (Q_0, q)) \in \mathcal{G}_{\Phi, C, c_1, c_2, r, t}$ for any $P_0, Q_0 \in [0, 1]$ that satisfy condition A0.

These assumptions on f are satisfied for **Gaussian location-scale families**, where the base density is the standard Gaussian density with $f(x) = -x^2 - \frac{1}{2}\log 2\pi$, and **Laplace location-scale families**, where the base density is the standard Laplace density with $f(x) = -|x| - \log 2$.

Example 5.2 (Scale family over $[0, \infty)$). *Let* $f : [0, \infty) \to \mathbb{R}$ *be a continuously differentiable function such that* $\int_0^\infty e^{f(x)}\, dx = 1$. *Suppose*

(a) $|f^{(k)}(x)|$ is bounded for some $k \geq 2$, and

(b) there exist $c, M > 0$ such that $f'(x) < -M$ for $x > c$.

For any $\sigma > 0$, define $f_\sigma(x) := f\left(\frac{x}{\sigma}\right) - \log \sigma$.

Then there exists $c_\sigma > 1$ such that, with $\Theta := [\frac{1}{c_\sigma}, c_\sigma]$, the family $\{f_\sigma\}_{\sigma \in \Theta}$ satisfies conditions B1–B5 in Proposition 5.1 with respect to ϕ defined in equation (3), and some universal constants C^{**}, c_1, c_2, r, and t. As a direct consequence of Proposition 5.1, for some universal constant $C > 0$, if we fix any $(\sigma_1, \sigma_0) \in \Theta^2$ and define

$$p(x) = \frac{1}{\sigma_1} \exp\left(f\left(\frac{x}{\sigma_1}\right)\right), \quad \text{and} \quad q(x) = \frac{1}{\sigma_0} \exp\left(f\left(\frac{x}{\sigma_0}\right)\right),$$

then $((P_0, p), (Q_0, q)) \in \mathcal{G}_{\Phi, C, c_1, c_2, r, t}$ for any $P_0, Q_0 \in [0, 1]$ that satisfy condition A0.

These assumptions on f are satisfied for **exponential scale families**, where the base density is the standard exponential density with $f(x) = -x$.

Proposition 5.1 also applies to the family of Gamma distributions, see Proposition E.3 in the appendix.

In this paper, we only study continuous edge weights in detail; in practice, discrete edge weights such as counts are also important. Although Theorem 5.1 does not apply directly to such cases, our analysis is relevant to some instances of SBMs with discrete edge weights. In Appendix F, we discuss a crude way to handle count-valued edge weights, with particular attention toward Poisson-distributed edge weights.

5.2 Lower bound

Our information-theoretic lower bound applies to any permutation equivariant estimators (Definition 2.5). Before stating the result, we define an appropriate subset of $\mathcal{P}^2$ to capture the conditions we need on $((P_0, p), (Q_0, q))$. Let $C^* \in [1, \infty)$, and let $\mathcal{G}^*_{C^*} \subset \mathcal{P}^2$ be such that $((P_0, p), (Q_0, q)) \in \mathcal{G}^*$ if and only if

A0* $\frac{1}{C^*} \leq \frac{P_0}{Q_0} \leq C^*$, and

A1* $\int_S (p(x) + q(x))\left|\log \frac{p(x)}{q(x)}\right|^2 dx \leq C^* \int_S (p(x)^{1/2} - q(x)^{1/2})^2\, dx$.

14

Condition A1* is similar to A2 and A3 in the definition of the set of regular distributions $\mathcal{G}_{\Phi,C,c_1,c_2,r,t}$ that appears in the upper bound (Theorem 5.1). In fact, if $\int_S (p^{1/2} - q^{1/2})^2 dx$ is bounded away from 0, then there exists C^* such that A1* is equivalent to A2. Thus, although $\mathcal{G}_{C^*}^*$ is in general not a superset of $\mathcal{G}_{\Phi,C,c_1,c_2,r,t}$, the set $\mathcal{G}_{C^*}^* \cap \mathcal{G}_{\Phi,C,c_1,c_2,r,t}$ contains important and interesting examples. For instance, any family that satisfies the conditions of Proposition 5.1 belongs to the intersection, as is verified in the proof (cf. Appendix E).

Theorem 5.2. *Let $C^* \geq 1$ and let $\sigma_0 : [n] \to [K]$ be a clustering such that one cluster is of size $\frac{n}{\beta K}$ and another is of size $\frac{n}{\beta K} + 1$. Let I'_n be any sequence such that $nI'_n \to \infty$, and let $C = 2\log 2$. Then there exists $\zeta_n \to 0$ and $c' > 0$ such that, for any permutation equivariant algorithm $\hat{\sigma}$,*

$$\inf_{\substack{((P_0,p),(Q_0,q)) \in \mathcal{G}_{C^*}^* \\ I'_n \leq I((P_0,p),(Q_0,q)) \leq C}} \mathbb{E}_{\substack{(P_0,p) \\ (Q_0,q)}} \left[\ell\big(\hat{\sigma}(A), \sigma_0\big) \right] \exp\left(\frac{n}{\beta K} I((P_0,p),(Q_0,q))(1+\zeta_n) \right) \geq c'.$$

Furthermore, for any $c > 0$, there exists $c' > 0$ such that for any permutation equivariant algorithm $\hat{\sigma}$,

$$\inf_{\substack{((P_0,p),(Q_0,q)) \in \mathcal{G}_{C^*}^* \\ I((P_0,p),(Q_0,q)) \leq c/n}} \mathbb{E}_{\substack{(P_0,p) \\ (Q_0,q)}} \left[\ell\big(\hat{\sigma}(A), \sigma_0\big) \right] \geq c'.$$

Theorem 5.2 shows that if $nI_n \to \infty$, the misclustering risk of any permutation equivariant algorithm is at least $\exp\left(-(1+o(1)) \frac{nI\big((P_0,p),(Q_0,q)\big)}{\beta K} \right)$. If $nI\big((P_0,p),(Q_0,q)\big) = O(1)$, any permutation invariant algorithm is inconsistent.

Remark 5.4. *Rather than being a minimax lower bound that applies to the worst case, Theorem 5.2 applies to* any *parameter $((P_0,q),(Q_0q)) \in \mathcal{G}_{C^*}^*$; we thus have an infimum over the parameter space rather than a supremum. This is possible because the permutation equivariance condition excludes the trivial case where $\hat{\sigma} = \sigma_0$.*

Proof sketch of Theorem 5.2: The full proof of the theorem is provided in Appendix G; we highlight key points here. The proof borrows elements from Yun and Proutiere [42] and Zhang and Zhou [43]. One key difference is that Theorem 5.2 holds for any parameters in the parameter space, rather than adopting a minimax framework, as in Zhang and Zhou [43], or assuming a prior on σ_0, as in Yun and Proutiere [42].

Crucial to the proof is the notion of a *misclustered* node. Let

$$S_K[\hat{\sigma}(A), \sigma_0] := \operatorname*{arg\,min}_{\rho \in S_K} d_H(\rho \circ \hat{\sigma}(A), \sigma_0).$$

It is straightforward to define a set of misclustered nodes when $S_K[\hat{\sigma}(A), \sigma_0]$ is a singleton, but care must be taken when $S_K[\hat{\sigma}(A), \sigma_0]$ contains multiple elements. We define the set of *misclustered nodes* as

$$\mathcal{E}[\hat{\sigma}(A), \sigma_0] := \left\{ v : (\rho \circ \hat{\sigma}(A))(v) \neq \sigma_0(v), \text{ for some } \rho \in S_K[\hat{\sigma}(A), \sigma_0] \right\}. \tag{5}$$

To see an example of the subtlety that arises when $S_K[\hat{\sigma}(A), \sigma_0]$ is not a singleton, note that the "for some $\rho \in S_K[\hat{\sigma}(A), \sigma_0]$" qualifier in the definition of $\mathcal{E}[\hat{\sigma}(A), \sigma_0]$ cannot be replaced with "for all $\rho \in S_K[\hat{\sigma}(A), \sigma_0]$." Otherwise, in the case where the clusters in σ_0 all have the same size, and where $\hat{\sigma}(A)$ is a trivial algorithm that maps all nodes to cluster 1, the set $S_K[\hat{\sigma}(A), \sigma_0]$ equals S_K and $\mathcal{E}[\hat{\sigma}(A), \sigma_0]$ would be empty.

With the definition of $\mathcal{E}[\hat{\sigma}(A), \sigma_0]$, we may formalize symmetry properties of permutation equivariant estimators. For example, if A is distributed according to a weighted SBM with σ_0 as the true

cluster assignment, and if nodes u and v lie in clusters of equal sizes, then $P(u \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]) = P(v \in \mathcal{E}[\hat{\sigma}(A), \sigma_0])$ for any permutation equivariant $\hat{\sigma}$ (cf. Corollary G.1).

Without loss of generality, let cluster 1 and 2 be the clusters in σ_0 that have sizes $\frac{n}{\beta K} + 1$ and $\frac{n}{\beta K}$, respectively, and let node 1 belong to cluster 1. Let σ^* be a random cluster assignment where $\sigma^*(u) = \sigma_0(u)$ for all $u \neq 1$, and $\sigma^*(1)$ is 1 or 2 with probability $\frac{1}{2}$ each. Let $\mathbb{P}_{SBM}(\cdot \mid \sigma^*)$ denote the distribution on $\mathbb{R}^{n \times n}$ induced by the random cluster assignment σ^*. We perturb $\mathbb{P}_{SBM}(\cdot \mid \sigma^*)$ to define a new distribution $\mathbb{P}_\Psi(\cdot)$ on $\mathbb{R}^{n \times n}$, where under P_Ψ, the A_{uv}'s are independent; and if $u = 1$ and v is in cluster 1 or 2, then A_{uv} is distributed according to a new probability distribution Y^* instead of P or Q. If $u \neq 1$ or if v is not in cluster 1 or 2, then A_{uv} is distributed according to $P_{SBM}(A \mid \sigma^*)$. We take $Y^* := \arg\min_Y \max \left\{ \int_S \log \frac{dY}{dP} dY, \int_S \log \frac{dY}{dQ} dY \right\}$, constructed to be similar to both P and Q (see Lemma G.2).

Under $\mathbb{P}_\Psi(\cdot)$, we can show that it is impossible to consistently cluster node 1 with respect to σ^* as the true cluster assignment. We then use the fact that $\mathbb{P}_\Psi(A)$ and $\mathbb{P}_{SBM}(A \mid \sigma^*)$ are similar to deduce the difficulty of correctly clustering node 1 under $\mathbb{P}_{SBM}(A \mid \sigma^*)$. Permutation equivariance translates this into a result on the number of misclustered nodes in a way similar to Lemma 2.1 in Zhang and Zhou [43]—the distinction being that Zhang and Zhou [43] uses a uniform prior over the true clustering to transform arbitrary estimators into permutation equivariant ones, whereas we place no prior on the true clustering, but restrict estimators to be permutation equivariant. Finally, we finish the proof by using permutation equivariance again to show that the misclustering error does not depend on whether $\sigma^*(1) = 1$ or $\sigma^*(1) = 2$.

5.3 Adaptivity

Let $\mathcal{F}_n^{p.e.}$ be the class of permutation equivariant clustering algorithm on networks with n nodes. Theorems 5.1 and 5.2 directly imply the following corollary, which sharply characterizes the optimal performance of $\mathcal{F}_n^{p.e.}$:

Corollary 5.1. *Let $\sigma_0 : [n] \to [K]$, and suppose one cluster is of size $\frac{n}{\beta K}$ and another is of size $\frac{n}{\beta K} + 1$. Let $C^*, C \geq 1, c_1, c_2 > 0, r > 0$, and $t \in (2/r, 1)$, and let Φ be a transformation function. Write $\mathcal{G}_\Phi := \mathcal{G}_{\Phi, C, c_1, c_2, r, t}$, $\mathcal{G}^* := \mathcal{G}_{C^*}^*$, and $\Lambda := \{\beta, K, C^*, C, c_1, c_2, r, t, \Phi\}$. Let $\left((P_{0,n}, p_n), (Q_{0,n}, q_n)\right) \in \mathcal{G}_\Phi \cap \mathcal{G}^*$ for every $n \in \mathbb{N}$.*

(i) If $\limsup_n I\left((P_{0,n}, p_n), (Q_{0,n}, q_n)\right) \frac{n}{\beta K \log n} \leq 1$, there exists $\zeta_n \to 0$, depending only on Λ, such that

$$\inf_{\hat{\sigma} \in \mathcal{F}_n^{p.e.}} \mathbb{E}_{\substack{(P_{0,n}, p_n) \\ (Q_{0,n}, q_n)}} \left[l(\hat{\sigma}(A), \sigma_0) \right] = \exp\left(-\frac{nI\left((P_{0,n}, p_n), (Q_{0,n}, q_n)\right)}{\beta K} (1 + \zeta_n) \right).$$

(ii) If $\liminf_n I\left((P_{0,n}, p_n), (Q_{0,n}, q_n)\right) \frac{n}{\beta K \log n} > 1$, there exists $\zeta_n \to 0$, depending only on Λ, such that $\inf_{\hat{\sigma} \in \mathcal{F}_n^{p.e.}} \mathbb{P}_{(P_{0,n}, p_n), (Q_{0,n}, q_n)}\left(l(\hat{\sigma}(A), \sigma_0) > 0 \right) \leq \zeta_n$.

(iii) If there exists $c > 0$ such that $\limsup_n I\left((P_{0,n}, p_n), (Q_{0,n}, q_n)\right) n < c$, there exists $c' > 0$ such that $\liminf_{n \to \infty} \inf_{\hat{\sigma} \in \mathcal{F}_n^{p.e.}} \mathbb{E}_{(P_{0,n}, p_n), (Q_{0,n}, q_n)}\left[l(\hat{\sigma}(A), \sigma_0) \right] > c'$.

The algorithm $\hat{\sigma}$ described in Section 4.1 with discretization level L_n diverging sufficiently slowly achieves the optimal rate in part (i) and (ii) for *any* $\left((P_{0,n}, p_n), (Q_{0,n}, q_n)\right) \in \mathcal{G}_\Phi \cap \mathcal{G}^*$. Thus, $\hat{\sigma}$ adapts to the edge probabilities $P_{0,n}$ and $Q_{0,n}$ and the edge weight densities p_n and q_n: Although $\hat{\sigma}$ has no knowledge of the parameters $\left((P_{0,n}, p_n), (Q_{0,n}, q_n)\right)$, it achieves the same optimal rate as if $\left((P_{0,n}, p_n), (Q_{0,n}, q_n)\right)$ were known.

In particular, this implies that one does not have to pay a price for taking the nonparametric approach. This seemingly counterintuitive phenomenon arises because the cost of discretization is reflected in the lower-order ζ_n term in the exponent. As an illustrative example, suppose $1 - P_{0,n} = 1 - Q_{0,n} = a\frac{\log n}{n}$ for some $a > 0$, and the densities p_n and q_n are of $N(\mu_1, \sigma_1^2)$ and $N(\mu_0, \sigma_0^2)$,

respectively. Then $I_n = (1 + o(1))\frac{a \log n}{n}\theta$, where $\theta = 2\left(1 - \sqrt{\frac{2\sigma_1^2\sigma_0^2}{\sigma_1^2+\sigma_0^2}}e^{-\frac{1}{4}\frac{(\mu_1-\mu_0)^2}{\sigma_1^2+\sigma_0^2}}\right)$, and the optimal

rate is $n^{-(1+o(1))\frac{2\theta}{\beta K}}$, which is attained by the nonparametric discretization estimator $\hat{\sigma}$.

Similarly, if $1 - P_{0,n} = 1 - Q_{0,n} = \frac{a \log n}{n}$ and the densities p_n and q_n are $\text{Exp}(\lambda_1)$ and $\text{Exp}(\lambda_0)$, respectively, then $I_n = (1 + o(1))\frac{\log n}{n}\theta'$, where $\theta' = 2\left(1 - \sqrt{\frac{\lambda_1\lambda_0}{\lambda_1+\lambda_0}}\right)$. The optimal rate $n^{-(1+o(1))\frac{2\theta'}{\beta K}}$ is again achieved by the nonparametric discretization estimator $\hat{\sigma}$.

6 Proof sketch: Recovery algorithm

A large portion of the Appendix is devoted to proving that our recovery algorithm succeeds and achieves the optimal error rates. We provide an outline of the proofs here.

We divide our argument into propositions that focus on successive stages of our algorithm. A birds-eye view of our method reveals that it contains two major components: (1) convert a weighted network into a labeled network, and then (2) run a community recovery algorithm on the labeled network. The first component involves two steps, transformation and discretization. Step (1) comprises the red and green steps in Figure 1 and outputs an adjacency matrix with discrete edge weights. Step (2) is denoted in blue.
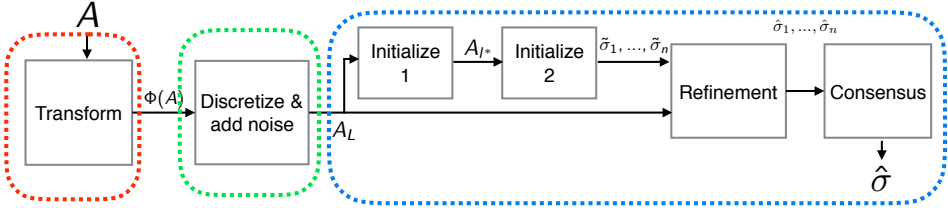

Figure 2: Analysis of the right-most blue region is contained in Section 6.1, of the middle green region in Section 6.2, and of the left-most red region in Section 6.3
.

In our algorithm, we use a single discretization level L throughout for ease of presentation. In practice, one could use different discretization levels for the initialization stage and for the refinement stage. By comparing Proposition 6.1, Proposition 6.2, and Theorem 5.1, we can see that the bias introduced by discretization is a second-order effect compared to the variance, which is why the discretization level should be small in both stages. The discretization level for the initialization stage can, however, be chosen to be larger than that of the refinement stage, because the initialization stage aims to produce a consistent estimator rather than an optimal one, and can thus tolerate greater variance. More precisely, the theoretical requirements on discretization for the initialization stage are $L \to \infty$ and $\frac{nI'_n}{L} \to \infty$, whereas the requirements for the refinement stage are $L \to \infty$ and $\frac{nI'_n}{Le^{L^{\tau/2}}} \to \infty$ (note that I'_n is defined in Theorem 5.1); L is required to be of smaller order to control the ratio $\frac{P_l}{Q_l}$ of the discretized probabilities.

6.1 Analysis of community recovery on a labeled network

We first examine the second component of our algorithm, which is a subroutine (right-most region in Figure 1) for recovering communities in a network where the edges have discrete labels $l = 1, \ldots, L_n$. The following proposition characterizes the rate of convergence of the output of the subroutine, where within-community edges are assigned edge labels with probabilities $\{P_l\}$, and between-community

edges are assigned edge labels according to $\{Q_l\}$. For convenience, if an edge does not exist between u and v, we assign the label 0 to A_{uv}, so P_0 and Q_0 are the edge absence probabilities.

Formally, for $L \in \mathbb{N}$, define $\mathcal{P}_L := \{(P_0, ..., P_L) \in [0, 1]^{L+1} : \sum_{l=1}^L P_l = 1\}$. For a clustering $\sigma_0 : [n] \to [K]$ and $(\{P_l\}, \{Q_l\}) \in \mathcal{P}_L^2$, we define a Labeled Stochastic Block Model $LSBM(\sigma_0, \{P_l\}, \{Q_l\})$ as a distribution on $\{0, \ldots, L\}^{n \times n}$ such that if $A \sim LSBM(\sigma_0, \{P_l\}, \{Q_l\})$, then for any $u, v \in [n]$ such that $u > v$,

$$A_{uv} \sim \begin{cases} \{P_l\} & \text{if } \sigma_0(u) = \sigma_0(v), \\ \{Q_l\} & \text{if } \sigma_0(u) \neq \sigma_0(v). \end{cases}$$

For $\rho > 1$, let $\mathcal{G}_{L,\rho} \subset \mathcal{P}_L^2$ be such that $(\{P_l\}, \{Q_l\}) \in \mathcal{G}_{L,\rho}$ if and only if $\frac{1}{\rho} \leq \frac{P_l}{Q_l} \leq \rho$ for all $l = 0, ..., L$. For a pair $(\{P_l\}, \{Q_l\}) \in \mathcal{P}_L$, we define $I(\{P_l\}, \{Q_l\}) := -2 \log \sum_{l=0}^L \sqrt{P_l Q_l}$.

In the next proposition, for a given clustering σ_0 and $(\{P_l\}, \{Q_l\}) \in \mathcal{P}_L^2$, we let the random network A have the distribution $LSBM(\sigma_0, \{P_l\}, \{Q_l\})$.

Proposition 6.1. *Let $\sigma_0 \in \mathcal{C}(\beta, K)$. Let $\{I_n, I'_n, \rho_n, L_n\}_{n \in \mathbb{N}}$ be any sequences such that $I_n \to 0$, $\rho_n \geq 2$, $L_n \geq 1$, and $\frac{nI'_n}{(L_n+1)\rho_n^2 \log \rho_n} \to \infty$. Then there exists a sequence $\zeta_n \to 0$ such that*

$$\lim_{n \to \infty} \sup_{\substack{(\{P_l\}, \{Q_l\}) \in \mathcal{G}_{L_n, \rho_n} \\ I'_n \leq I(\{P_l\}, \{Q_l\}) \leq I_n}} \mathbb{P}_{(\{P_l\}, \{Q_l\})} \left(l(\hat{\sigma}(A), \sigma_0) \leq \exp\left(-(1 - \zeta_n)\frac{n}{\beta K} I(\{P_l\}, \{Q_l\})\right) \right) = 1.$$

Furthermore, if $\frac{nI_n}{\beta K \log n} \leq 1$, then

$$\sup_{\substack{(\{P_l\}, \{Q_l\}) \in \mathcal{G}_{L_n, \rho_n} \\ I'_n \leq I(\{P_l\}, \{Q_l\}) \leq I_n}} \mathbb{E}\left[l(\hat{\sigma}(A), \sigma_0)\right] \exp\left((1 - \zeta_n)\frac{n}{\beta K} I(\{P_l\}, \{Q_l\})\right) \leq 1.$$

Remark 6.1. *This result resembles that of Yun and Proutiere [42], who also study an SBM where the edges carry discrete labels. They state their results using a seemingly different divergence, but it coincides with the Renyi divergence when specialized to our setting (cf. Lemma G.2). Proposition 6.1 differs critically from Yun and Proutiere [42] in two respects, however. First, they hold the number of labels L_n to be fixed and assume that the bound ρ_n on the probability ratio $\frac{P_{l,n}}{Q_{l,n}}$ is fixed, whereas we allow both L_n and ρ_n to diverge. Second, they assume that $\sum_{l=1}^{L_n}(P_{l,n} - Q_{l,n})^2$ is sufficiently large when compared to $\max_{l=1,...,L_n} P_{l,n}$, whereas we do not make any assumptions of this form. These generalizations are crucial in analyzing the weighted SBM, since in order to achieve consistency for continuous distributions, the discretization level L_n and the bound ρ_n must increase with n.*

6.2 Discretization of the Renyi divergence

We now analyze the discretization step of the algorithm (green box in Figure 1). The input to this step is the weighted network $\Phi(A)$ in which all the edge weights are in $[0, 1]$. We use $\tilde{p}(z)$ and $\tilde{q}(z)$ for $z \in [0, 1]$ to denote the densities of the transformed edge weights; the next section shows the relationship between $\tilde{p}(z)$ and $p(x)$ and $\tilde{q}(z)$ and $q(x)$. The discretization step of the algorithm divides $[0, 1]$ into L_n uniform bins, denoted by $[a_l, b_l]$, for $1 \leq l \leq L_n$. The output is a network A_{L_n}, where each edge is assigned label $l = 1, \ldots, L_n$ with probability either

$$P_l := (1 - P_0) \int_{a_l}^{b_l} \tilde{p}(z)dz, \quad \text{or} \quad Q_l := (1 - Q_0) \int_{a_l}^{b_l} \tilde{q}(z)dz. \tag{6}$$

A missing edge is assigned the label 0. It is easy to show that discretization always leads to a loss of information; i.e., $I(\{P_l\}, \{Q_l\}) \leq I((P_0, \tilde{p}), (Q_0, \tilde{q}))$.

Let $\tilde{\mathcal{P}}$ denote the set of probability distributions on $[0, 1]$ whose singular part is a point mass at 0. Let $\tilde{C} \in (0, \infty)$, $\tilde{c}_1, \tilde{c}_2 \in (0, 1/2)$, $r > 2$, and $t > 0$, and define the set $\tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t} \subset \tilde{\mathcal{P}}^2$ such that $((P_0, \tilde{p}), (Q_0, \tilde{q})) \in \tilde{\mathcal{G}}$ if and only if the following hold:

C0 We have $\frac{1}{\tilde{C}} \leq \frac{1-P_0}{1-Q_0} \leq \tilde{C}$ and $\frac{1}{\tilde{C}} \leq \frac{P_0}{Q_0} \leq \tilde{C}$.

C1 For all $z \in (0,1)$, we have $0 < \tilde{p}(z), \tilde{q}(z) \leq \tilde{C}$.

C2 There exists a quasi-convex $\tilde{g} : [0,1] \to [0,\infty)$ such that $\tilde{g}(z) \geq \left|\log \frac{\tilde{p}(z)}{\tilde{q}(z)}\right|$ and $\int_0^1 \tilde{g}(z)^r \, dz \leq \tilde{C}$.

C3 Denoting $\tilde{\alpha} := \left\{ \int_0^1 (\sqrt{\tilde{p}(z)} - \sqrt{\tilde{q}(z)})^2 \, dz \right\}^{1/2}$ and $\tilde{\gamma}(z) := \frac{\tilde{p}(z) - \tilde{q}(z)}{\tilde{\alpha}}$, we have

$$\int_0^1 \left\{ \frac{\tilde{\gamma}(z)}{\tilde{p}(z) + \tilde{q}(z)} \right\}^r (\tilde{p}(z) + \tilde{q}(z)) \, dz \leq \tilde{C}.$$

C4 There exists a quasi-convex function $\tilde{h} : [0,1] \to [0,\infty)$ such that

$$\tilde{h}(z) \geq \max \left\{ \left| \frac{\tilde{\gamma}(z)}{\tilde{p}(z) + \tilde{q}(z)} \right|, \left| \frac{\tilde{p}'(z)}{\tilde{p}(z)} \right|, \left| \frac{\tilde{q}'(z)}{\tilde{q}(z)} \right|, \left| \frac{\tilde{\gamma}'(z)}{\tilde{p}(z) + \tilde{q}(z)} \right| \right\}$$

and $\int_0^1 \tilde{h}(z)^t dz < \tilde{C}$.

C5 We have $\tilde{p}'(z), \tilde{q}'(z) \geq 0$ for all $z < \tilde{c}_1$, and $\tilde{p}'(z), \tilde{q}'(z) \leq 0$ for all $z > 1 - \tilde{c}_2$.[3]

Proposition 6.2. *Let* $\tilde{C} \in (0,\infty)$, $\tilde{c}_1, \tilde{c}_2 \in (0,1/2)$, $r > 2$, *and* $t > 0$. *For any* $((P_0, \tilde{p}), (Q_0, \tilde{q})) \in \tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t}$, *for any* $L \in \mathbb{N}$ *such that* $L \geq \tilde{c}_1^{-1} \vee \tilde{c}_1^{-1}$, *and for* $\{P_l, Q_l\}$ *defined in equation* (6), *we have* $\frac{1}{2\tilde{C}\exp((2\tilde{C}L)^{1/r})} \leq \frac{P_l}{Q_l} \leq 2\tilde{C}\exp((2\tilde{C}L)^{1/r})$ *for all* $l \in \{0, \dots L\}$. *Furthermore,*

$$\lim_{L \to \infty} \sup_{((P_0, \tilde{p}), (Q_0, \tilde{q})) \in \tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t}} \left| 1 - \frac{I(\{P_l\}, \{Q_l\})}{I\big((P_0, \tilde{p}), (Q_0, \tilde{q})\big)} \right| = 0.$$

We prove Proposition 6.2 in Appendix C.

6.3 Analysis of the transformation function

Proposition 6.2 considers densities supported on $[0,1]$. In conjunction with Proposition 6.1, this suffices to obtain Theorem 5.1, because the densities of the transformed edge weights are compactly supported and, importantly, the Renyi divergence is invariant with respect to the transformation function Φ.

To be precise, let $p(\cdot)$ and $q(\cdot)$ denote probability densities on S, and for $X \sim p$ and $Y \sim q$, let $\tilde{p}(\cdot)$ and $\tilde{q}(\cdot)$ denote the densities of $\Phi(X)$ and $\Phi(Y)$. We then have $\tilde{p}(z) = \frac{p(\Phi^{-1}(z))}{\phi(\Phi^{-1}(z))}$ and $\tilde{q}(z) = \frac{q(\Phi^{-1}(z))}{\phi(\Phi^{-1}(z))}$ for $z \in [0,1]$. Therefore, via the change of variable $z = \Phi^{-1}(x)$, we have

$$\int_S \sqrt{p(x)q(x)} \, dx = \int_0^1 \sqrt{\tilde{p}(z)\tilde{q}(z)} \, dz, \text{ and}$$
$$I\big((P_0, p), (Q_0, q)\big) = I\big((P_0, \tilde{p}), (Q_0, \tilde{q})\big).$$

7 Simulation studies

We start with a toy example that illustrates the intuition behind our discretization-based algorithm. In this example, we have $n = 1000$ nodes, $K = 2$ clusters, and $P_0 = Q_0 = 0.5$. We also set $p(\cdot)$ and $q(\cdot)$ as the normal density $N(0, 1.3^2 + 1)$ and mixture of normals $\frac{1}{2}N(-1.3, 1) + \frac{1}{2}N(1.3, 1)$, respectively (see Figure 3). Observe that $\int_{\mathbb{R}} x dP = \int_{\mathbb{R}} x dQ = 0$ and $\int_{\mathbb{R}} x^2 dP = \int_{\mathbb{R}} x^2 dQ = \frac{1}{2}(1.3^2 + 1)$. The true clustering σ_0 maps the first 500 nodes to cluster 1 and the rest to cluster 2.

---

[3]If $\tilde{g}$ is non-decreasing, we need only $\tilde{p}'(z), \tilde{q}'(z) \leq 0$ for all $z > 1 - c_2'$.
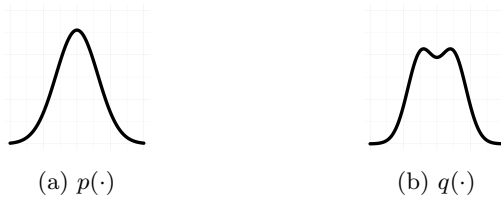
(a) $p(\cdot)$　　　　　　(b) $q(\cdot)$

Figure 3

In Figure 4a, we generate a random weighted network A and display the adjacency matrix *without randomly permuting the rows and columns*. It is difficult to discern the block structure because (P_0, p) and (Q_0, q) have equal mean and variance. In Figure 4b, 4c, and 4d, we discretize A using the transformation $\Phi(x) = \int_{-\infty}^{x} \frac{1}{4} e^{-|t|/2} \, dt$ and $L = 3$ bins and show the discretized network A^1, A^2, A^3; recall that A^1 is a binary adjacency matrix, where $A^1_{uv} = 1$ if $A_{uv} \neq 0$ and $\phi(A_{uv}) \in [0, 1/3)$, and $A^1_{uv} = 0$ otherwise, and likewise for A^2 and A^3. We observe that the block structure is clearly distinguishable in A^2 because the densities $p(\cdot)$ and $q(\cdot)$ differ most around the origin; the block structure is somewhat visible in A^1 and A^3, but to a lesser extent. These figures illustrate why the discretization and initialization stages are useful.

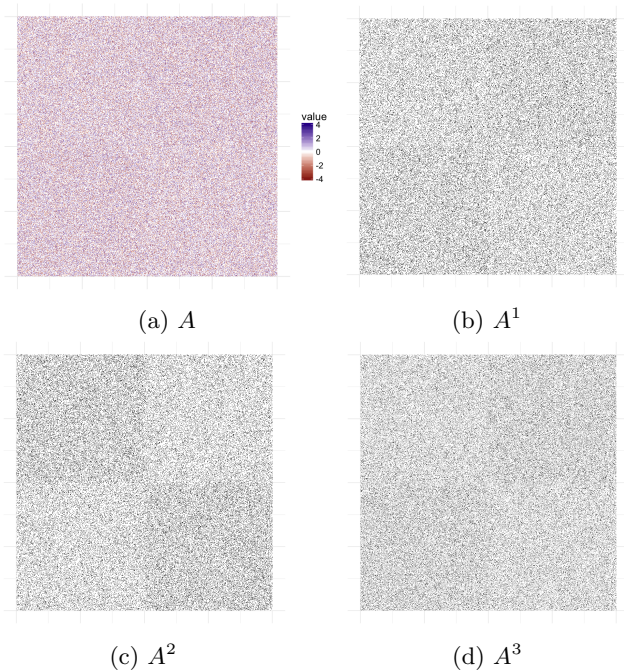(a) A　　　　　　(b) A^1

(c) A^2　　　　　　(d) A^3

Figure 4

In Figure 5a, we test how the performance of our algorithm scales with the network size n. We use the same setting as our first simulation, except we let $n \in \{400, 600, 800, \ldots, 2000\}$ and $L_n = \lfloor 0.4(\log(\log n))^4 \rfloor$. For each value of n, we perform 100 trials, where we generate a random network A, perform our clustering algorithm, and calculate the misclustering error. The misclustering errors are averaged across the 100 random trials and the aggregated medians are shown, with deviations, in Figure 5a. In Figure 5a, we observe the same threshold behavior that arises in the unweighted setting: the misclustering error is around 0.5—equivalent to random guessing—for low n, and drops sharply to 0 as the value of n passes a threshold (around $n = 1000$ in this case). We note that

for this and our next simulation study, we use a simplified version of our algorithm as described in Remark 4.2; we observed no difference in performance between the full version and the simplified version of the algorithm.
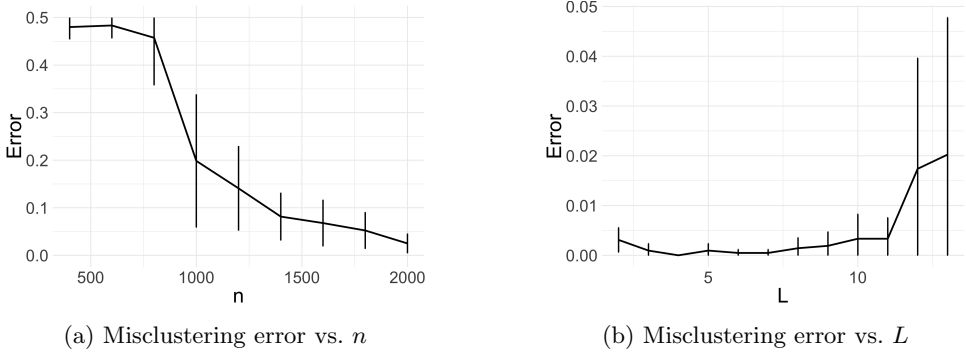

(a) Misclustering error vs. n (b) Misclustering error vs. L

Figure 5

In Figure 5b, we study the sensitivity of our algorithm to the choice of discretization level L. We let $K = 3$, $n = 2100$, $P_0 = 0.3$, and $Q_0 = 0.27$, and let $p(\cdot)$ be the density of $N(0.3, 0.8^2)$, and $q(\cdot)$ be the density of $N(0, 1)$. We let $L \in \{1, 2, 3, \ldots, 12, 13\}$ and, for each setting of L, we perform 100 random trials in which we generate a random network A, perform our clustering algorithm, and calculate the misclustering error; the results are shown in Figure 5b; the error for $L = 1$, in which we discard the edge weights, exceeds 0.56 and is thus omitted from the plot. We observe that the algorithm performs best when L is chosen to be small, though not too small, as is suggested by our theoretical analysis.

In Figure 6a, we compare our approach against treating a weighted network as an unweighted one by discarding the edge weights. In this setting, we let $n = 1500$, $P_0 = 0.3$, $Q_0 = 0.23$, and $K = 3$. We choose $q(\cdot)$ as the density of $N(0, 1)$ and $p(\cdot)$ as the density of $N(\mu, 1)$ where we let $\mu \in \{0, 0.05, 0.1, 0.15, 0.2, 0.25\}$. We perform 100 trials and aggregate the result in Figure 6a. In red, we plot the misclustering proportion error incurred by our WSBM clustering algorithm with $L = 5$; in blue, we plot the misclustering error incurred by ignoring the edge weights entirely and treating the network as an unweighted one. As we expect, when μ is close to 0, the edge weights are uninformative and it is better to ignore the edge weights. As μ increases, however, the advantage of using the weights become significant.

In Figure 6b, we compare our algorithm against clustering an unweighted network formed by optimally thresholding the edge weights. We let $K = 3$, $P_0 = 0.3$, and $Q_0 = 0.27$, and let $p(\cdot)$ be the density of $N(0.3, 0.8)$ and $q(\cdot)$ be the density of $N(0, 1)$. For $\tau \in \mathbb{R}$, we define the thresholded
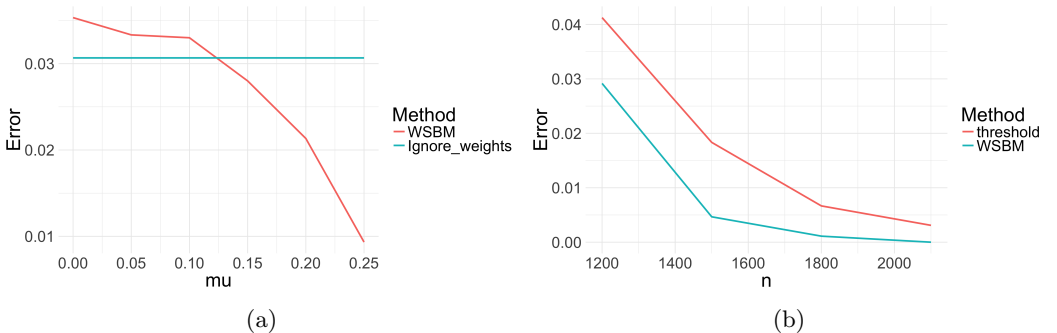

(a)   (b)

Figure 6

network $A_\tau \in \{0,1\}^{n \times n}$ as $A_{\tau,uv} = 1$ if $A_{uv} \neq 0$ and $A_{uv} \geq \tau$, and $A_{\tau,uv} = 0$ if $A_{uv} = 0$ or if $A_{uv} < \tau$. For each $\tau \in \{-2, -1.8, -1.6, \ldots, 1.6, 1.8, 2.0\}$, we form A_τ, extract the cluster, and compute the misclustering error. We then report the lowest misclustering error among all A_τ for $\tau \in \{-2, -1.8, -1.6, \ldots, 1.6, 1.8, 2.0\}$ as the red line in Figure 6b; this approach is of course impossible to implement in practice, and we use it only for the purpose of comparison. The turquoise line is the misclustering error incurred by our algorithm, using $L_n = \lfloor 0.4(\log(\log n))^4 \rfloor$.

8   Conclusion

We have provided a rate-optimal community estimation algorithm for the homogeneous weighted stochastic block model. Our algorithm includes a preprocessing step consisting of transforming and discretizing the (possibly) continuous edge weights to obtain a simpler graph with edge weights supported on a finite, discrete set. This approach may be useful for other network data analysis problems involving continuous distributions, where discrete versions of the problem are simpler to analyze.

Our paper provides a step toward understanding the weighted SBM under the same mathematical framework that has been exceptionally fruitful in the case of unweighted models. It is far from comprehensive, however, and many open questions remain. We describe a few here:

1. An important extension is the *heterogenous* stochastic block model, where edge weight distributions depend on the exact community assignments of both endpoints. In such a setting, Abbe and Sandon [5] and Yun and Proutiere [42] have shown that a generalized information divergence—the CH divergence—governs the intrinsic difficulty of community recovery. We believe that a similar discretization-based approach should lead to analogous results in the case of a heterogeneous weighted SBM.

2. Real-world networks often have nodes with very high degrees, which may adversely affect the accuracy of recovery methods for the stochastic block model. To solve this problem, degree-corrected SBMs [45, 16] have been proposed as an effective alternative to regular SBMs. It remains to extend the concept of degree-correction to the weighted SBM.

Acknowledgments

A   Proof of Proposition 6.1

We structure the proof according to the flow of our algorithm. Since this proposition addresses the case of discrete labels, we do not need to consider the "transformation and discretization" step. We will prove the proposition by constructing a sequence $\zeta_n \to 0$ such that the statements of the proposition are satisfied.

Since $\frac{nI'_n}{(L_n+1)\rho_n^2 \log \rho_n} \to \infty$ and $I_n \to 0$, there exists $N_0 \in \mathbb{N}$ such that for all $n \geq N_0$, we have

$$\frac{n-1}{\rho_n^2 \beta^2 K^5 \log(\rho_n^2 \beta K)} \frac{I'_n}{L_n+1} \geq c_{init}^{-1} C_{spec2} C_{mis}, \tag{7}$$

$$\frac{n-1}{\beta^2 K^6} \frac{I'_n}{L_n+1} \geq c_{init}^{-1} C_{spec1}, \quad \frac{L_n+1}{n} \leq \frac{1}{8}, \quad \text{and } I_n \leq \frac{1}{2}, \tag{8}$$

where C_{spec1} and C_{spec2} are universal constants defined in Proposition B.3, c_{init} is a universal constant defined in Proposition B.11, and C_{mis} is a universal constant defined in Proposition B.12. For $n \leq N_0$, we define ζ_n such that $\exp\left(-\frac{nI_n}{\beta K}(1 + \zeta_n)\right) = 1$.

Now, suppose $n \geq N_0$, so inequalities (7) and (8) hold. Let us arbitrarily fix $(\{P_l\}, \{Q_l\}) \in \mathcal{G}_{L_n, \rho_n}$ such that $I'_n \leq I(\{P_l\}, \{Q_l\}) \leq I_n$. We consider Algorithm 4.2. Let $\delta = 2\frac{L+1}{n}$ and define

$$P'_l := P_l(1 - \delta) + \frac{\delta}{L+1}, \quad \text{and} \quad Q'_l := Q_l(1 - \delta) + \frac{\delta}{L+1}.$$

Since the input network A_L has the distribution $LSBM(\sigma_0, (\{P_l\}, \{Q_l\}))$, the output A_L of Algorithm 4.2 has the distribution $LSBM(\sigma_0, (\{P'_l\}, \{Q'_l\}))$. It is then clear that $P'_l, Q'_l \geq \frac{2}{n}$ for all $l \in \{0, \ldots, L+1\}$, and furthermore, by Lemma B.2, we know that

$$I(\{P'_l\}, \{Q'_l\}) \geq \frac{I(\{P_l\}, \{Q_l\})}{1 + 2I(\{P_l\}, \{Q_l\})}\left(1 - 4\frac{L+1}{n}\right) \geq \frac{I'_n}{4}, \tag{9}$$

$$I(\{P'_l\}, \{Q'_l\}) \leq I(\{P_l\}, \{Q_l\}) \leq I_n. \tag{10}$$

Let $l^* \in \{0, \ldots, L\}$ be the output of the first stage of INITIALIZATION (Algorithm 4.3). Define E_1 as the event that

$$\frac{\Delta_{l^*}^2}{P_{l^*} \vee Q_{l^*}} \geq c_{init}\frac{I(\{P'_l\}, \{Q'_l\})}{L+1}. \tag{11}$$

Since

$$\frac{nI(\{P'_l\}, \{Q'_l\})}{(L+1)\beta^2 K^6} \geq \frac{nI'_n}{4(L+1)\beta^2 K^6} \geq \frac{1}{4}c_{init}^{-1}C_{spec1} \geq 2C_{spec1},$$

$P'_l \vee Q'_l \geq \frac{1}{n}$, and $\frac{n}{\log n} \geq 2^{15} \vee 30\beta K$, we may apply Claim 3 of Proposition B.11 to conclude that $\mathbb{P}(E_1) \geq 1 - 12(L+1)^2 n^{-5}$.

Let $\{\tilde{\sigma}_u\}_{u \in [n]}$ be the initial clusterings output by the second stage of INITIALIZATION (Algorithm 4.3). Define E_2 as the event that

$$\text{for all } u \in [n], \, l_{\setminus\{u\}}(\tilde{\sigma}_u, \sigma_0) \leq C_{spec2}\frac{\beta K^4}{n-1}\frac{P_{l^*} \vee Q_{l^*}}{\Delta_{l^*}^2},$$

where $l_{\setminus\{u\}}(\cdot, \cdot)$ is the misclustering error defined on $[n]\setminus\{u\}$. For $u \in [n]$ and $l \in \{0, \ldots, L\}$, define $\tilde{\sigma}_u^{(l)}$ as the result of applying spectral clustering (with $\mu = 4\beta$ and $\tau = 40K\bar{d}$ where $\bar{d}$ is defined with respect to $A_{l,-u}$) on $A_{l,-u} \in \{0, 1\}^{(n-1)\times(n-1)}$, that is, the network excluding node u with only the edges whose label is l. By Proposition B.3 and a union bound, we have that, with probability at least $1 - Ln(n-1)^{-5}$,

$$\max_{u \in [n]} \quad \max_{l \in \{0, \ldots L\} : \frac{n}{\beta^2 K^6}\frac{(P_l - Q_l)^2}{P_l \vee Q_l} \geq C_{spec1}} l_{\setminus\{u\}}(\tilde{\sigma}_u^{(l)}, \sigma_0)C_{spec2}^{-1}\frac{n}{\beta K^4}\frac{(P_l - Q_l)^2}{P_l \vee Q_l} \leq 1.$$

Since, under event E_1,

$$\frac{n-1}{\beta^2 K^6}\frac{\Delta_{l^*}^2}{P_{l^*} \vee Q_{l^*}} \geq c_{init}\frac{nI(\{P'_l\}, \{Q'_l\})}{L_n + 1} \geq c_{init}\frac{n-1}{\beta^2 K^6}\frac{I'_n}{L_n + 1} \geq C_{spec1},$$

since $P_{l^*} \vee Q_{l^*} \geq \frac{1}{n-1}$, and since $n - 1 \geq 8\beta^2 K^2 \vee 2000$, we obtain

$$\mathbb{P}(E_1 \cap E_2) \geq 1 - 12(L+1)^2 n^{-5} - Ln(n-1)^{-5}. \tag{12}$$

23

Note that under event $E_1 \cap E_2$, for all $u \in [n]$, we have

$$l_{\backslash\{u\}}(\tilde{\sigma}_u, \sigma_0) \leq C_{spec2} \frac{\beta K^4}{n-1} \frac{P_{l^*} \vee Q_{l^*}}{\Delta_{l^*}^2} \leq c_{init}^{-1} C_{spec2} \frac{\beta K^4}{n-1} \frac{L_n + 1}{I(\{P_l'\}, \{Q_l'\})}$$

$$\leq 4c_{init}^{-1} C_{spec2} \frac{\beta K^4}{n-1} \frac{L_n + 1}{I_n'} \overset{(a)}{\leq} \left(C_{mis}\rho_n^2 \beta K \log(\rho_n^2 \beta K)\right)^{-1}, \tag{13}$$

where (a) follows from inequality (7). Since $C_{mis} > 16$ and $\rho_n \geq 1$, we have $l_{\backslash\{u\}}(\tilde{\sigma}_u, \sigma_0) < \frac{1}{16\beta K}$. Recall the definition (99) of $S_K[\tilde{\sigma}_u, \sigma_0]$. Since the smallest cluster of σ_0 is of size at least $\frac{n}{\beta K}$, we have by Lemma B.6 that $S_K[\tilde{\sigma}_u, \sigma_0]$ is a singleton; we let π_u denote the only element of $S_K[\tilde{\sigma}_u, \sigma_0]$.

Since $\hat{\sigma}_u = \tilde{\sigma}_u$ on $[n]\backslash\{u\}$, we thus have, for all $u \in [n]$, that

$$\frac{1}{n}d_H(\pi_u \circ \hat{\sigma}_u, \sigma_0) \leq \frac{1}{n}\left\{d_H(\pi_u \circ \tilde{\sigma}_u, \sigma_0) + 1\right\} \leq \frac{1}{16\beta K} \frac{n-1}{n} + \frac{1}{n} < \frac{1}{8\beta K}. \tag{14}$$

By Lemma B.6 again, we know that π_u is the only element of $S_K[\hat{\sigma}_u, \sigma_0]$.

Since the smallest cluster of σ_0 is of size at least $\frac{n}{\beta K}$, the smallest cluster of $\hat{\sigma}_u$ is of size at least $\frac{n}{2\beta K}$. Furthermore, we have

$$l(\hat{\sigma}_1, \hat{\sigma}_u) \leq \frac{1}{n}d_H(\pi_1 \circ \hat{\sigma}_1, \pi_u \circ \hat{\sigma}_u) \leq \frac{1}{n}\left(d_H(\pi_1 \circ \hat{\sigma}_1, \sigma_0) + d_H(\pi_u \circ \hat{\sigma}_u, \sigma_0)\right) < \frac{1}{4\beta K}. \tag{15}$$

Therefore, from Lemma B.6, we conclude that $\pi_1^{-1} \circ \pi_u$ is the only element of $S_K[\hat{\sigma}_u, \hat{\sigma}_1]$ and

$$\hat{\sigma}(u) = \arg\max_{k \in [K]} \left|\{v \in [n] : \hat{\sigma}_u(v) = \hat{\sigma}_u(u)\} \cap \{v \in [n] : \hat{\sigma}_1(v) = k\}\right| = (\pi_1^{-1} \circ \pi_u)(\hat{\sigma}_u(u)). \tag{16}$$

Define $\gamma_u = l_{\backslash\{u\}}(\tilde{\sigma}_u, \sigma_0)$, $\eta_u' = 2\beta K \gamma_u \log \frac{eK}{\gamma_u} + 12\frac{\log n}{n}$, and $\eta_u = 10(\sqrt{\eta_u'} + \eta_u')$. Because inequality (13) holds under $E_1 \cap E_2$ and also $P_l' \vee Q_l' \geq \frac{1}{n-1}$ and $\frac{n}{\log n} \geq 2^8 \rho_n^2$, we can apply Proposition B.12, for each $u \in [n]$, to obtain

$$\mathbb{P}\left\{\exists \pi \in S_K[\hat{\sigma}_1, \sigma_0], \pi(\hat{\sigma}(u)) \neq \sigma_0(u)\right\}$$

$$\leq \mathbb{P}\left\{\exists \pi \in S_K[\hat{\sigma}_1, \sigma_0], \pi(\hat{\sigma}(u)) \neq \sigma_0(u) \,|\, E_1 \cap E_2\right\} + P(E_1^c \cup E_2^c)$$

$$\overset{(a)}{=} \mathbb{P}\left\{\pi_1(\hat{\sigma}(u)) \neq \sigma_0(u) \,|\, E_1 \cap E_2\right\} + P(E_1^c \cup E_2^c)$$

$$\overset{(b)}{=} \mathbb{P}\left\{\pi_u^{-1}(\sigma_0(u)) \neq \hat{\sigma}_u(u) \,|\, E_1 \cap E_2\right\} + P(E_1^c \cup E_2^c)$$

$$\leq (K-1)\exp\left(-(1 - C_{err}\beta K \rho_n \eta_u)\frac{n}{\beta K}I(\{P_l'\}, \{Q_l'\})\right) + 5(L+1)n^{-6} + 12(L+1)^2 n^{-5} - Ln(n-1)^-$$

$$\leq (K-1)\exp\left(-(1 - C_{err}\beta K \rho_n \eta_u)\frac{n}{\beta K}I(\{P_l'\}, \{Q_l'\})\right) + n^{-3},$$

where (a) follows because $S_K[\hat{\sigma}_1, \sigma_0]$ is a singleton under $E_1 \cap E_2$, and (b) follows from equation (16).

Define $\zeta_n' := C_{err}\beta K \rho_n \max_{u \in [n]} \eta_u$. By the penultimate statement in inequality (13) and the assumption that $\frac{nI_n'}{(L+1)\rho_n^2 \log \rho_n} \to \infty$, we have $\beta K \log(\beta K) \max_{u \in [n]} \gamma_u \to 0$, so $\max_{u \in [n]} \eta_u \to 0$ and $\zeta_n' \to 0$.

Observe that

$$
\mathbb{E}\big[l(\hat{\sigma}, \sigma_0)\big] = \mathbb{E}\left[\min_{\pi \in S_K} \frac{1}{n} \sum_{u=1}^{n} \mathbf{1}\{(\pi \circ \hat{\sigma})(u) \neq \sigma_0(u)\}\right]
$$

$$
\leq \mathbb{E}\left[\min_{\pi \in S_K[\hat{\sigma}_1, \sigma_0]} \frac{1}{n} \sum_{u=1}^{n} \mathbf{1}\{(\pi \circ \hat{\sigma})(u) \neq \sigma_0(u)\}\right]
$$

$$
\leq \mathbb{E}\left[\frac{1}{n} \sum_{u=1}^{n} \mathbf{1}\big\{\exists \pi \in S_K[\hat{\sigma}_1, \sigma_0],\, (\pi \circ \hat{\sigma})(u) \neq \sigma_0(u)\big\}\right]
$$

$$
\leq \frac{1}{n} \sum_{u=1}^{n} \mathbb{P}\big\{\exists \pi \in S_K[\hat{\sigma}_1, \sigma_0],\, \pi(\hat{\sigma}(u)) \neq \sigma_0(u)\big\}
$$

$$
\leq \exp\left(-(1 - \zeta_n') \frac{n}{\beta K} I(\{P_l'\}, \{Q_l'\})\right) + n^{-3},
$$

$$
\leq \exp\left(-(1 - \zeta_n'') \frac{n}{\beta K} I(\{P_l\}, \{Q_l\})\right) + n^{-3},
$$

where in the last inequality, we define $\zeta_n'' := 1 - (1 - \zeta_n')\big(1 - \frac{I(\{P_l\}, \{Q_l\}) - I(\{P_l'\}, \{Q_l'\})}{I(\{P_l\}, \{Q_l\})}\big)$. By inequalities (10) and (9), we have

$$
0 \leq \frac{I(\{P_l\}, \{Q_l\}) - I(\{P_l'\}, \{Q_l'\})}{I(\{P_l\}, \{Q_l\})} \leq 1 - \frac{1}{1 + 2I_n}(1 - 4\frac{L+1}{n}),
$$

so since $I_n \to 0$ by assumption, we have $\zeta_n'' \to 0$. For the second claim of Proposition 6.1, let $\zeta_n = \zeta_n''$ for all $n \geq N_0$. It is then clear that if $\frac{n I_n}{\beta K \log n} \leq 1$, we have

$$
\mathbb{E}l(\hat{\sigma}, \sigma_0) \leq \exp\left(-(1 - \zeta_n) \frac{n}{\beta K} I(\{P_l\}, \{Q_l\})\right).
$$

Since $(\{P_l\}, \{Q_l\})$ was chosen arbitrarily, the second claim of the proposition follows.

For the first claim, let us first suppose that $\exp\left(-(1 - \zeta_n'') \frac{n I(\{P_l\}, \{Q_l\})}{\beta K}\right) \geq n^{-2}$. Define $\tilde{\zeta}_n := \zeta_n'' + \big(\frac{\beta K}{n I_{n'}}\big)^{1/2}$. Then

$$
\mathbb{P}\left\{l(\hat{\sigma}, \sigma_0) > \exp\left(-(1 - \tilde{\zeta}_n) \frac{n I(\{P_l\}, \{Q_l\})}{\beta K}\right)\right\} \leq \frac{\mathbb{E}l(\hat{\sigma}, \sigma_0)}{\exp\left(-(1 - \tilde{\zeta}_n) \frac{n I(\{P_l\}, \{Q_l\})}{\beta K}\right)}
$$

$$
\leq \exp\left((\zeta_n'' - \tilde{\zeta}_n) \frac{n I(\{P_l\}, \{Q_l\})}{\beta K}\right) + \frac{n^{-3}}{\exp\left(-(1 - \tilde{\zeta}_n) \frac{n I(\{P_l\}, \{Q_l\})}{\beta K}\right)}
$$

$$
\leq \exp\left(-\big(\frac{n I(\{P_l\}, \{Q_l\})}{\beta K}\big)^{1/2}\right) + n^{-1} \leq \exp\left(-\big(\frac{n I_n'}{\beta K}\big)^{1/2}\right) + n^{-1}.
$$

25

Let us now suppose that $\exp\left(-(1-\zeta_n'')\frac{nI(\{P_l\},\{Q_l\})}{\beta K}\right) < n^{-2}$. Then

$$\mathbb{P}\left\{l(\hat\sigma, \sigma_0) \geq \exp\left(-(1-\zeta_n'')\frac{nI(\{P_l\},\{Q_l\})}{\beta K}\right)\right\} \leq \mathbb{P}(l(\hat\sigma, \sigma_0) > 0)$$

$$\leq \mathbb{P}\left\{\min_{\pi \in S_K} d_H(\pi \circ \hat\sigma, \sigma_0) > 0\right\}$$

$$\leq \mathbb{P}\left\{\min_{\pi \in S_K[\hat\sigma_1, \sigma_0]} d_H(\pi \circ \hat\sigma, \sigma_0) > 0\right\}$$

$$\leq \sum_{u=1}^{n} \mathbb{P}\left\{\exists \pi \in S_K[\hat\sigma_1, \sigma_0], \ \pi(\hat\sigma(u)) \neq \sigma_0(u)\right\}$$

$$\leq n \exp\left(-(1-\zeta_n'')\frac{nI(\{P_l\},\{Q_l\})}{\beta K}\right) + n^{-2} \leq 2n^{-1}.$$

We now let $\zeta_n = \tilde\zeta_n$. Since $\tilde\zeta_n \geq \zeta_n''$, we can conclude that

$$\mathbb{P}\left\{l(\hat\sigma, \sigma_0) > \exp\left(-(1-\zeta_n)\frac{n}{\beta K}I(\{P_l\},\{Q_l\})\right)\right\} \to 0.$$

B Supporting results for Proposition 6.1

We now provide proofs for the supporting results stated in Appendix A.

B.1 Analysis of estimation error of $\hat P_l$ and $\hat Q_l$

We begin with a proposition.

Proposition B.1. *Let $\sigma_0 \in \mathcal{C}(\beta, K)$. Let $L \in \mathbb{Z}^+$, let $(\{P_l\}, \{Q_l\}) \in \mathcal{P}_L^2$, and let $A \in \{0, \dots L\}^{n \times n}$ be a random labeled network with the distribution $LSBM(\sigma_0, (\{P_l\}, \{Q_l\}))$. Define $\Delta_l := |P_l - Q_l|$. For a clustering σ, define*

$$S(\sigma) := \{(u, v) \in [n]^2 \ : \ u \neq v, \ \sigma(u) = \sigma(v)\},$$

$$S(\sigma)^c := \{(u, v) \in [n]^2 \ : \ u \neq v, \ \sigma(u) \neq \sigma(v)\},$$

and also define the estimators

$$\hat P_l := \frac{1}{|S(\sigma)|} \sum_{(u,v) \in S(\sigma)} \mathbf{1}\{A_{uv} = l\}, \qquad \hat Q_l := \frac{1}{|S(\sigma)^c|} \sum_{(u,v) \in S(\sigma)^c} \mathbf{1}\{A_{uv} = l\}. \tag{17}$$

Let $\gamma \in [0, 1]$ and let $\eta := 8(\sqrt{\eta'} + \eta')$, where $\eta' := \beta K \gamma \log \frac{eK}{\gamma} + 6\frac{\log n}{n}$.

Then with probability at least $1 - 4(L+1)n^{-6}$, it holds that for any $\sigma : [n] \to [K]$ such that $l(\sigma, \sigma_0) \leq \gamma$, we have

$$\max\{|\hat P_l - P_l|, |\hat Q_l - Q_l|\} \leq \Delta_l + \eta\left(\frac{P_l \vee Q_l}{n}\right)^{1/2}.$$

Furthermore, if $\gamma \leq \frac{1}{4\beta K}$, then

1. *for all $l \in \{0, \dots, L\}$ where $P_l \vee Q_l \geq \frac{1}{n}$ and $\frac{n\Delta_l^2}{P_l \vee Q_l} \geq 1$, we have*

$$\max\{|\hat P_l - P_l|, |\hat Q_l - Q_l|\} \leq \eta\Delta_l,$$

2. and for all $l \in \{0, \ldots, L\}$ where $P_l \vee Q_l \geq \frac{1}{n}$ and $\frac{n\Delta_l^2}{P_l \vee Q_l} \leq 1$, we have

$$\max\{|\hat{P}_l - P_l|, |\hat{Q}_l - Q_l|\} \leq \eta \left(\frac{P_l \vee Q_l}{n} \right)^{1/2}.$$

Proof. We first fix a clustering $\sigma : [n] \to [K]$ satisfying $l(\sigma, \sigma_0) \leq \gamma$ and fix a color $l \in \{0, \ldots, L\}$. Then

$$\mathbb{E}\hat{P}_l - P_l = \frac{1}{|S(\sigma)|}\left(|S(\sigma) \cap S(\sigma_0)|P_l + |S(\sigma) \cap S(\sigma_0)^c|Q_l\right) - P_l = \frac{|S(\sigma) \cap S(\sigma_0)^c|}{|S(\sigma)|}(Q_l - P_l), \quad (18)$$

and

$$\mathbb{E}\hat{Q}_l - Q_l = \frac{1}{|S(\sigma)^c|}\left(|S(\sigma)^c \cap S(\sigma_0)|P_l + |S(\sigma)^c \cap S(\sigma_0)^c|Q_l\right) - P_l = \frac{|S(\sigma)^c \cap S(\sigma_0)|}{|S(\sigma)^c|}(P_l - Q_l). \quad (19)$$

Hence, we have

$$\max\{|\mathbb{E}\hat{P}_l - P_l|, |\mathbb{E}\hat{Q}_l - Q_l|\} \leq \Delta_l. \quad (20)$$

Now suppose $\gamma \leq \frac{1}{4\beta K}$. Note that for any $u, v \in [n]$ such that $\sigma(u) = \sigma(v)$ but $\sigma_0(u) \neq \sigma_0(v)$, we either have $\sigma(u) \neq \sigma_0(u)$ or $\sigma(v) \neq \sigma_0(v)$. Therefore,

$$|S(\sigma) \cap S(\sigma_0)^c| \leq |\{(u, v) \in [n]^2 \,:\, u \neq v, \sigma_0(u) \neq \sigma(u) \text{ or } \sigma_0(v) \neq \sigma(v)\}| \leq 2n^2\gamma.$$

By a symmetric argument, it follows that $|S(\sigma)^c \cap S(\sigma_0)| \leq 2n^2\gamma$, as well. Define $\hat{n}_k = |\{u \in [n] \,:\, \sigma(u) = k\}|$. Then

$$|S(\sigma)| = \sum_{k=1}^{K} |\{(u, v) \in [n] \,:\, u \neq v, \sigma(u) = \sigma(v) = k\}| = \sum_{k=1}^{K} \hat{n}_k^2 \geq \frac{n^2}{K}, \quad (21)$$

where the inequality holds by observing that $n = \sum_{k=1}^{K} n_k$ and applying the Cauchy-Schwarz inequality. We also have

$$|S(\sigma)^c| = \sum_{k=1}^{K} \hat{n}_k \sum_{k' \neq k} \hat{n}_{k'} \overset{(a)}{\geq} \sum_{k=1}^{K} \hat{n}_k(K-1)\frac{n}{2\beta K} \geq n^2 \frac{(K-1)}{2\beta K} \geq \frac{n^2}{4\beta}, \quad (22)$$

where (a) follows because $l(\sigma, \sigma_0) \leq \gamma \leq \frac{1}{4\beta K}$, so $\min_{k \in [K]} |\{u \in [n] \,:\, \sigma(u) = k\}| \geq \frac{n}{2\beta K}$.

Combining these bounds with equations (18) and (19), we conclude that if $\gamma \leq \frac{1}{4\beta K}$, we have

$$\mathbb{E}\hat{P}_l - P_l \leq K\gamma\Delta_l \quad \text{and} \quad \mathbb{E}\hat{Q}_l - Q_l \leq 4\beta\gamma\Delta_l. \quad (23)$$

We now bound the variance. We use the shorthand

$$T_{1,P} := (P_l \vee Q_l)|S(\sigma)|, \quad T_2 := \gamma n \log \frac{eK}{\gamma} + 6 \log n, \quad t_P := \sqrt{7/3}\{T_{1,P}T_2 \vee T_2^2\}^{1/2},$$

$$T_{1,Q} = (P_l \vee Q_l)|S(\sigma)^c|, \quad \text{and} \quad t_Q := \sqrt{7/3}\{T_{1,Q}T_2 \vee T_2^2\}^{1/2}.$$

We let $E_P(\sigma, l)$ be the event that $|\hat{P}_l - \mathbb{E}\hat{P}_l| \leq \frac{t_P}{\frac{1}{2}|S(\sigma)|}$ and let $E_Q(\sigma, l)$ be the event that $|\hat{Q}_l - \mathbb{E}\hat{Q}_l| \leq \frac{t_Q}{\frac{1}{2}|S(\sigma)^c|}$. For convenience, we also denote $\tilde{A}_{uv} := \mathbf{1}\{A_{ij} = l\}$. By Bernstein's inequality, we

have

$$P\big(E_P(\sigma,l)^c\big) \le P\Bigg(\bigg|\sum_{\substack{(u,v)\,:\,u<v,\\\sigma(u)=\sigma(v)}}(\tilde{A}_{uv}-\mathbb{E}\tilde{A}_{uv})\bigg| > t_P\Bigg) \le 2\exp\Bigg(-\frac{\frac{1}{2}t_P^2}{\sum_{\substack{(u,v)\,:\,u<v,\\\sigma(u)=\sigma(v)}}\mathbb{E}\tilde{A}_{uv}+\frac{1}{3}t_P}\Bigg)$$

$$\le \exp\bigg(-\frac{\frac{1}{2}t_P^2}{\frac{1}{2}(P_l\vee Q_l)|S(\sigma)|+\frac{2}{3}t_P}\bigg) = \exp\bigg(-\frac{t_P^2}{T_{1,P}+\frac{1}{3}t_P}\bigg). \quad (24)$$

We consider two cases:

1. Suppose $T_{1,P} \ge T_2$. Then $t_P^2 = (7/3)T_{1,P}T_2$ and

$$2\exp\bigg(-\frac{t_P^2}{T_{1,P}+\frac{2}{3}t_P}\bigg) \le 2\exp\bigg(-\frac{(7/3)T_{1,P}T_2}{T_{1,P}+\frac{4}{3}T_{1,P}}\bigg) \le 2\exp(-T_2).$$

2. Suppose $T_{1,P} \le T_2$. Then $t_P^2 = (7/3)T_2^2$, and the probability term is at most

$$2\exp\bigg(-\frac{4T_2^2}{A+\frac{4}{3}T_2}\bigg) \le 2\exp\bigg(-\frac{(7/3)T_2^2}{T_2+\frac{4}{3}T_2}\bigg) \le 2\exp(-T_2).$$

Combining the above with inequality (24), we have $P(E_P(\sigma,l)) \ge 1-2\exp(-T_2)$. By an identical argument, we can show that $P(E_Q(\sigma,l)) \ge 1-2\exp(-T_2)$, as well. Now note that

$$\frac{t_P}{\frac{1}{2}|S(\sigma)|} \le 2\sqrt{\frac{7}{3}}\frac{(T_1 T_2)^{1/2}}{|S(\sigma)|} + 2\sqrt{\frac{7}{3}}\frac{T_2}{|S(\sigma)|}$$

$$\le 4\bigg(\frac{(P_l\vee Q_l)T_2}{|S(\sigma)|}\bigg)^{1/2} + 4\frac{T_2}{|S(\sigma)|}$$

$$\overset{(a)}{\le} 4\bigg(\frac{K(P_l\vee Q_l)}{n}\frac{T_2}{n}\bigg)^{1/2} + 4\frac{K}{n}\frac{T_2}{n}$$

$$\overset{(b)}{\le} 4\bigg(K\frac{T_2}{n}\bigg)^{1/2}\bigg(\frac{P_l\vee Q_l}{n}\bigg)^{1/2} + 4K\frac{T_2}{n}\bigg(\frac{P_l\vee Q_l}{n}\bigg)^{1/2},$$

where (a) follows from inequality (21) and (b) holds because $P_l\vee Q_l \ge \frac{1}{n}$. In an identical manner, we can use inequality (22) to show that

$$\frac{t_Q}{2|S(\sigma)^c|} \le 4\bigg(4\beta\frac{T_2}{n}\bigg)^{1/2}\bigg(\frac{P_l\vee Q_l}{n}\bigg)^{1/2} + 16\beta\frac{T_2}{n}\bigg(\frac{P_l\vee Q_l}{n}\bigg)^{1/2}.$$

Let $\eta' := 2\beta K\frac{T_2}{n}$. Using the fact that $4\beta\vee K \le 2\beta K$, we have

$$\max\{|\hat{P}_l - \mathbb{E}\hat{P}_l|, |\hat{Q}_l - \mathbb{E}\hat{Q}_l|\} \le 4(\sqrt{\eta'}+\eta')\bigg(\frac{P_l\vee Q_l}{n}\bigg)^{1/2}, \quad (25)$$

under event $E_P(\sigma,l)\cap E_Q(\sigma,l)$.

We now take a union bound over all colors l and over all clusterings σ satisfying $l(\sigma,\sigma_0)\le\gamma$. There are at most $\binom{n}{\gamma n}K^{\gamma n}$ possible σ's satisfying the error bound. Since

$$\log\bigg(\binom{n}{\gamma n}K^{\gamma n}\bigg) \le \log\bigg(\frac{n^{\gamma n}e^{\gamma n}}{(\gamma n)^{\gamma n}}\frac{1}{\sqrt{2\pi\gamma n}}\bigg) + \gamma n\log K$$

$$\le \log\bigg(\frac{e^{\gamma n}}{\gamma^{\gamma n}}\bigg) - \frac{1}{2}\log 2\pi\gamma n + \gamma n\log K$$

$$\le \gamma n\log\frac{e}{\gamma} + \gamma n\log K = \gamma n\log\frac{eK}{\gamma},$$

we conclude that

$$\mathbb{P}\left(\bigcap_{l \in \{0,\ldots,L} \bigcap_{\sigma \,:\, l(\sigma,\sigma_0) \leq \gamma} (E_P(\sigma,l) \cap E_Q(\sigma,l)) \right) \geq 1 - 4(L+1)\exp(-6\log n)$$

$$\geq 1 - 4(L+1)n^{-6}.$$

Combining inequalities (20), (23), and (25), we conclude that, with probability at least $1 - 4(L+1)n^{-6}$, for all clusterings σ such that $l(\sigma,\sigma_0) \leq \gamma$ and for all $l \in \{0,\ldots,L\}$,

$$\max\{|P_l - \hat{P}_l|,\, |Q_l - \hat{Q}_l|\} \leq \Delta_l + \eta\left(\frac{P_l \vee Q_l}{n}\right)^{1/2}.$$

If also $\gamma \leq \frac{1}{4K\beta}$, then

$$\max\left\{|P_l - \hat{P}_l|,\, |Q_l - \hat{Q}_l|\right\} \leq \eta\Delta_l + \eta\left(\frac{P_l \vee Q_l}{n}\right)^{1/2},$$

where $\eta := (2\beta K\gamma) \vee 4(\sqrt{\eta'} + \eta') \leq 4(\sqrt{\eta'} + \eta')$ and $\eta' = 2\beta K\gamma \log\frac{eK}{\gamma} + 6\frac{\log n}{n}$. The statement of the theorem follows immediately. $\qquad\square$

Proposition B.2. *Let $L, A, (\{\hat{P}_l\}, \{\hat{Q}_l\})$ be defined as in Proposition B.1 and suppose $n \geq 2$. Then with probability at least $1 - (L+1)\exp\left(-\frac{n}{5\beta K}\right)$, we have that, for any clustering σ, and for any $l \in \{0,\ldots,L\}$ such that $P_l \vee Q_l \geq \frac{1}{n}$,*

$$\hat{P}_l \vee \hat{Q}_l \geq \frac{1}{\beta K}(P_l \vee Q_l).$$

Proof. Fix a clustering σ, fix $l \in \{0,\ldots,L\}$, and suppose $P_l \vee Q_l \geq \frac{1}{n}$. Define $\hat{T}_l := \frac{2}{n(n-1)}\sum_{(u,v)\,:\,u<v}(A_l)_{uv}$. Let $S(\sigma)$ and $S(\sigma)^c$ be defined as in the statement of Proposition B.1. Then

$$\hat{P}_l \vee \hat{Q}_l \geq \frac{|S(\sigma)|\hat{P}_l}{n(n-1)} + \frac{|S(\sigma)^c|\hat{Q}_l}{n(n-1)} = \hat{T}_l. \tag{26}$$

Note that $\hat{T}_l$ does not depend on σ. Let $t := \frac{1}{2}\mathbb{E}\hat{T}_l$, and let $E(l)$ be the event that $\hat{T}_l \geq t$. Note that

$$t = \frac{1}{n(n-1)}(|S(\sigma_0)P_l + |S(\sigma_0)^c|Q_l) \stackrel{(a)}{\geq} \frac{1}{K}P_l + \frac{1}{2\beta}Q_l \geq \frac{1}{\beta K}(P_l \vee Q_l). \tag{27}$$

In the above derivations, (a) follows because we can use similar reason as in inequalities (21) and (22) to show that $|S(\sigma_0)| \geq \frac{n^2}{K}$ and $|S(\sigma_0)^c| \geq \frac{n^2}{2\beta}$. The last inequality follows because $2\beta \vee K \leq \beta K$. Thus, under event $E(l)$, for any σ, we have $\hat{P}_l \vee \hat{Q}_l \geq \hat{T}_l \geq \frac{1}{\beta K}(P_l \vee Q_l)$. By Bernstein's inequality, we have

$$\mathbb{P}(E(l)^c) \leq \mathbb{P}\left(\left|\sum_{(u,v)\,:\,u<v}(A_l)_{uv} - \mathbb{E}(A_l)_{uv}\right| \geq t\right)$$

$$\leq \exp\left(\frac{-\frac{1}{2}t^2}{\frac{1}{3}t + \sum_{(u,v)\,:\,u<v}\mathbb{E}(A_l)_{uv}}\right) \leq \exp\left(-\frac{t}{5}\right) \stackrel{(a)}{\leq} \exp\left(-\frac{n}{5\beta K}\right),$$

where (a) follows from inequality (27). A union bound over all colors $l \in \{0,\ldots,L\}$ finishes the proof. $\qquad\square$

B.2 Analysis of spectral clustering

Proposition B.3. *Let $\sigma_0 \in \mathcal{C}(\beta, K)$, let $p, q \in [1/n, 1]$, and let $A \in \{0,1\}^{n \times n}$ be a random matrix with the distribution $SBM(\sigma_0, p, q)$. Let $C_{spec1} := 2^{34}$ and $C_{spec2} := 2^{29}$. Suppose $n \geq 8\beta^2 K^2 \vee 2000$ and*

$$\frac{n}{\beta^2 K^6} \frac{(p-q)^2}{p \vee q} \geq C_{spec1}. \tag{28}$$

Then the output σ of Algorithm 4.4 with parameters $\mu = 4\beta$ and $\tau = 40K\bar{d}$ satisfies

$$l(\sigma, \sigma_0) \leq C_{spec2} \frac{\beta K^4}{n} \frac{p \vee q}{(p-q)^2},$$

with probability at least $1 - n^{-5}$.

Proof. Let E_1 be the event that $\frac{1}{2K} n(p \vee q) \leq \bar{d} \leq 3n(p \vee q)$. By Proposition B.4 and the assumption that $K \leq n^{1/6}$, we have $P(E_1) \geq 1 - \exp\left(-\frac{n}{12K}\right) \geq 1 - \exp\left(-\frac{n^{5/6}}{12}\right)$. Under event E_1, we have $\tau = 40K\bar{d} \geq 20n(p \vee q)$, so we may apply Lemma B.1 to obtain

$$P(E_2 \cap E_1) \geq 1 - \exp\left(-\frac{n^{5/6}}{12}\right) - n^{-6},$$

where E_2 is defined to be the event that

$$\|T_\tau(A) - P\|_2 \leq 2^{11} K \sqrt{n(p \vee q)}.$$

Note that $n \geq 2000$ implies that $\frac{3}{2} \frac{n}{\log n} \geq 6 + \log(2)$.

Now suppose the event $E_1 \cap E_2$ holds. Then

$$D_{total}^2 := \sum_{u=1}^n \|\hat{A}_u - P_u\|^2 \leq \|\hat{A} - P\|_F^2 \leq K\|\hat{A} - P\|_2^2$$

$$\leq K\left(\|T_\tau(A) - P\|_2 + \|\hat{A} - T_\tau(A)\|_2\right)^2$$

$$\overset{(a)}{\leq} 4K\|T_\tau(A) - P\|_2^2 \leq 2^{24} K^3 n(p \vee q). \tag{29}$$

Let $\sigma_0 : [n] \to [K]$ denote the true clustering function, and let $\{\mathcal{Z}_1, \ldots, \mathcal{Z}_K\}$ denote the K unique rows of P, where we use the indexing convention that $P_u = \mathcal{Z}_k$ if and only if $\sigma_0(u) = k$. We also let $\{\mathcal{S}_1, \ldots \mathcal{S}_K\}$ denote the set $\{\hat{A}_u\}_{u \in S}$, where we use the indexing convention that, for any $k \in [K]$, we have $\arg\min_{\mathcal{Z}_{k'} \in \{\mathcal{Z}_1, \ldots, \mathcal{Z}_K\}} \|\mathcal{S}_k - \mathcal{Z}_{k'}\|_2 = \mathcal{Z}_k$. For each node $u \in [n]$, we define

$$\mathcal{S}^*(u) := \arg\min_{\mathcal{S}_k \in \{\mathcal{S}_1, \ldots \mathcal{S}_K\}} \|\hat{A}_u - \mathcal{S}_k\|_2, \quad \text{and} \quad \mathcal{Z}^*(u) := \arg\min_{\mathcal{Z}_k \in \{\mathcal{Z}_1, \ldots \mathcal{Z}_K\}} \|\hat{A}_u - \mathcal{Z}_k\|_2.$$

We also define the shorthand $D_{sep}^2 := 2\frac{n}{\beta K}(p-q)^2$. For a node $u \in [n]$, we call u *valid* if $\|\hat{A}_u - \mathcal{Z}^*(u)\|_2 \leq D_{sep}/8$. We now make several claims that we use in the proof.

Claim 1: For $k, k' \in [K]$ such that $k \neq k'$, we have $\|\mathcal{Z}_k - \mathcal{Z}_{k'}\|_2 \geq D_{sep}$.

Claim 2: For any $k \in [K]$, we have $\left|\left\{u \in \mathcal{C}_k : D(u) < \frac{2D_{total}}{\left(\frac{n}{\mu\beta K^2}\right)^{1/2}}, \|\hat{A}_u - \mathcal{Z}_k\|_2 < \frac{D_{total}}{\left(\frac{n}{\mu\beta K^2}\right)^{1/2}}\right\}\right| \geq |\mathcal{C}_k|(1 - 1/(\mu K))$. Thus, it follows that the $(1 - 1/(\mu K))$-quantile of $\{D(u) : u \in [n]\}$ is less than $\frac{2D_{total}}{\left(\frac{n}{\mu\beta^2 K}\right)^{1/2}}$.

Claim 3: If a node $u \in S$, then u is valid.

Claim 4: For any $u, v \in S$ such that $u \neq v$, we have $\mathcal{Z}^*(u) \neq \mathcal{Z}^*(v)$.

Claim 5: For $k, k' \in [K]$ such that $k \neq k'$, we have $\|\mathcal{S}_k - \mathcal{S}_{k'}\|_2 \geq \frac{3}{4} D_{sep}$.

Claim 6: If u is valid, then $\|\hat{A}_u - \mathcal{S}^*(u)\|_2 \leq D_{sep}/4$.

Claim 1 follows from the SBM definition and the fact that the smallest cluster has at least $\frac{n}{\beta K}$ elements. To derive Claim 2, let $k \in [K]$, define $\mathcal{C}_k := \{u : \sigma_0(u) = k\}$, and let $C' \subset \mathcal{C}_k$ be such that $|C'| = \lfloor |\mathcal{C}_k|/(\mu K) \rfloor$ and $\max_{u \in \mathcal{C}_k \setminus C'} \|\hat{A}_u - \mathcal{Z}_k\|_2 \leq \min_{u \in C'} \|\hat{A}_u - \mathcal{Z}_k\|_2$. Then

$$D_{total}^2 \geq \sum_{v \in C'} \|\hat{A}_v - \mathcal{Z}_k\|_2^2 \geq |C'| \min_{v \in C'} \|\hat{A}_v - \mathcal{Z}_k\|_2^2 \overset{(a)}{>} \left(\frac{n}{\mu \beta K^2} - 1\right) \max_{v \in \mathcal{C}_k \setminus C'} \|\hat{A}_v - \mathcal{Z}_k\|_2^2,$$

where (a) follows because $|\mathcal{C}_k| \geq \frac{n}{\beta K}$. Thus, we have shown that for any $v \in \mathcal{C}_k \setminus C'$, we have $\|\hat{A}_v - \mathcal{Z}_k\|_2 < \frac{D_{total}}{\left(\frac{n}{\mu \beta K^2} - 1\right)^{1/2}}$. Since $|\mathcal{C}_k \setminus C'| = \mathcal{C}_k(1 - 1/(\mu K)) \geq \frac{n}{2\beta K} \geq \lceil \frac{n}{\mu K} \rceil$, we have that, for any $v \in \mathcal{C}_k \setminus C'$,

$$D(v) \leq \max_{v' \in \mathcal{C}_k \setminus C'} \|\hat{A}_v - \hat{A}_{v'}\|_2 \leq \max_{v' \in \mathcal{C}_k \setminus C'} \left(\|\hat{A}_v - \mathcal{Z}_k\|_2 + \|\hat{A}_{v'} - \mathcal{Z}_k\|_2\right) < \frac{2 D_{total}}{\left(\frac{n}{\mu \beta K^2} - 1\right)^{1/2}}.$$

Claim 2 follows by again noting that $|\mathcal{C}_k \setminus C'| \geq |\mathcal{C}_k|(1 - 1/(\mu K))$.

To argue Claim 3, let u be an arbitrary invalid node and let $N(u) = \{v \in [n] : \|\hat{A}_u - \hat{A}_v\| \leq D(u)\}$; it follows from the definition of $D(u)$ that $|N(u)| \geq \lceil \frac{n}{\mu K} \rceil$. Then

$$D_{total}^2 \geq \sum_{v \in N(u)} \|\hat{A}_v - \mathcal{Z}^*(v)\|_2^2 \geq \sum_{v \in N(u)} \max\{\|\hat{A}_u - \mathcal{Z}^*(v)\|_2 - \|\hat{A}_v - \hat{A}_u\|_2, 0\}^2$$

$$\geq \frac{n}{\mu K} \max\{D_{sep}/8 - D(u), 0\}^2.$$

Therefore,

$$D(u) \geq \frac{D_{sep}}{8} - \frac{D_{total}}{\left(\frac{n}{\mu K}\right)^{1/2}} \overset{(a)}{\geq} \frac{D_{sep}}{16} \overset{(b)}{\geq} \frac{2 D_{total}}{\left(\frac{n}{\mu \beta K^2} - 1\right)^{1/2}} \overset{(c)}{>} (1 - 1/(\mu K))\text{-quantile}\{D(u) : u \in [n]\},$$

where (a) and (b) follows from hypothesis (28) of the proposition, and (c) follows from Claim 2. Claim 3 can then be shown from the definition of the set S.

We argue Claim 4 by induction. Let $\{u_1, \ldots u_K\} = S$, where u_1 is the first node added to S, u_2 is the second, etc. Let r be a positive integer such that $2 \leq r < K$, and suppose $\mathcal{Z}^*(u_s) \neq \mathcal{Z}^*(u_{s'})$ for all $s, s' \in [r]$. Since $r < K$, there exists $k \in [K]$ such that $\mathcal{Z}^*(u_s) \neq \mathcal{Z}_k$ for all $s \in [r]$. Since $\frac{n}{\mu K} \leq \frac{n}{2\beta K}$, by Claim 2 and the Pigeonhole Principle, there must exist a node $u' \in \mathcal{C}_k$ such that $\|\hat{A}_{u'} - \mathcal{Z}_k\|_2 \leq \frac{D_{total}}{\left(\frac{n}{\mu \beta K^2} - 1\right)^{1/2}} \leq D_{sep}/8$, where the latter inequality is true by hypothesis (28) of the proposition and the fact that $D(u') \leq (1 - 1/(\mu K))\text{-quantile}\{D(u) : u \in [n]\}$. Let $u_{r+1} \in S$. Then by the definition of S, for any $s \in [r]$, we have

$$\|\hat{A}_{u_s} - \hat{A}_{u_{r+1}}\|_2 \geq \|\hat{A}_{u_s} - \hat{A}_{u'}\|_2 \geq \|\mathcal{Z}_k - \mathcal{Z}^*(u_s)\|_2 - \|\hat{A}_{u_s} - \mathcal{Z}^*(u_s)\|_2 - \|\hat{A}_{u'} - \mathcal{Z}_k\|_2 \overset{(a)}{\geq} \frac{3}{4} D_{sep},$$

where (a) follows by Claim 1 and Claim 3. At the same time, let v be a valid node such that $\mathcal{Z}^*(v) = \mathcal{Z}^*(u_s)$ for some $s \in [r]$. Then

$$\|\hat{A}_v - \hat{A}_{u_s}\|_2 \leq \|\hat{A}_v - \mathcal{Z}^*(v)\|_2 + \|\hat{A}_{u_s} - \mathcal{Z}^*(u_s)\|_2 \overset{(a)}{\leq} D/4,$$

where (a) follows again by Claim 3. Therefore, by definition of u_{r+1}, it must be that $\mathcal{Z}^*(u_{r+1}) \neq \mathcal{Z}^*(u_s)$ for any $s \in [r]$. Claim 4 follows by induction.

Claim 5 is true because, for any $k \neq k'$, we have

$$\|\mathcal{S}_k - \mathcal{S}_{k'}\|_2 \geq \|\mathcal{Z}_{k'} - \mathcal{Z}_k\|_2 - \|\mathcal{Z}_k - \mathcal{S}_k\|_2 - \|\mathcal{Z}_{k'} - \mathcal{S}_{k'}\|_2 \overset{(a)}{\geq} (3/4)D_{sep},$$

where (a) holds because of Claim 1, Claim 3, Claim 4, and the indexing convention for $\{\mathcal{S}_1, \ldots \mathcal{S}_K\}$.

To see that Claim 6 is true, let u be valid and suppose $\mathcal{Z}^*(u) = \mathcal{Z}_k$ for some $k \in [K]$. Then

$$\|\hat{A}_u - \mathcal{S}^*(u)\|_2 \leq \|\hat{A}_u - \mathcal{S}_k\|_2 \leq \|\hat{A}_u - \mathcal{Z}_k\|_2 + \|\mathcal{Z}_k - \mathcal{S}_k\|_2 \leq \frac{D_{sep}}{4}.$$

We have proved all six claims and now proceed to the proof of the proposition. We say that a node $u \in [n]$ is *incorrect* if u is valid and if $\mathcal{S}^*(u) \neq \mathcal{S}_{\sigma_0(u)}$. Suppose u is incorrect and suppose without the loss of generality that $\sigma_0(u) = k$. Then

$$\begin{aligned}
\|\hat{A}_u - \mathcal{Z}_k\|_2 &\geq \|\hat{A}_u - \mathcal{S}_k\|_2 - \|\mathcal{S}_k - \mathcal{Z}_k\|_2 \\
&\geq \|\mathcal{S}_k - \mathcal{S}^*(u)\|_2 - \|\hat{A}_u - \mathcal{S}^*(u)\|_2 - \|\mathcal{S}_k - \mathcal{Z}_k\|_2 \\
&\overset{(a)}{\geq} \frac{3}{4}D_{sep} - \frac{1}{4}D_{sep} - \frac{1}{8}D_{sep} \geq \frac{3}{8}D_{sep},
\end{aligned}$$

where (a) holds because of Claim 5 and Claim 6. Define a permutation $\tau : [K] \to [K]$ such that for $i, k \in [K]$, we have $\tau(i) = k$ if $S[i] = \mathcal{S}_k$. Then

$$\begin{aligned}
l(\sigma_0, \sigma) \leq d_H(\sigma_0, \tau \circ \sigma) &\leq \frac{|\{u \in [n] : u \text{ is invalid or incorrect}\}|}{n} \\
&\leq \frac{1}{n} \frac{D_{total}^2}{(D_{sep}/8)^2} \leq 2^{29} \frac{\beta K^4}{n} \frac{p \vee q}{(p - q)^2}.
\end{aligned}$$

We finish the proof by noting that if $n \geq 2000$, then $P(E_1 \cap E_2) \geq 1 - e^{-n^{5/6}/12} - n^{-6} \geq n^{-5}$. $\quad\square$

The following Lemma is Lemma 3.3 in [12] and also as Lemma 5 in [**?**]. We transcribe the full statement and proof here to make the paper self-contained.

Lemma B.1. *Let $P \in [0,1]^{n \times n}$ be a symmetric matrix, let $p_{max} := \max_{u \geq v} P_{uv}$, and suppose $np_{\max} \geq 1$. Let A be an adjacency matrix such that $A_{uu} = 0$ and $A_{uv} \sim Ber(P_{uv})$ for $u < v$. Let $\tilde{\tau} \geq 20$ and $\tau := \tilde{\tau} n p_{\max}$. Then for any $C' > 0$ and any n such that $\frac{3}{2} \frac{n}{\log n} \geq C' + \log 2$, we have*

$$\|T_\tau(A) - P\|_2 \leq \frac{4}{3}(461 + 160C' + 16\tilde{\tau})\sqrt{np_{\max}},$$

with probability at least $1 - n^{-C'}$.

Proof. Let $S := \{u \in [n] : d_u < \tau\}$ and let $A_{SS} := T_\tau(A)$. Let P_{SS} be the result of setting row/column u of P to zero for every $u \in S^c$. Observe then that

$$\|T_\tau(A) - P\|_2 \leq \|P_{SS} - P\|_2 + \|A_{SS} - P_{SS}\|_2.$$

We first bound $\|P_{SS} - P\|_2$. By Proposition B.5, with probability at least $1 - 2\exp(-n/2)$, we have

$$\begin{aligned}
\|P_{SS} - P\|_2 &\leq \|P_{SS} - P\|_F \leq \big|\{(u,v) : u \in S^c \text{ or } v \in S^c\}\big|^{1/2} p_{\max} \\
&\leq 2^{1/2} \frac{n}{\tau^{1/2}} p_{\max} \overset{(a)}{\leq} \sqrt{np_{\max}},
\end{aligned}$$

where (a) follows because $\tilde{\tau} \geq 20$. Define E_1 to be the event that $\|P_{SS} - P\|_2 \leq \sqrt{np_{\max}}$.

Also define $M := A_{SS} - P_{SS}$. Let $\mathcal{C}$ be the minimal $1/8$-covering of S^{n-1}; it follows that $|\mathcal{C}| \leq 64^n$. For any $x, y \in S^{n-1}$, let $\tilde{x}, \tilde{y} \in \mathcal{C}$ be such that $\|x - \tilde{x}\|, \|y - \tilde{y}\| \leq 1/8$. Then

$$x^\top M y = \tilde{x}^\top M \tilde{y} + (x - \tilde{x})^\top M y + \tilde{x}^\top M(y - \tilde{y}) \leq \tilde{x}^\top M \tilde{y} + \frac{1}{4}\|M\|_2.$$

Taking the supremum over $x, y \in S^{n-1}$, we have

$$\|M\|_2 = \sup_{x,y \in S^{n-1}} x^\top M y \leq \frac{4}{3} \sup_{\tilde{x}, \tilde{y} \in \mathcal{C}} \tilde{x}^\top M \tilde{y}. \tag{30}$$

For any $x, y \in \mathcal{C}$, define $H(x,y) := \{(u,v) \,:\, |x_u y_v| \geq \sqrt{\frac{p_{\max}}{n}}\}$ and $L(x,y) := \{(u,v) \,:\, |x_u y_v| < \sqrt{\frac{p_{\max}}{n}}\}$. It follows that for any $x, y \in \mathcal{C}$, we have

$$x^\top M y = \sum_{(u,v) \in H(x,y)} x_u y_v M_{uv} + \sum_{(u,v) \in L(x,y)} x_u y_v M_{uv}. \tag{31}$$

Define E_2 to be the event in which $\max_{x,y \in \mathcal{C}} \sum_{(u,v) \in L(x,y)} x_u y_v M_{uv} \leq 12\sqrt{p_{\max} n}$. By Proposition B.6 and a union bound over all $x, y \in \mathcal{C}$, we have $P(E_2) \geq 1 - e^{-n}$. For any $x, y \in \mathcal{C}$, $\sum_{(u,v) \in H(x,y)} x_u y_v \leq \sum_{(u,v) \in H(x,y)} x_u^2 y_v^2 \sqrt{\frac{n}{p_{\max}}} \leq \sqrt{n p_{\max}}$. Thus,

$$\sum_{(u,v) \in H(x,y)} x_u y_v (P_{SS})_{uv} \leq \sqrt{n p_{\max}}. \tag{32}$$

For any $S', T' \subseteq [n]$, define $e(S', T'|A) := \sum_{(u,v)\,:\,(u,v) \in (S' \times T') \cup (T' \times S')} A_{uv}$. Let E_3 be the event that, for any $S', T' \subseteq [n]$, either

$$\frac{e(S', T'|A)}{|S'||T'|p_{\max}} \leq 4$$

or

$$\frac{e(S', T'|A)}{|S'||T'|p_{\max}} \log \frac{e(S', T'|A)}{|S'||T'|p_{\max}} \leq 3(4 + 2C') \frac{1}{|S'|} \log \frac{en}{|T'|}.$$

Then by Proposition B.7, we have $P(E_3) \geq 1 - n^{-2C'}$. For any $S', T' \subseteq [n]$, we have $e(S', T'|A_{SS}) \leq e(S', T'|A)$, so in the event E_3, for any $S', T' \subseteq [n]$, we have either $\frac{e(S',T'|A_{SS})}{|S'||T'|p_{\max}} \leq 4$ or $\frac{e(S',T'|A_{SS})}{|S'||T'|p_{\max}} \log \frac{e(S',T'|}{|S'||T'|}$. $3(4 + 2C') \frac{1}{|S'|} \log \frac{en}{|T'|}$. By Proposition B.8 with $d = np_{\max}$, C_2 set to 8, $C_3 = 2(4 + 2C')$, and $\tau = \tilde{\tau} np_{\max}$, it is known that, in the event E_3, we have

$$\max_{x,y \in \mathcal{C}} \sum_{(u,v) \in H(x,y)} x_u y_v (A_{SS})_{uv} \leq (448 + 160C' + 16\tilde{\tau})\sqrt{n p_{\max}}. \tag{33}$$

Under the event $E_1 \cap E_2 \cap E_3$, by combining inequalities (30), 31, 32, 33, and the definition of E_2, we have

$$\|M\|_2 \leq \frac{4}{3}(461 + 160C' + 16\tilde{\tau})\sqrt{n p_{\max}}.$$

To finish the proof, we take a union bound over the events E_1^c, E_2^c, and E_3^c, and observe that $2e^{-n/2} + e^{-n} + n^{-2C'} \leq n^{-C'}$ when $\frac{3}{2}\frac{n}{\log n} \geq C' + \log 2$. $\qquad\square$

Proposition B.4. *Let A, P, and $p_{\max}$ be defined as in Lemma B.1. Let $C_1 > 0$ be defined such that $\sum_{u<v} P_{uv} = C_1 n^2 p_{\max}$. Let $\bar{d} := \frac{1}{n}\sum_{u \neq v} A_{uv}$ be the average degree. Then with probability at least $\exp(-Cn)$ where $C := \frac{C_1/2}{1/3+2}$, we have*

$$C_1 n p_{\max} \leq \bar{d} \leq 3 C_1 n p_{\max}.$$

In the case of $SBM(K, \beta, p, q)$, we have $\frac{1}{2K} \leq C_1 \leq 1$.

Proof. We use the shorthand $t := \frac{1}{2}C_1 n^2 p_{\max}$. By Bernstein's inequality (Proposition B.10), we have

$$\mathbb{P}\big(C_1 n p_{\max} \leq \bar{d} \leq 3C_1 n p_{\max}\big) \leq \mathbb{P}\bigg(\Big|\sum_{(u,v)\,:\,u<v}(A_{uv} - \mathbb{E}A_{uv})\Big| \geq t\bigg)$$

$$\leq \exp\bigg(-\frac{(1/2)t^2}{(1/3)t + C_1 n^2 p_{\max}}\bigg)$$

$$\leq \exp\bigg(-\frac{(1/2)C_1 n}{1/3 + 2}\bigg),$$

where the last inequality follows because $np_{\max} \geq 1$ by assumption. For the second claim, define $N_w := \sum_{k=1}^{K}\frac{n_k(n_k-1)}{2}$ and $N_b := \frac{n(n-1)}{2} - N_w$. Then

$$C_1 = \frac{N_w}{n^2}\frac{p}{p_{\max}} + \frac{N_b}{n^2}\frac{q}{p_{\max}}.$$

The lower bound on C_1 follows from the fact that $N_w \leq n(n/K - 1)$ and the assumption that $K \leq n/2$. The upper bound on C_1 follows from the fact that $N_w + N_b = n(n-1)/2$. $\qquad\square$

B.3   Supporting results for Lemma B.1

The following proposition is Lemma 3.1 in [12] and Lemma 11 in [17]. We transcribe the full statement and the proof here to make the paper self-contained.

Proposition B.5. *Let A, P, and p_{max} be defined as in Lemma B.1. For a node u, let $d_u := \sum_{v \neq u} A_{uv}$ be the degree. Let $\tilde{\tau} \geq 20$ and let $\tau := \tilde{\tau} n p_{\max}$. Then with probability at least $1 - 2\exp(-n/2)$, we have*

$$\big|\{u \in \{1,\ldots,n\}\,:\,d_u \geq \tau\}\big| \leq \frac{n}{\tau}.$$

Proof. Let $S \subseteq [n]$ and define $e(S,V) := \frac{1}{2}\sum_{(u,v)\,:\,u\in S \text{ or } v\in S} A_{uv}$. Since

$$\big|\{(u,v)\,:\,u \in S \text{ or } v \in S\}\big| \leq 2|S|n,$$

we have $\mathbb{E}e(S,V) \leq |S|np_{\max}$. Therefore, by Chernoff's bound (Proposition B.9), we have

$$\mathbb{P}\big(d_u \geq \tau,\,\forall u \in S\big) \leq \mathbb{P}\big(e(S,V) \geq \frac{\tau}{2}|S|\big)$$

$$\leq \mathbb{P}\bigg(e(S,V) \geq \mathbb{E}e(S,V) + \big(\frac{\tilde{\tau}}{2} - 1\big)p_{\max}n|S|\bigg)$$

$$\leq \exp\bigg(-\big(\frac{\tilde{\tau}}{2} - 1\big)p_{\max}n\bigg(\log\big(1 + \frac{(\frac{\tilde{\tau}}{2}-1)p_{\max}n}{\mathbb{E}e(S,V)}\big) - 1\bigg)\bigg)$$

$$\overset{(a)}{\leq} \exp\big(-\big((\frac{\tilde{\tau}}{2} - 1)\log 10\big)p_{\max}n\big) \overset{(b)}{\leq} \exp\big(-\frac{\tilde{\tau}}{4}p_{\max}n|S|\big). \qquad (34)$$

In the above, (a) follows because $\frac{(\tilde{\tau}/2-1)p_{\max}n}{\mathbb{E}e_1(S,V)} \geq 9$ and $\log(x) - 1 \geq \log(x)/2$ for all $x \geq 10$, and (b) follows because $(\tilde{\tau}/2 - 1)\log 10 \geq \tilde{\tau}$ for all $\tilde{\tau} \geq 20$.

Taking a union bound over all subsets of size greater than $\frac{n}{\tau}$, we have

$$
\begin{aligned}
\mathbb{P}\big(\exists S, \, |S| > \frac{n}{\tau}, \, d_u \geq \tau, \, \forall u \in S\big) &\leq \sum_{l=\lceil n/\tau \rceil}^{n} \sum_{S \subseteq V, |S|=l} \exp\big(-\tilde{\tau} p_{\max} n l\big) \\
&\overset{(a)}{\leq} \sum_{l=\lceil n/\tau \rceil}^{n} \exp\big(-\tilde{\tau} p_{\max} n l + l \log \frac{en}{l}\big) \\
&\overset{(b)}{\leq} \sum_{l=\lceil n/\tau \rceil}^{n} \exp\bigg(-l\big(\tilde{\tau} p_{\max} n - \log(e\tilde{\tau} p_{\max} n)\big)\bigg) \\
&\overset{(c)}{\leq} \sum_{l=\lceil n/\tau \rceil}^{n} \exp\big(-\frac{\tilde{\tau}}{2} p_{\max} n l\big) \overset{(d)}{\leq} 2\exp(-n/2).
\end{aligned}
$$

In the above derivations, (a) follows from the Stirling approximation; (b) follows because $\frac{n}{l} > \tau$; (c) follows because $\tilde{\tau} - \log(e\tilde{\tau}) \geq \tilde{\tau}/2$ for all $\tilde{\tau} \geq 20$ and $p_{\max} n \geq 1$; (d) follows because $\sum_{l=\lceil n/\tau \rceil}^{n} \exp\big(-\frac{\tilde{\tau}}{2} p_{\max} n l\big)$ is a geometric series the largest term of which is bounded above by $\exp(-n/2)$, and the ratio is $\exp(-\frac{\tilde{\tau}}{2} p_{\max} n) \leq e^{-10} \leq 1/2$. $\qquad\square$

Proposition B.6. *Let $A, P,$ and $p_{\max}$ be defined as in Lemma B.1. Let $x, y \in S^{n-1}$ and let $L := \{(u,v) : |x_u y_v| \leq \sqrt{\frac{p_{\max}}{n}}\}$. Then with probability at least $\exp(-8n)$,*

$$
\sup_{S \subseteq [n]} \sum_{(u,v) \in L} x_u y_v (A_{SS} - P_{SS})_{uv} \leq 12\sqrt{p_{\max} n}.
$$

Proof. First fix $S \subseteq [n]$. Since $\max_{(u,v) \in L} |x_u y_v (A_{SS} - P_{SS})_{uv}| \leq \sqrt{\frac{p_{\max}}{n}}$ and $\sum_{(u,v) \in L} |x_u y_v|^2 \mathbb{E}((A_{SS} - P_{SS})_{uv}^2) \leq p_{\max}$, we may apply Bernstein's inequality (Proposition B.10) to obtain $\sum_{(u,v) \in L} x_u y_v (A_{SS} - P_{SS})_{uv} \geq 12\sqrt{p_{\max} n}$, with probability at most $\exp(-9n)$. We now take a union bound to obtain

$$
\mathbb{P}\bigg(\max_{S \subseteq [n]} \sum_{(u,v) \in L} x_u y_v (A_{SS} - P_{SS})_{uv} \geq 12\sqrt{p_{\max} n}\bigg) \leq \exp\big(-9n + n\log 2\big) \leq \exp(-8n).
$$

$\qquad\square$

The following Lemma is Lemma A.3 in [12]. We transcribe the full statement and proof here to make the paper self-contained.

Proposition B.7. *Let $A, p_{\max}$ be defined as in Lemma B.1. Let $d := np_{\max}$. Let $C' > 0$. For any $S, T \subseteq [n]$, define $e(S,T) := \frac{1}{2} \sum_{(u,v) \,:\, (u,v) \in (S \times T) \cup (T \times S)} A_{uv}$. Then with probability at least $1 - n^{-C'}$, for any $S, T \subseteq [n]$, we have*

$$
\frac{e(S,T)}{|S||T|\frac{d}{n}} \leq 8, \qquad or \qquad \frac{e(S,T)}{|S||T|\frac{d}{n}} \log \frac{e(S,T)}{|S||T|\frac{d}{n}} \leq 2(4 + C')\frac{1}{|S|} \log \frac{en}{|T|}.
$$

Proof. Fix $C' > 0$. Let us first fix $S, T \subseteq [n]$ and, without loss of generality, suppose that $|T| \geq |S|$. Since $|\{(u,v) : u \in S, v \in T \text{ or } v \in S, u \in T\}| \leq 2|S||T|$, we have $\mathbb{E}e(S,T) \leq p_{\max}|S||T|$. By a Chernoff bound (Proposition B.9), for $l \geq 8$, we have $\frac{e(S,T)}{|S||T|p_{\max}} \leq l$ with probability at least $1 - \exp\big(-\frac{1}{2}|S||T|p_{\max} l \log l\big)$. Let $l' > 0$ be a positive real number such that $l' \log l' = \frac{2(4+C')|T|}{|S||T|p_{\max}} \log \frac{en}{|T|}$. Let $l = 8 \vee l'$. It is clear that $l \log l \geq \frac{2(4+C')|T|}{|S||T|p_{\max}} \log \frac{en}{|T|}$. Taking a union bound over all subsets, we

35

obtain

$$\mathbb{P}\left(\exists S, T \subseteq V, e(S,T) \geq l p_{\max} |S||T|\right) \leq \sum_{(s,t)\in[n]\times[n]\,:\,t\geq s} \exp\left(-(4+C')t\log\frac{en}{t}\right)\binom{n}{s}\binom{n}{t}$$

$$\leq \sum_{(s,t)\in[n]\times[n],\,t\geq s} \exp\left(-(2+C')t\log\frac{en}{t}\right)$$

$$\overset{(a)}{\leq} \sum_{(s,t)\in[n]\times[n],\,t\geq s} n^{-(2+C')} \leq n^{-C'},$$

where (a) follows because $t\log\frac{en}{t} \geq \log en$ for $1 \leq t \leq n$. The proposition follows. $\qquad\square$

The following Lemma is Lemma A.2 in [12]. We transcribe the full statement and proof here to make the paper self-contained.

Proposition B.8. *Let $\tau > 0$. Let A be a graph on n nodes such that the maximum degree is τ. For any $S, T \subseteq [n]$, let $e(S,T)$ be defined as in Proposition B.7. Let $d > 0$ be a positive number such that, for any $S, T \subset [n]$, we have either*

$$\frac{e(S,T)}{|S||T|\frac{d}{n}} \leq C_2, \qquad or \qquad \frac{e(S,T)}{|S||T|\frac{d}{n}}\log\frac{e(S,T)}{|S||T|\frac{d}{n}} \leq C_3 \frac{1}{|S|}\log\frac{en}{|T|},$$

where $C_2 \geq e$ and $C_3 > 0$ are positive universal constants. Let $x, y \in S^{n-1}$, and let $H = \{(u,v) : |x_u y_v| \geq \sqrt{d}/n\}$. Then

$$\sum_{(u,v)\in H} x_u y_v A_{uv} \leq \left(16 C_2 + 40 C_3 + 8 + 16\frac{\tau}{d}\right)\sqrt{d}.$$

Proof. Fix $x, y \in S^{n-1}$. Since $A_{uv} \geq 0$, we may assume without loss of generality that $x_u, y_v \geq 0$ for all $u, v \in [n]$. For $i, j \in \{1, \ldots, \lceil\log_2 \frac{n}{\sqrt{d}}\rceil\}$, let $\gamma_0 = \sqrt{d/n}$, $\gamma_i := 2^i \gamma_0$, and $\gamma_j := 2^j \gamma_0$. We also define

$$S_i := \left\{u : \frac{\gamma_{i-1}}{\sqrt{n}} \leq x_u \leq \frac{\gamma_i}{\sqrt{n}}\right\}, \qquad T_j := \left\{v : \frac{\gamma_{j-1}}{\sqrt{n}} \leq y_v \leq \frac{\gamma_j}{\sqrt{n}}\right\}.$$

Define the shorthand $s_i := |S_i|$ and $t_j := |T_j|$. Note that since $|x_u y_v| \geq \frac{\sqrt{d}}{n}$ and $x_u, y_v \leq 1$ for all $(u,v) \in H$, we have $H \subseteq \bigcup_{(i,j)\,:\,\gamma_i\gamma_j\geq\sqrt{d}} S_i \times T_j$. Hence,

$$\sum_{(u,v)\in H} x_u y_v A_{uv} \leq \sum_{\substack{(i,j):\\ \gamma_i\gamma_j\geq\sqrt{d}}} \sum_{u\in S_i,\,v\in T_j} x_u y_v A_{uv} \leq \sum_{\substack{(i,j):\\ \gamma_i\gamma_j\geq\sqrt{d}}} \frac{\gamma_i\gamma_j}{n}\frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} s_i t_j \frac{d}{n}. \tag{35}$$

Also note that since $x, y \in S^{n-1}$, we have $\sum_i \frac{\gamma_i^2}{n}s_i \leq 4$ and $\sum_j \frac{\gamma_j^2}{n}t_j \leq 4$. Define the shorthand $H^* := \{(i,j) : \gamma_i\gamma_j \geq \sqrt{d}\}$ and define $H_1 := \{(i,j) \in H^* : \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \leq C_2 \frac{\gamma_i\gamma_j}{\sqrt{d}}\}$. Then

$$\sum_{(i,j)\in H_1} \frac{\gamma_i\gamma_j}{n}\frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} s_i t_j \frac{d}{n} \leq C_2\sqrt{d} \sum_{(i,j)\in H_1} \frac{\gamma_i^2}{n}\frac{\gamma_j^2}{n}s_i t_j \leq 16 C_2\sqrt{d}. \tag{36}$$

Now define $H_2 := \{(i,j) \in H^* : \frac{\gamma_i}{\gamma_j} \geq \sqrt{d}\}$. Note that because the maximum degree is bounded by

36

τ, we have $\frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \leq \frac{s_i \tau}{s_i t_j \frac{d}{n}} \leq \frac{\tau}{d} \frac{n}{t_j}$. Thus,

$$\sum_{(i,j) \in H_2} \frac{\gamma_i \gamma_j}{n} \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} s_i t_j \frac{d}{n} \leq \frac{\tau}{d} \sum_{(i,j) \in H_2} \frac{\gamma_i \gamma_j}{n} s_i d$$

$$\leq \frac{\tau}{d} \sum_i \frac{\gamma_i^2}{n} s_i d \sum_{j\,:\,(i,j) \in H_2} \frac{\gamma_j}{\gamma_i}$$

$$\overset{(a)}{\leq} 2\frac{\tau}{d}\sqrt{d} \sum_i \frac{\gamma_i^2}{n} s_i \leq 8\frac{\tau}{d}\sqrt{d}, \tag{37}$$

where in the above derivations, (a) follows because, for any fixed i, the sum $\sum_{(i,j) \in H_2} \frac{\gamma_j}{\gamma_i}$ is a geometric series whose term-to-term ratio is 2. Let $j^* := \max\{j : (i,j) \in H_2\}$. Then $\frac{\gamma_{j^*}}{\gamma_i} \leq \frac{1}{\sqrt{d}}$. Therefore, we have $\sum_{(i,j) \in H_2} \frac{\gamma_j}{\gamma_i} \leq \frac{2}{\sqrt{d}}$. Now define $H_2' := \{(i,j) \in H^* : \frac{\gamma_j}{\gamma_i} \geq \sqrt{d}\}$. By similar reasoning as above, we have

$$\sum_{(i,j) \in H_2'} \frac{\gamma_i \gamma_j}{n} \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} s_i t_j \frac{d}{n} \leq 8\frac{\tau}{d}\sqrt{d}. \tag{38}$$

Define $H_3 := \left\{(i,j) \in H^* : \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \geq C_2 \frac{\gamma_i \gamma_j}{\sqrt{d}}, \frac{1}{\sqrt{d}} \leq \frac{\gamma_i}{\gamma_j} \leq \sqrt{d}, \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \geq \left(\frac{n}{t_j}\right)^{1/2}\right\}$. Since $\frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \geq C_2 \frac{\gamma_i \gamma_j}{\sqrt{d}} \geq C_2$ for all $(i,j) \in H_3$, we have $\frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \leq \frac{C_3}{\log \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}}} \frac{n}{d s_i} \log \frac{n}{t_j} \leq 2C_3 \frac{n}{d s_i}$. Thus,

$$\sum_{(i,j) \in H_3} \frac{\gamma_i \gamma_j}{n} \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} s_i t_j \frac{d}{n} \leq 2C_3 \sqrt{d} \sum_{(i,j) \in H_3} \frac{\gamma_j^2}{n} t_j \frac{\gamma_i}{\gamma_j} \frac{1}{\sqrt{d}} \leq 8C_3 \sqrt{d}. \tag{39}$$

Define $H_4 := \left\{(i,j) \in H^* : \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \geq C_2 \frac{\gamma_i \gamma_j}{\sqrt{d}}, \frac{1}{\sqrt{d}} \leq \frac{\gamma_i}{\gamma_j} \leq \sqrt{d}, \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \leq \left(\frac{n}{t_j}\right)^{1/2}, \gamma_j^2 \leq \frac{n}{t_j \gamma_j^2}\right\}$. Note that if $(i,j) \in H_3$, then $\frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \leq \left(\frac{n}{t_j \gamma_j^2}\right)^{1/2} \gamma_j \leq \frac{n}{t_j \gamma_j^2}$. Therefore,

$$\sum_{(i,j) \in H_4} \frac{\gamma_i \gamma_j}{n} \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} s_i t_j \frac{d}{n} \leq \sqrt{d} \sum_{(i,j) \in H_4} \frac{\gamma_i^2}{n} s_i \frac{\sqrt{d}}{\gamma_i \gamma_j} \leq \sqrt{d} \sum_i \frac{\gamma_i^2}{n} s_i \sum_{j\,:\,(i,j) \in H_4} \frac{\sqrt{d}}{\gamma_i \gamma_j} \leq 8\sqrt{d}. \tag{40}$$

The last inequality follows because, for every i, the sum $\sum_{j\,:\,(i,j) \in H_4} \frac{\sqrt{d}}{\gamma_i \gamma_j}$ is a geometric series where the term-to-term ratio is 2 and the largest term is at most 1. Define $H_5 := \left\{(i,j) \in H^* : \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \geq C_2 \frac{\gamma_i \gamma_j}{\sqrt{d}}, \frac{1}{\sqrt{d}} \leq \frac{\gamma_i}{\gamma_j} \leq \sqrt{d}, \gamma_j^2 \geq \frac{n}{t_j \gamma_j^2}\right\}$. For any $(i,j) \in H_5$, we have $4\log \gamma_j \geq \log \frac{n}{t_j}$. Therefore, $\frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} \leq 4C_3 \frac{n}{d s_i} \frac{\log \gamma_j}{\log\left(C_2 \frac{\gamma_i \gamma_j}{\sqrt{d}}\right)}$, and

$$\sum_{(i,j) \in H_5} \frac{\gamma_i \gamma_j}{n} \frac{e(S_i,T_j)}{s_i t_j \frac{d}{n}} s_i t_j \frac{d}{n} \leq 4C_3 \sqrt{d} \sum_{(i,j) \in H_5} \frac{\gamma_j^2}{n} t_j \frac{\gamma_i \gamma_j}{\sqrt{d}} \frac{1}{\gamma_j^2} \frac{\log \gamma_j}{\log\left(C_2 \frac{\gamma_i \gamma_j}{\sqrt{d}}\right)}$$

$$\leq 4C_3 \sqrt{d} \sum_j \frac{\gamma_j^2}{n} t_j \sum_{i\,:\,(i,j) \in H_5} \frac{\gamma_i \gamma_j}{\sqrt{d}} \frac{1}{\gamma_j^2} \frac{\log \gamma_j}{\log\left(C_2 \frac{\gamma_i \gamma_j}{\sqrt{d}}\right)}$$

$$\overset{(a)}{\leq} 32C_3 \sqrt{d}. \tag{41}$$

In the above derivations, (a) follows because for all $(i,j) \in H_5$, we have $\gamma_j^2 \geq \frac{\gamma_i \gamma_j}{\sqrt{d}} \geq 1$. Since $C_2 \geq e$ and $x \mapsto \frac{x}{\log(C_2 x)}$ is non-decreasing for all $x \geq 1$, we have $\frac{\gamma_i \gamma_j}{\sqrt{d}} \frac{1}{\gamma_j^2} \frac{\log \gamma_j}{\log(C_2 \frac{\gamma_i \gamma_j}{\sqrt{d}})} \leq \frac{\gamma_i \gamma_j}{\sqrt{d}} \frac{1}{\gamma_j^2} \frac{\log(C_2 \gamma_j^2)}{\log(C_2 \frac{\gamma_i \gamma_j}{\sqrt{d}})} \leq 1$.

For any j, the quantity $\sum_{i\,:\,(i,j)\in H_5} \frac{\gamma_i \gamma_j}{\sqrt{d}} \frac{1}{\gamma_j^2} \frac{\log \gamma_j}{\log\left(C_2 \frac{\gamma_i \gamma_j}{\sqrt{d}}\right)}$ is a geometric series with largest term bounded by 1.

Since $H^* = H_1 \cup H_2 \cup H_2' \cup H_3 \cup H_4 \cup H_5$, we may combine inequalities (35), 36, 37, 38, 39, 40, and 41 to obtain the bound

$$\sum_{(u,v)\in H} x_u y_v A_{uv} \le (16C_2 + 16\frac{\tau}{d} + 8 + 40C_3)\sqrt{d}.$$

$\square$

Proposition B.9 (Chernoff bound). *For $i = 1, \ldots, n$, let $p_i \in [0,1]$. Let $X_i \sim Ber(p_i)$, and let $p = \frac{1}{n}\sum_{i=1}^n p_i$. Then for any $t > 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^n X_i \ge pn + t\right) \le \exp\left(t - (t + pn)\log\left(1 + \frac{t}{pn}\right)\right).$$

A bound that we frequently use is $\mathbb{P}\left(\sum_{i=1}^n X_i \ge pn + t\right) \le \exp\left(-(t + pn)\left(\log\left(1 + \frac{t}{pn}\right) - 1\right)\right)$.

Proposition B.10 (Bernstein's inequality). *Let $X_1, \ldots, X_n$ be real-valued random variables bounded in absolute value by $M > 0$. Suppose $\mathbb{E}X_i = 0$ for all $i = 1, \ldots, n$. Then for any $t > 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^n X_i \ge t\right) \le \exp\left(-\frac{(1/2)t^2}{(1/3)Mt + \sum_{i=1}^n \mathbb{E}X_i^2}\right).$$

If $t \ge \frac{3}{M}\sum_{i=1}^n \mathbb{E}X_i^2$, then $\mathbb{P}\left(\sum_{i=1}^n X_i \ge t\right) \le \exp\left(-(3/4)\frac{t}{M}\right)$.

B.4 Choosing the label l^*

First, we show that for sufficiently well-separated labels, $\hat{I}_l$ is close to $\frac{(P_l - Q_l)^2}{P_l \vee Q_l}$. If the probabilities are not well-separated, we claim that $\hat{I}_l$ is negligibly small.

Proposition B.11. *Let $\sigma_0 \in \mathcal{C}(\beta, K)$, let $L \in \mathbb{Z}^+$, $(\{P_l\}, \{Q_l\}) \in \mathcal{P}_{L}^2$, and let $A \in \{0, \ldots, L\}^{n \times n}$ have the distribution $LSBM(\sigma_0, (\{P_l\}, \{Q_l\}))$. For $l = 0, \ldots, L$, let $\tilde{A}^l \in \{0,1\}^{n \times n}$ be defined as $\tilde{A}_{ij}^l := \mathbf{1}\{A_{ij} = l\}$. Let σ^l be the output of SPECTRAL CLUSTERING (Algorithm 4.4 with parameters $\mu = 4\beta$ and $\tau = 20\bar{d}$) on $\tilde{A}^l$, and let $\hat{P}_l$ and $\hat{Q}_l$ be estimates of P_l and Q_l constructed from σ^l. Suppose $\frac{n}{\log n} \ge 2^{15} \vee 30\beta K$.*

Let $C_{spec1} > 0$ be the universal constant defined in Proposition B.3. Then the following claims are true with probability at least $1 - 12(L+1)^2 n^{-5}$:

1. *For all labels l satisfying $P_l \vee Q_l \ge \frac{1}{n}$ and $\frac{n\Delta_l^2}{P_l \vee Q_l} \ge C_{spec1}\beta^2 K^6$, we have*

$$\frac{1}{4}\frac{|P_l - Q_l|}{\sqrt{P_l \vee Q_l}} \le \frac{|\hat{P}_l - \hat{Q}_l|}{\sqrt{\hat{P}_l \vee \hat{Q}_l}} \le 2\frac{|P_l - Q_l|}{\sqrt{P_l \vee Q_l}}. \tag{42}$$

2. *For all labels satisfying $P_l \vee Q_l \ge \frac{1}{n}$ and $\frac{n\Delta_l^2}{P_l \vee Q_l} \le C_{spec1}\beta^2 K^6$, we have*

$$\frac{|\hat{P}_l - \hat{Q}_l|}{\sqrt{\hat{P}_l \vee \hat{Q}_l}} \le 16\sqrt{C_{spec1}}\beta^2 K^4 \sqrt{\frac{1}{n}}. \tag{43}$$

3. *Suppose $P_l \vee Q_l \geq \frac{1}{n}$ for all $l \in \{0, \dots, L\}$ and $\frac{nI(\{P_l\}, \{Q_l\})}{L\beta^2 K^6} \geq 2C_{spec1}$. Let $c_{init} := 2^{-7}$. Then, with $l^* := \arg\min_{l \in \{0, \dots, L\}} \frac{|\hat{P}_l - \hat{Q}_l|}{\sqrt{\hat{P}_l \vee \hat{Q}_l}}$, we have*

$$\frac{\Delta_{l^*}^2}{P_{l^*} \vee Q_{l^*}} \geq c_{init} \frac{I(\{P_l\}, \{Q_l\})}{L}. \tag{44}$$

Proof. Let $C_{spec1} = 2^{34}$. Let us fix $l \in \{0, \dots, L\}$. Let us first suppose that $P_l \vee Q_l \geq \frac{1}{n}$ and

$$\frac{n\Delta_l^2}{P_l \vee Q_l} \geq C_{spec1}\beta^2 K^6. \tag{45}$$

Let σ_l be the result of performing SPECTRAL CLUSTERING (Algorithm 4.4) on A_l, and let $\hat{P}_l$ and $\hat{Q}_l$ be the subsequent estimators of P_l and Q_l. Let $E_1(l)$ be the event that $l(\sigma_0, \sigma_l) \leq \gamma$, where $\gamma := \frac{1}{2^{14}\beta K \log(\beta K)}$. Under assumption (45), we can verify that the hypothesis of Proposition B.3 holds and thus apply Proposition B.3 to show that $\mathbb{P}(E_1(l)) \geq 1 - n^{-5}$. Define $E_2(l)$ to be the event that

$$\max\{\hat{P}_l - P_l, \hat{Q}_l - Q_l\} \leq \frac{1}{4}\Delta_l. \tag{46}$$

It is straightforward to show that, under $E_1(l)$ and the hypothesis that $6\frac{\log n}{n} \leq 2^{-12}$, we can bound $\eta \leq \frac{1}{4}$, where η is defined in the statement of Proposition B.1. Therefore, by Proposition B.1 and a union bound, we have $\mathbb{P}(E_1(l) \cap E_2(l)) \geq 1 - 4(L+1)n^{-5}$. Under $E_1(l) \cap E_2(l)$, we have from inequality (46) that

$$|\hat{P}_l - \hat{Q}_l| \leq |\hat{P}_l - P_l| + |P_l - Q_l| + |\hat{Q}_l - Q_l| \leq 2\eta\Delta_l + \Delta_l \leq \frac{3}{2}\Delta,$$

$$|\hat{P}_l - \hat{Q}_l| \geq |P_l - Q_l| - |\hat{Q}_l - Q_l| - |\hat{P}_l - P_l| \geq \Delta_l - 2\eta\Delta_l \geq \frac{1}{2}\Delta_l.$$

Furthermore, we can again apply inequality (46) to obtain

$$\hat{P}_l \vee \hat{Q}_l \leq (P_l \vee Q_l) + \eta\Delta_l \leq (P_l \vee Q_l) + \eta(P_l \vee Q_l) \leq \frac{5}{4}(P_l \vee Q_l),$$

$$\hat{P}_l \vee \hat{Q}_l \geq (P_l \vee Q_l) - \eta\Delta_l \geq \frac{3}{4}(P_l \vee Q_l).$$

We can thus prove inequality (42):

$$\frac{1}{\sqrt{5}} \frac{\Delta_l}{\sqrt{P_l \vee Q_l}} \leq \frac{|\hat{P}_l - \hat{Q}_l|}{\sqrt{\hat{P}_l \vee \hat{Q}_l}} \leq \frac{3}{\sqrt{3}} \frac{\Delta_l}{\sqrt{P_l \vee Q_l}}.$$

Now suppose $P_l \vee Q_l \geq \frac{1}{n}$ and $\frac{n\Delta_l^2}{P_l \vee Q_l} < C_{spec1}\beta^2 K^6$. Let $E_1(l)$ denote the entire probability space, and let $E_2(l)$ be the event that

$$\max\{\hat{P}_l - P_l, \hat{Q}_l - Q_l\} \leq \Delta_l + 4\beta^2 K^2 \left(\frac{P_l \vee Q_l}{n}\right)^{1/2}, \quad \text{and} \quad \hat{P}_l \vee \hat{Q}_l \geq \frac{1}{\beta K}(P_l \vee Q_l). \tag{47}$$

Since $\gamma \leq 1$ we have $\eta \leq 4\beta K^2$, where η is defined in the statement of Proposition B.1. Hence, by Proposition B.1 and Proposition B.2, we have

$$\mathbb{P}(E_2(l)) \geq 1 - 4(L+1)n^{-6} - (L+1)\exp\left(-\frac{n}{5\beta K}\right) \overset{(a)}{\geq} 1 - 8(L+1)n^{-6},$$

39

where (a) follows under the hypothesis that $\frac{n}{\log n} \geq 30\beta K$. Under $E_2(l) = E_1(l) \cap E_2(l)$, we have

$$|\hat{P}_l - \hat{Q}_l| \leq \Delta_l + |\hat{P}_l - P_l| + |\hat{Q}_l - Q_l|$$

$$\leq 3\Delta_l + 8\beta K^2 \left(\frac{P_l \vee Q_l}{n}\right)^{1/2} \leq 16\sqrt{C_{spec1}}\beta K^3 \left(\frac{P_l \vee Q_l}{n}\right)^{1/2}.$$

Therefore,

$$\frac{|\hat{P}_l - \hat{Q}_l|}{\sqrt{\hat{P}_l \vee \hat{Q}_l}} \leq 16\sqrt{C_{spec1}}\beta^2 K^4 \left(\frac{1}{n}\right)^{1/2}.$$

A union bound gives us $\mathbb{P}\left(\bigcap_{l=0}^{L} E_1(l) \cap E_2(l)\right) \geq 1 - 12(L+1)^2 n^{-5}$, which proves the first and the second claims of the proposition. For the third claim (44), suppose $\bigcap_{l=0}^{L} E_1(l) \cap E_2(l)$ holds and let $l' := \arg\max_{l \in \{0,\dots,L\}} \frac{\Delta_{l'}^2}{P_{l'} \vee Q_{l'}}$. Observe that

$$\frac{n\Delta_{l'}^2}{P_{l'} \vee Q_{l'}} \geq \frac{n}{L} \sum_{l=0}^{L} \frac{\Delta_l^2}{P_l \vee Q_l} \geq \frac{n}{L} \sum_{l=0}^{L} (\sqrt{P_l} - \sqrt{Q_l})^2 \overset{(a)}{\geq} \frac{n}{2L} I(\{P_l\}, \{Q_l\}) \overset{(b)}{\geq} C_{spec1}\beta^2 K^6, \qquad (48)$$

where (a) holds by Lemma H.2 and (b) follows from the hypothesis of the proposition. Therefore, from the definition of l^* and from inequality (42), we have

$$\frac{|\hat{P}_{l^*} - \hat{Q}_{l^*}|}{\sqrt{\hat{P}_{l^*} \vee \hat{Q}_{l^*}}} \geq \frac{|\hat{P}_{l'} - \hat{Q}_{l'}|}{\sqrt{\hat{P}_{l'} \vee \hat{Q}_{l'}}} \geq \frac{1}{4} \frac{\Delta_{l'}}{\sqrt{P_{l'} \vee Q_{l'}}} \geq \sqrt{C_{spec1}}\beta K^3 > 16\sqrt{C_{spec1}}\beta^2 K^4 n^{-1/2}. \qquad (49)$$

We may deduce from inequality (43) that $\frac{n\Delta_{l^*}^2}{P_{l^*} \vee Q_{l^*}} \geq C_{spec1}\beta^2 K^6$. Hence, by inequality (42), we have

$$\frac{\Delta_{l^*}}{\sqrt{P_{l^*} \vee Q_{l^*}}} \geq \frac{1}{2} \frac{|\hat{P}_{l^*} - \hat{Q}_{l^*}|}{\sqrt{\hat{P}_{l^*} \vee \hat{Q}_{l^*}}} \geq \frac{1}{8} \frac{\Delta_{l'}}{\sqrt{P_{l'} \vee Q_{l'}}} \geq \frac{1}{8} \left(\frac{I(\{P_l\}, \{Q_l\})}{2L}\right)^{1/2}. \qquad (50)$$

$\square$

B.5 Analysis of error probability for a single node

Proposition B.12. *Let $(\{P_l\}, \{Q_l\}) \in \mathcal{G}_{L,\rho}$ be such that $P_l \vee Q_l \geq \frac{1}{n-1}$ for all $l \in \{0,\dots,L\}$. Let $\sigma_0 \in \mathcal{C}(\beta, K)$, and let A be a random labeled matrix taking values in $\{0,\dots,L\}^{n\times n}$, with the distribution $LSBM(\sigma_0, (\{P_l\}, \{Q_l\}))$. Let $u \in [n]$, and let $\tilde{\sigma}_u$ be the output clustering on $\{1,\dots,n\}\backslash\{u\}$ of INITIALIZATION (Algorithm 4.3). For $l \in \{0,\dots,L\}$, let $\hat{P}_l, \hat{Q}_l$ be computed by equation (17) on $A_{l^*}\backslash\{u\}$ with respect to $\tilde{\sigma}_l$. Suppose $\pi_u \in S_K$ satisfies*

$$l(\sigma_0, \tilde{\sigma}_u) = \frac{1}{n-1} d_H(\sigma_0, \pi_u \circ \tilde{\sigma}_u),$$

where both l and d are taken with respect to the set $\{1, 2, \dots, n\} \backslash \{u\}$. Let $C_{mis} := 2^{15} \log(2^{15})$, let $\gamma_0 := \left(C_{mis}\rho^2 K\beta \log(e\rho^2 K\beta)\right)^{-1}$, let $\gamma \in [0, \gamma_0]$, and suppose that $l(\sigma_0, \tilde{\sigma}_u) \leq \gamma$. Suppose also that $\frac{n}{\log n} \geq 2^{16}\rho^2$ and $\frac{nI(\{P_l\}, \{Q_l\})}{L+1} \geq 1$. Let $\eta' := 2\beta K\gamma \log \frac{eK}{\gamma} + 12\frac{\log n}{n}$ and $\eta := 10(\sqrt{\eta'} + \eta')$.
   Then there exists a universal constant $C > 0$ such that, with probability at least

$$1 - 5(L+1)n^{-6} - (K-1)\exp\left(-(1 - C\beta K\rho\eta)\frac{n}{\beta K} I(\{P_l\}, \{Q_l\})\right),$$

we have
$$\pi_u^{-1}(\sigma_0(u)) = \arg\max_{k \in [K]} \sum_{v\,:\,\tilde{\sigma}_u(v)=k} \sum_{l=0}^{L} \log \frac{\hat{P}_l}{\hat{Q}_l} \mathbf{1}(A_{uv} = l).$$

Proof. Define the shorthand $C_{tmp1} := 4 \cdot 40^2 \rho^2 \beta K$ and $C_{tmp2} := eK$. Since $\gamma_0 \leq \left(C_{mis}\rho^2 \beta K \log(e\rho^2 \beta K)\right)^{-}$ with $C_{mis} = 2^{15} \log(2^{15})$, we have $\gamma \leq \gamma_0 \leq \left(4C_{tmp1} \log(C_{tmp1} + C_{tmp2} + 4)\right)^{-1}$, which implies that

$$4 \cdot 40^2 \rho^2 \beta K \gamma \log \frac{eK}{\gamma} = C_{tmp1} \gamma \log \frac{C_{tmp2}}{\gamma} \leq 1,$$

so $2\beta K \gamma \log \frac{eK}{\gamma} \leq \frac{1}{2} \frac{1}{40^2} \frac{1}{\rho^2}$. Additionally, since $\frac{n}{\log n} \geq 2^{16}\rho^2$ by assumption, we have $\frac{12 \log n}{n} \leq \frac{1}{2} \frac{1}{40^2} \frac{1}{\rho^2}$, as well. Thus,

$$\eta' \leq \frac{1}{40^2}\rho^2 < 1, \text{ and } \eta \leq 20\sqrt{\eta'} \leq \frac{1}{2\rho}. \tag{51}$$

Let us assume without loss of generality that $\sigma_0(u) = 1$, and π_u is the identity. Observe that since $2\beta \geq \frac{n-1}{K} \frac{1}{\frac{n}{\beta K}-1}$, the minimum cluster size of σ_0 with respect to $\{1, \ldots, n\}\setminus\{u\}$ is at least $\frac{n}{2\beta K}$. Define E_1 to be the event that for any $l \in \{0, \ldots, L\}$,

$$\max\{|\hat{P}_l - P_l|, |\hat{Q}_l - Q_l|\} \leq \eta\left(\Delta_l \vee \left(\frac{P_l \vee Q_l}{n}\right)^{1/2}\right). \tag{52}$$

Since $l(\tilde{\sigma}_u, \sigma_0) \leq \gamma \leq \frac{1}{8\beta K}$, we have by Proposition B.1 that

$$P(E_1) \geq 1 - 4(L+1)(n-1)^{-6} \geq 1 - 5(L+1)n^{-6}.$$

For $k \in [K]$, define $E_2(k)$ to be the event

$$\sum_{v \in [n]\setminus\{u\}\,:\,\tilde{\sigma}_u(v)=1} \sum_{l=0}^{L} \log \frac{\hat{P}_l}{\hat{Q}_l} \mathbf{1}\{A_{uv} = l\} > \sum_{v \in [n]\setminus\{u\}\,:\,\tilde{\sigma}_u(v)=k} \sum_{l=0}^{L} \log \frac{\hat{P}_l}{\hat{Q}_l} \mathbf{1}\{A_{uv} = l\}. \tag{53}$$

For any $v \in [n]\setminus\{u\}$, define $\bar{A}_{uv} := \sum_{l=0}^{L} \log \frac{\hat{P}_l}{\hat{Q}_l} \mathbf{1}(A_{uv} = l)$. Note that $\{\bar{A}_{uv}\}_{v\in[n]\setminus\{u\}}$ is a set of independent random variables. If $v \in \sigma_0^{-1}(1)$, then for $l \in \{0, \ldots, L\}$, we have $\bar{A}_{uv} = \log \frac{\hat{P}_l}{\hat{Q}_l}$ with probability P_l. On the other hand, if $v \notin \sigma_0^{-1}(1)$, then for $l \in \{0, \ldots, L\}$, $\bar{A}_{uv} = \log \frac{\hat{P}_l}{\hat{Q}_l}$ with probability Q_l. Now define the shorthand

$$V_1 := \{v \in [n]\setminus\{u\} : \sigma_0(v) = 1, \tilde{\sigma}_u(v) = 1\}, \quad V_1' := \{v \in [n]\setminus\{u\} : \sigma_0(v) \neq 1, \tilde{\sigma}_u(v) = 1\}, \tag{54}$$
$$V_k := \{v \in [n]\setminus\{u\} : \sigma_0(v) = 1, \tilde{\sigma}_u(v) = k\}, \quad V_k' := \{v \in [n]\setminus\{u\} : \sigma_0(v) \neq 1, \tilde{\sigma}_u(v) = k\}. \tag{55}$$

41

Then for any $t \geq 0$, we have

$$\mathbb{P}(E_1^c \cup E_2(k)^c) - \mathbb{P}(E_1^c)$$

$$\leq \mathbb{P}\bigg(\sum_{v \in V_k} \bar{A}_{uv} + \sum_{v \in V_k'} \bar{A}_{uv} - \sum_{v \in V_1} \bar{A}_{uv} - \sum_{v \in V_1'} \bar{A}_{uv} \geq 0 \bigg)$$

$$\leq \mathbb{P}\bigg\{ \exp\bigg(t\bigg(\sum_{v \in V_k} \bar{A}_{uv} + \sum_{v \in V_k'} \bar{A}_{uv} - \sum_{v \in V_1} \bar{A}_{uv} - \sum_{v \in V_1'} \bar{A}_{uv} \bigg) \bigg) \geq 1 \bigg\}$$

$$\leq \mathbb{E}\bigg[\exp\bigg(t\bigg(\sum_{v \in V_k} \bar{A}_{uv} + \sum_{v \in V_k'} \bar{A}_{uv} - \sum_{v \in V_1} \bar{A}_{uv} - \sum_{v \in V_1'} \bar{A}_{uv} \bigg) \bigg) \bigg]$$

$$\leq \bigg(\sum_{l=0}^{L} e^{t \log \frac{\hat{P}_l}{\hat{Q}_l}} P_l \bigg)^{|V_k|} \bigg(\sum_{l=0}^{L} e^{t \log \frac{\hat{P}_l}{\hat{Q}_l}} Q_l \bigg)^{|V_k'|}$$

$$\bigg(\sum_{l=0}^{L} e^{-t \log \frac{\hat{P}_l}{\hat{Q}_l}} P_l \bigg)^{|V_1|} \bigg(\sum_{l=0}^{L} e^{-t \log \frac{\hat{P}_l}{\hat{Q}_l}} Q_l \bigg)^{|V_1'|}.$$

Setting $t = 1/2$, we have

$$\mathbb{P}(E_1^c \cup E_2(k)^c) - \mathbb{P}(E_1^c) \tag{56}$$

$$\leq \bigg(\sum_{l=0}^{L} \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} P_l \bigg)^{|V_k|} \bigg(\sum_{l=0}^{L} \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} Q_l \bigg)^{|V_k'|} \bigg(\sum_{l=0}^{L} \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} P_l \bigg)^{|V_1|} \bigg(\sum_{l=0}^{L} \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} Q_l \bigg)^{|V_1'|}$$

$$= \bigg(\frac{\sum_{l=0}^{L} \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} P_l}{\sum_{l=0}^{L} \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} Q_l} \bigg)^{|V_k|} \bigg(\frac{\sum_{l=0}^{L} \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} Q_l}{\sum_{l=0}^{L} \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} P_l} \bigg)^{|V_1'|}$$

$$\bigg(\sum_{l=0}^{L} \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} Q_l \bigg)^{|V_k'|-|V_k|} \bigg(\sum_{l=0}^{L} \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} P_l \bigg)^{|V_1|-|V_1'|}. \tag{57}$$

Since $\frac{1}{n-1} d_H(\sigma_0, \tilde{\sigma}_u) \leq \gamma$, we have

$$\min(|V_1|, |V_k'|) \geq \frac{n}{\beta K} - 1 - \gamma(n-1) \geq \frac{n}{2\beta K} - \gamma n, \tag{58}$$

$$\max(|V_1'|, |V_k|) \leq \gamma(n-1) \leq \gamma n. \tag{59}$$

Observe that the bounds (51) and (52) satisfy the conditions of Lemmas B.3 and B.4. Define

42

$L_1 := \{l \in \{0, \ldots, L\} : \frac{n\Delta_l^2}{P_l \vee Q_l} \geq 1\}$. Then

$$\left| 1 - \frac{\sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} P_l}{\sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} Q_l} \right| = \left| \frac{\sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} (P_l - Q_l)}{\sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} Q_l} \right| \overset{(a)}{\leq} \frac{4}{\sum_{l=0}^L \sqrt{P_l Q_l}} \left| \sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} (P_l - Q_l) \right|$$

$$\overset{(b)}{\leq} 8 \left| \sum_{l=0}^L \left(\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} - 1 \right) (P_l - Q_l) \right|$$

$$\leq 8 \left| \sum_{l \in L_1} \left(\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} - 1 \right) (P_l - Q_l) \right| + 8 \sum_{l \notin L_1} \left| \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} - 1 \right| |P_l - Q_l|$$

$$\overset{(c)}{\leq} 8(1 + 6\eta\rho) \sum_{l \in L_1} \frac{\Delta_l^2}{Q_l} + 32\rho \sum_{i \notin L_1} \frac{\Delta_l}{\sqrt{n(P_l \vee Q_l)}}$$

$$\leq 8(1 + 6\eta\rho)\rho \sum_{l \in L_1} \frac{\Delta_l^2}{P_l \vee Q_l} + 32\rho \sum_{l \notin L_1} \frac{\Delta_l}{\sqrt{n(P_l \vee Q_l)}}$$

$$\overset{(d)}{\leq} 2^6 \rho I(\{P_l\}, \{Q_l\}) + 32\rho \frac{L+1}{n} \leq 2^7 \rho I(\{P_l\}, \{Q_l\}),$$

where (a) follows from Claim 3 of Lemma B.3, (b) follows because $\sum_{l=0}^L \sqrt{P_l Q_l} \geq e^{-I(\{P_l\}, \{Q_l\})} \geq \frac{1}{2}$ by assumption, (c) follows by Lemma B.4, and (d) follows from Lemma B.5 and the fact that $2\eta\rho \leq 1$. An identical analysis shows that $\left| 1 - \frac{\sum_{l=0}^L \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} Q_l}{\sum_{l=0}^L \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} P_l} \right| \leq 2^7 \rho I(\{P_l\}, \{Q_l\})$. Using the fact that $|x| \leq \exp(|1 - x|)$ for all $x \in \mathbb{R}$, we have

$$\left(\frac{\sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} P_l}{\sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} Q_l} \right)^{|V_k|} \left(\frac{\sum_{l=0}^L \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} Q_l}{\sum_{l=0}^L \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} P_l} \right)^{|V_1'|} \leq \exp\big((|V_k| + |V_1'|) 2^7 \rho I(\{P_l\}, \{Q_l\}) \big)$$

$$\leq \exp\big(2^7 \gamma \rho n I(\{P_l\}, \{Q_l\}) \big)$$

$$\leq \exp\big(2^7 \rho \beta K \gamma \frac{n}{\beta K} I(\{P_l\}, \{Q_l\}) \big). \tag{60}$$

We now use the shorthand $\hat{I} := -\log\left\{ \left(\sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} Q_l \right) \left(\sum_{l=0}^L \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} P_l \right) \right\}$, and define $m_k := |V_k'| - |V_k|$ and $m_1 := |V_1| - |V_1'|$. Then

$$\left(\sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} Q_l \right)^{|V_k'| - |V_k|} \left(\sum_{l=0}^L \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} P_l \right)^{|V_1| - |V_1'|}$$

$$= \exp\left(-\frac{m_1 + m_k}{2} \hat{I} \right) \left(\sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} Q_l \right)^{\frac{m_k - m_1}{2}} \left(\sum_{l=0}^L \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} P_l \right)^{\frac{m_1 - m_k}{2}}. \tag{61}$$

We now have

$$\hat{I} - I(\{P_l\}, \{Q_l\}) = -\log \frac{\left(\sum_{l=0}^L \sqrt{\frac{\hat{P}_l}{\hat{Q}_l}} Q_l \right) \left(\sum_{l=0}^L \sqrt{\frac{\hat{Q}_l}{\hat{P}_l}} P_l \right)}{\left(\sum_{l=0}^L \sqrt{P_l Q_l} \right)^2}. \tag{62}$$

Let us first consider the numerator:

$$\left(\sum_{l=0}^{L}\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}Q_l\right)\left(\sum_{l=0}^{L}\sqrt{\frac{\hat{Q}_l}{\hat{P}_l}}P_l\right) = \left(\sum_{l=0}^{L}\sqrt{P_lQ_l}\sqrt{\frac{\hat{P}_l}{P_l}\frac{Q_l}{\hat{Q}_l}}\right)\left(\sum_{l=0}^{L}\sqrt{P_lQ_l}\sqrt{\frac{P_l}{\hat{P}_l}\frac{\hat{Q}_l}{Q_l}}\right)$$

$$= \sum_{l=0}^{L}P_lQ_l + 2\sum_{l<l'}\sqrt{P_lQ_lP_{l'}Q_{l'}} + \sum_{l<l'}\sqrt{P_lQ_lP_{l'}Q_{l'}}\left(\sqrt{T_{l,l'}} + \frac{1}{\sqrt{T_{l,l'}}} - 2\right)$$

$$= \left(\sum_{l=0}^{L}\sqrt{P_lQ_l}\right)^2 + \sum_{l<l'}\sqrt{P_lQ_lP_{l'}Q_{l'}}\left(\sqrt{T_{l,l'}} + \frac{1}{\sqrt{T_{l,l'}}} - 2\right),$$

where, we write $T_{l,l'} := \frac{\hat{P}_l}{P_l}\frac{Q_l}{\hat{Q}_l}\frac{P_{l'}}{\hat{P}_{l'}}\frac{\hat{Q}_{l'}}{Q_{l'}}$. Furthermore, since $\sum_{l=0}^{L}\sqrt{P_lQ_l} \geq 1/2$, we have

$$\hat{I} - I(\{P_l\},\{Q_l\})$$

$$= -\log\left\{1 + \frac{\sum_{l<l'}\sqrt{P_lQ_lP_{l'}Q_{l'}}\left(\sqrt{T_{l,l'}} + \frac{1}{\sqrt{T_{l,l'}}} - 2\right)}{\left(\sum_{l=0}^{L}\sqrt{P_lQ_l}\right)^2}\right\}$$

$$\geq -\log\left\{1 + 4\sum_{l<l'}\sqrt{P_lQ_lP_{l'}Q_{l'}}\left(\sqrt{T_{l,l'}} + \frac{1}{\sqrt{T_{l,l'}}} - 2\right)\right\}$$

$$\geq -4\sum_{l<l'}\sqrt{P_lQ_lP_{l'}Q_{l'}}\left(\sqrt{T_{l,l'}} + \frac{1}{\sqrt{T_{l,l'}}} - 2\right). \tag{63}$$

We now bound $|T_{l,l'} - 1|$:

$$|T_{l,l'} - 1| = \left|\frac{\hat{P}_l}{P_l}\frac{Q_l}{\hat{Q}_l}\frac{P_{l'}}{\hat{P}_{l'}}\frac{\hat{Q}_{l'}}{Q_{l'}} - 1\right|$$

$$= \left|\left(1 - \frac{P_l - \hat{P}_l}{P_l}\right)\left(1 - \frac{\hat{Q}_l - Q_l}{\hat{Q}_l}\right)\left(1 - \frac{\hat{P}_{l'} - P_{l'}}{\hat{P}_{l'}}\right)\left(1 - \frac{Q_{l'} - \hat{Q}_{l'}}{Q_{l'}}\right) - 1\right|$$

$$\overset{(a)}{\leq} 4\left(\frac{|P_l - \hat{P}_l|}{P_l} + \frac{|\hat{Q}_l - Q_l|}{\hat{Q}_l} + \frac{|\hat{P}_{l'} - P_{l'}|}{\hat{P}_{l'}} + \frac{|Q_{l'} - \hat{Q}_{l'}|}{Q_{l'}}\right)$$

$$\overset{(b)}{\leq} 8\left(\frac{|P_l - \hat{P}_l|}{P_l} + \frac{|\hat{Q}_l - Q_l|}{Q_l} + \frac{|\hat{P}_{l'} - P_{l'}|}{P_{l'}} + \frac{|Q_{l'} - \hat{Q}_{l'}|}{Q_{l'}}\right),$$

where (a) and (b) follow from Lemma B.3. Suppose without loss of generality that for all $l = \{0,\ldots,L-1\}$, we have $\frac{|\hat{P}_l - P_l|}{P_l} + \frac{|\hat{Q}_l - Q_l|}{Q_l} \geq \frac{|\hat{P}_{l+1} - P_{l+1}|}{P_{l+1}} + \frac{|\hat{Q}_{l+1} - Q_{l+1}|}{Q_{l+1}}$. Then for $l, l' \in \{0,\ldots,L\}$ such that $l < l'$, we have, by Proposition B.1, that

$$|T_{l,l'} - 1| \leq 8\left(\frac{|\hat{P}_l - P_l|}{P_l} + \frac{|\hat{Q}_l - Q_l|}{Q_l}\right) \overset{(a)}{\leq} \begin{cases} 8\eta\frac{\Delta_l}{P_l \vee Q_l}, & \text{if } l \in L_1, \\ 8\eta\frac{1}{\sqrt{n(P_l \vee Q_l)}}, & \text{if } l \notin L_1. \end{cases} \tag{64}$$

Applying Taylor's theorem on the function $g(x) = x^{1/2} + x^{-1/2} - 2$ and using the fact that $\left|\frac{3}{8}x^{-5/2} - \frac{15}{8}x^{-7/2}\right| \leq 24$ for all $x \in [1/2, 3/2]$, we have, for some $\tau_{l,l'} \in [-3,3]$, that $\sqrt{T_{l,l'}} + \frac{1}{\sqrt{T_{l,l'}}} - 2 =$

$\tau_{l,l'}(T_{l,l'}-1)^2$. Continuing from inequality (63), we obtain

$$\hat{I} - I(\{P_l\}, \{Q_l\})$$

$$\geq -4\sum_{l<l'}\sqrt{P_lQ_lP_{l'}Q_{l'}}\left(\sqrt{T_{l,l'}} + \frac{1}{\sqrt{T_{l,l'}}} - 2\right)$$

$$\geq -4\sum_{l\in L_1}\sum_{l':l'>l}\sqrt{P_lQ_lP_{l'}Q_{l'}}\left(\sqrt{T_{l,l'}} + \frac{1}{\sqrt{T_{l,l'}}} - 2\right)$$

$$-4\sum_{l\notin L_1}\sum_{l':l'>l}\sqrt{P_lQ_lP_{l'}Q_{l'}}\left(\sqrt{T_{l,l'}} + \frac{1}{\sqrt{T_{l,l'}}} - 2\right)$$

$$\geq -24\eta\sum_{l\in L_1}\sum_{l'>l}\sqrt{P_lQ_lP_{l'}Q_{l'}}\left(\frac{\Delta_l}{P_l\vee Q_l}\right)^2 - 24\eta\sum_{l\notin L_1}\sum_{l'>l}\sqrt{P_lQ_lP_{l'}Q_{l'}}\frac{1}{n(P_l\vee Q_l)}$$

$$\geq -24\eta\left(\sum_{l\in L_1}\frac{\Delta_l^2\sqrt{P_lQ_l}}{(P_l\vee Q_l)^2}\right)\left(\sum_{l'=0}^L\sqrt{P_{l'}Q_{l'}}\right) - 24\eta\left(\sum_{l\notin L_1}\frac{\sqrt{P_lQ_l}}{n(P_l\vee Q_l)}\right)\left(\sum_{l'=0}^L\sqrt{P_{l'}Q_{l'}}\right)$$

$$\geq -24\eta\left(\sum_{l\in L_1}\frac{\Delta_l^2}{P_l\vee Q_l}\right)\left(\sum_{l'=0}^L\sqrt{P_{l'}Q_{l'}}\right) - 24\eta\left(\sum_{l\notin L_1}\frac{1}{n}\right)\left(\sum_{l'=0}^L\sqrt{P_{l'}Q_{l'}}\right)$$

$$\overset{(a)}{=} -2^7\eta I(\{P_l\}, \{Q_l\}),$$

where (a) follows from the assumptions that $I(\{P_l\}, \{Q_l\}) \leq 2\log 2$ and $I(\{P_l\}, \{Q_l\}) \geq \frac{L+1}{n}$, and also from Lemma B.5. Thus, using the fact that $m_k = |V_k'| - |V_k| \geq \frac{n}{2\beta K} - 2\gamma n$ and $m_1 = |V_1| - |V_1'| \geq \frac{n}{2\beta K} - 2\gamma n$, where both inequalities follow from (58) and (59), we have

$$\exp\left(-\frac{(m_1+m_k)}{2}\hat{I}\right) \leq \exp\left(-\left(\frac{n}{\beta K} - \gamma n\right)I(\{P_l\}, \{Q_l\})(1-2^7\eta)\right)$$

$$\leq \exp\left(-(1-2^7\eta)(1-\beta K\gamma)\frac{n}{\beta K}I(\{P_l\}, \{Q_l\})\right). \tag{65}$$

Now we bound the last two terms of equation (61). We first assume that $m_k \geq m_1$. Then

$$\left(\sum_{l=0}^L\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}Q_l\right)^{\frac{m_k-m_1}{2}}\left(\sum_{l=0}^L\sqrt{\frac{\hat{Q}_l}{\hat{P}_l}}P_l\right)^{\frac{m_1-m_k}{2}}$$

$$=\left(\frac{\sum_{l=0}^L\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}Q_l}{\sum_{l=0}^L\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}\hat{Q}_l}\right)^{\frac{m_k-m_1}{2}}\left(\frac{\sum_{l=0}^L\sqrt{\frac{\hat{Q}_l}{\hat{P}_l}}P_l}{\sum_{l=0}^L\sqrt{\frac{\hat{Q}_l}{\hat{P}_l}}\hat{P}_l}\right)^{\frac{m_1-m_k}{2}}$$

$$=\left(1+\frac{\sum_{l=0}^L\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}(Q_l-\hat{Q}_l)}{\sum_{l=0}^L\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}\hat{Q}_l}\right)^{\frac{m_k-m_1}{2}}\left(1+\frac{\sum_{l=0}^L\sqrt{\frac{\hat{Q}_l}{\hat{P}_l}}(\hat{P}_l-P_l)}{\sum_{l=0}^L\sqrt{\frac{\hat{Q}_l}{\hat{P}_l}}P_l}\right)^{\frac{m_k-m_1}{2}}. \tag{66}$$

By Lemma B.4, we have $\sum_{l=0}^L\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}\hat{Q}_l \geq \frac{1}{2}\sqrt{P_lQ_l} \geq \frac{1}{4}$ and $\sum_{l=0}^L\sqrt{\hat{Q}_l\hat{P}_l}P_l \geq \frac{1}{4}\sqrt{P_lQ_l} \geq \frac{1}{8}$. We also

have

$$\left|\sum_{l=0}^{L}\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}(Q_l-\hat{Q}_l)\right| = \left|\sum_{l=0}^{L}\left(\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}-1\right)(Q_l-\hat{Q}_l)\right|$$

$$\leq \sum_{l\in L_1}\left|\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}-1\right||Q_l-\hat{Q}_l| + \sum_{l\notin L_1}\left|\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}-1\right||Q_l-\hat{Q}_l|$$

$$\overset{(a)}{\leq} \eta(1+6\eta\rho)\sum_{l\in L_1}\frac{\Delta_l^2}{Q_l}+4\eta\rho\frac{L+1}{n}$$

$$\overset{(b)}{\leq} \eta\rho(1+6\eta\rho)\sum_{l\in L_1}\frac{\Delta_l^2}{Q_l\vee P_l}+4\eta\rho I(\{P_l\},\{Q_l\}) \leq 12\eta\rho I(\{P_l\},\{Q_l\}), \tag{67}$$

where (a) follows from Proposition B.1 and Lemma B.4, (b) follows from the assumption that $\frac{nI(\{P_l\},\{Q_l\})}{L+1}\geq 1$, and the last inequality follows from Lemma B.5. By an identical argument, we have

$$\left|\sum_{l=0}^{L}\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}(Q_l-\hat{Q}_l)\right| \leq 12\eta\rho I(\{P_l\},\{Q_l\}). \tag{68}$$

Using inequalities (66), (67), and (68), we have

$$\left(\sum_{l=0}^{L}\sqrt{\frac{\hat{P}_l}{\hat{Q}_l}}Q_l\right)^{\frac{m_k-m_1}{2}}\left(\sum_{l=0}^{L}\sqrt{\frac{\hat{Q}_l}{\hat{P}_l}}P_l\right)^{\frac{m_1-m_k}{2}}$$

$$\leq \exp\big((m_k-m_1)\log(1+2^7\eta\rho I(\{P_l\},\{Q_l\}))\big)$$

$$\leq \exp\big((m_k+m_1)2^7\eta\rho I(\{P_l\},\{Q_l\})\big) \leq \exp\big(2^8\eta\rho n I(\{P_l\},\{Q_l\})\big)$$

$$\leq \exp\big(2^8\beta K\rho\eta\frac{n}{\beta K}I(\{P_l\},\{Q_l\})\big). \tag{69}$$

If $m_1\geq m_k$, inequality (69) still holds by an identical argument.

Finally, we combine inequalities (57), (60), (61), (65), and (69) to obtain

$$\mathbb{P}(E_1(k)^c\cup E_2(k)^c) \leq \mathbb{P}(E_1(k)^c)+\exp\left(-(1-C\eta\rho)\frac{n}{\beta K}I(\{P_l\},\{Q_l\})\right)$$

$$\leq 5(L+1)n^{-6}+\exp\left(-(1-C\beta K\rho\eta)\frac{n}{\beta K}I(\{P_l\},\{Q_l\})\right).$$

A union bound over $\{E_1^c\cup E_2(k)^c : k\in[K]\backslash\{1\}\}$ finishes the proof. $\qquad\square$

B.6 Additional lemmas for Proposition 6.1

Lemma B.2. *Let $L\in\mathbb{Z}^+$ and $(\{P_l\},\{Q_l\})\in\mathcal{P}_L^2$. Suppose $I(\{P_l\},\{Q_l\})\leq 2\log 2$. Let $c\in[0,\infty)$, let $\delta := c\frac{L+1}{n}$, and let*

$$P_l' := P_l(1-\delta)+\frac{\delta}{L+1}, \quad and \quad Q_l' := Q_l(1-\delta)+\frac{\delta}{L+1}.$$

Then $P_l', Q_l'\geq\frac{c}{n}$ for all $l\in\{0,\dots,L\}$, and

$$1-c\frac{L+1}{n} \geq \frac{I(\{P_l'\},\{Q_l'\})}{I(\{P_l\},\{Q_l\})} \geq \frac{1}{1+2I(\{P_l\},\{Q_l\})}\left(1-c\frac{L+1}{n}\right).$$

Proof. Define $P''_l := \frac{1}{L+1}$ and $Q''_l := \frac{1}{L+1}$ for all $l \in \{0, \ldots, L\}$. Then $(\{P'_l\}, \{Q'_l\}) = (1 - \delta)(\{P_l\}, \{Q_l\}) + \delta(\{P''_l\}, \{Q''_l\})$. Since $I : \mathcal{P}_L \times \mathcal{P}_L \to [0, \infty]$ is convex, we have

$$I(\{P'_l\}, \{Q'_l\}) \le (1 - \delta)I(\{P_l\}, \{Q_l\}) + \delta I(\{P''_l\}, \{Q''_l\}) \le \left(1 - \frac{L+1}{n}\right) I(\{P_l\}, \{Q_l\}).$$

Define the shorthand $H := \sum_{l=0}^{L}(\sqrt{P_l} - \sqrt{Q_l})^2$ and $H' := \sum_{l=0}^{L}(\sqrt{P'_l} - \sqrt{Q'_l})^2$. We then have

$$
\begin{aligned}
H - H' &= \sum_{l=0}^{L}(\sqrt{P_l} - \sqrt{Q_l})^2 \left(1 - \frac{(\sqrt{P'_l} - \sqrt{Q'_l})^2}{(\sqrt{P_l} - \sqrt{Q_l})^2}\right) \\
&= \sum_{l=0}^{L}(\sqrt{P_l} - \sqrt{Q_l})^2 \left(1 - \left(1 - c\frac{L+1}{n}\right)^2 \left(\frac{\sqrt{P_l} + \sqrt{Q_l}}{\sqrt{P'_l} + \sqrt{Q'_l}}\right)^2\right) \\
&\overset{(a)}{\le} 2c\frac{L+1}{n}\sum_{l=0}^{L}(\sqrt{P_l} - \sqrt{Q_l})^2 \le 2c\frac{L+1}{n}H,
\end{aligned}
$$

where (a) is true because $(\sqrt{P'_l} + \sqrt{Q'_l})^2 \le (\sqrt{P_l} + \sqrt{Q_l})^2$. Note also that since $I(\{P_l\}, \{Q_l\}) \le 2\log 2$ and $I(\{P_l\}, \{Q_l\}) = -2\log(1 - H/2)$, we have $H \le 1$, so by Lemma H.1, we conclude that $H \le I(\{P_l\}, \{Q_l\}) \le H(1 + 2H)$.

Therefore, by Lemma H.1, we have

$$
\begin{aligned}
I(\{P'_l\}, \{Q'_l\}) \ge H' &\ge H\left(1 - 2c\frac{L+1}{n}\right) \overset{(a)}{\ge} I(\{P_l\}, \{Q_l\})\frac{1}{1 + 2H}\left(1 - 2c\frac{L+1}{n}\right) \\
&\overset{(b)}{\ge} I(\{P_l\}, \{Q_l\})\frac{1}{1 + 2I(\{P_l\}, \{Q_l\})}\left(1 - 2c\frac{L+1}{n}\right).
\end{aligned}
$$

$\square$

We often use the bound $\frac{1}{2}P \le \hat{P}_l \le 2P_l$, justified in the following lemma:

Lemma B.3. *Let $L \in \mathbb{Z}^+$, $\rho \in [1, \infty)$, let $(\{P_l\}, \{Q_l\}) \in \mathcal{G}_{L,\rho}$, and suppose $P_l \vee Q_l \ge \frac{1}{n}$ for all $l \in \{0, \ldots, L\}$. Suppose there exists $\eta \in [0, \frac{1}{2\rho})$ such that, for all $l \in \{0, \ldots, L\}$, we have*

$$\max\{|\hat{P}_l - P_l|, |\hat{Q}_l - Q_l|\} \le \eta\left(\Delta_l \vee \left(\frac{P_l \vee Q_l}{n}\right)^{1/2}\right).$$

Then for all $l \in \{0, \ldots, L\}$:

1. *It holds that*

$$\max\left\{\frac{|\hat{P}_l - P_l|}{P_l}, \frac{|\hat{Q}_l - Q_l|}{Q_l}\right\} \le \eta\rho.$$

2. *It holds that $\frac{1}{2}P_l \le \hat{P}_l \le 2P_l$ and $\frac{1}{2}Q_l \le \hat{Q}_l \le 2Q_l$.*

3. *It holds that $\left(\frac{\hat{P}_l}{\hat{Q}_l}\right)^{1/2}Q_l \ge \frac{1}{2}\sqrt{\hat{P}_l\hat{Q}_l} \ge \frac{1}{4}\sqrt{P_lQ_l}.$*

Proof. Fix an $l \in \{0, \ldots, L\}$ arbitrarily. We prove the first claim for P_l; the same argument applies to Q_l. Since $n(P_l \vee Q_l) \ge 1$, we have

$$\frac{|\hat{P}_l - P_l|}{P_l} \le \eta\max\left\{\frac{\Delta_l}{P_l}, \frac{\sqrt{(P_l \vee Q_l)}}{\sqrt{n}P_l}\right\} \le \eta\rho.$$

The second claim follows from Claim 1, since $\eta\rho \leq \frac{1}{2}$ by assumption. The third claim follows because, by Claim 2, we have

$$\left(\frac{\hat{P}_l}{\hat{Q}_l}\right)^{1/2} Q_l \geq \sqrt{\hat{P}_l \hat{Q}_l} \frac{Q_l}{\hat{Q}_l} \geq \frac{1}{2}\sqrt{\hat{P}_l \hat{Q}_l} \geq \frac{1}{4}\sqrt{P_l Q_l}.$$

$\square$

Lemma B.4. *Let* $L \in \mathbb{Z}^+$, $\rho \in [1, \infty)$, *let* $(\{P_l\}, \{Q_l\}) \in \mathcal{G}_{L,\rho}$, *and suppose* $P_l \vee Q_l \geq \frac{1}{n}$ *for all* $l \in \{0, \ldots, L\}$. *Suppose there exists* $\eta \in [0, \frac{1}{2\rho})$ *such that, for all* $l \in \{0, \ldots, L\}$, *we have*

$$\max\{|\hat{P}_l - P_l|, |\hat{Q}_l - Q_l|\} \leq \eta\left(\Delta_l \vee \left(\frac{P_l \vee Q_l}{n}\right)^{1/2}\right).$$

Then:

1. *For all* l *satisfying* $n\frac{\Delta_l^2}{P_l \vee Q_l} \geq 1$, *we have*

$$\left|\left(\frac{\hat{P}_l}{\hat{Q}_l}\right)^{1/2} - 1\right| \leq \frac{\Delta}{Q_l}(1 + 6\eta\rho) \quad and \quad \left|\left(\frac{\hat{Q}_l}{\hat{P}_l}\right)^{1/2} - 1\right| \leq \frac{\Delta}{P_l}(1 + 6\eta\rho). \tag{70}$$

2. *For all* l *satisfying* $n\frac{\Delta_l^2}{P_l \vee Q_l} < 1$, *we have*

$$\left|\left(\frac{\hat{P}_l}{\hat{Q}_l}\right)^{1/2} - 1\right| \leq 4\rho\frac{1}{\sqrt{n(P_l \vee Q_l)}} \quad and \quad \left|\left(\frac{\hat{Q}_l}{\hat{P}_l}\right)^{1/2} - 1\right| \leq 4\rho\frac{1}{\sqrt{n(P_l \vee Q_l)}}. \tag{71}$$

Proof. Fix an $l \in \{0, \ldots, L\}$, and suppose $\Delta_l \geq \left(\frac{P_l \vee Q_l}{n}\right)^{1/2}$. Define the shorthand

$$\eta_1 := \frac{\hat{P}_l - P_l}{P_l - Q_l}, \quad \eta_2 := \frac{Q_l - \hat{Q}_l}{P_l - Q_l}, \quad \eta_3 := \frac{Q_l}{\hat{Q}_l} - 1.$$

Then

$$\frac{\hat{P}_l}{\hat{Q}_l} - 1 = \frac{\hat{P}_l - P_l + (P_l - Q_l) + (Q_l - \hat{Q}_l)}{Q_l}\frac{Q_l}{\hat{Q}_l} = \frac{P_l - Q_l}{Q_l}(1 + \eta_1 + \eta_2)(1 + \eta_3).$$

Now note that $|\eta_1| \leq \frac{|\hat{P}_l - P_l|}{\Delta_l} \leq \eta$. Likewise, we have $|\eta_2| \leq \eta$. Finally, we know that $|\eta_3| \leq \frac{|\hat{Q}_l - Q_l|}{Q_l}\frac{Q_l}{\hat{Q}_l} \leq 2\eta\rho$ by Lemma B.3. Therefore, we have $\left|\frac{\hat{P}_l}{\hat{Q}_l} - 1\right| \leq \left|\frac{P_l - Q_l}{Q_l}\right|(1 + 2\eta)(1 + 2\eta\rho)$. We then use the fact that $2\eta\rho \leq 1$ and apply the inequality $|\sqrt{x} - 1| \leq |x - 1|$ for all $x \geq 0$ to prove the first inequality of the first case (70). The second inequality holds by symmetry.

Now suppose $\Delta_l < \left(\frac{P_l \vee Q_l}{n}\right)^{1/2}$. Define the shorthand

$$\eta_1 := \frac{\hat{P}_l - P_l}{Q_l}\sqrt{n(P_l \vee Q_l)}, \quad \eta_2 := \frac{Q_l - \hat{Q}_l}{Q_l}\sqrt{n(P_l \vee Q_l)}, \quad \eta_3 := \frac{Q_l}{\hat{Q}_l} - 1, \quad \tau := \frac{P_l - Q_l}{Q_l}\sqrt{n(P_l \vee Q_l)}.$$

Then

$$\frac{\hat{P}_l}{\hat{Q}_l} - 1 = \frac{\hat{P}_l - P_l + (P_l - Q_l) + (Q_l - \hat{Q}_l)}{Q_l}\frac{Q_l}{\hat{Q}_l} = \frac{1}{\sqrt{n(P_l \vee Q_l)}}(\eta_1 + \eta_2 + \tau)(1 + \eta_3).$$

Observe $|\eta_1| \leq \eta\frac{P_l \vee Q_l}{Q_l} \leq \eta\rho$, and likewise $|\eta_2| \leq \eta\rho$. We use Lemma B.3, the fact that $Q_l n \geq 1$, and $\rho \geq 1$ to bound $|\eta_3| \leq 2\eta\frac{\sqrt{P_l \vee Q_l}}{Q_l\sqrt{n}} \leq 2\eta\rho^{1/2} \leq 2\eta\rho$. Finally, we have $|\tau| \leq \frac{\Delta_l}{Q_l}\sqrt{n(P_l \vee Q_l)} \leq \rho$.

Therefore, we have $\left|\frac{\hat{P}_l}{\hat{Q}_l} - 1\right| \le \frac{1}{\sqrt{n(P_l \vee Q_l)}}(2\eta\rho + \rho)(1 + 2\eta\rho)$. Since $2\eta\rho \le 1 \le \rho$ and $|\sqrt{x} - 1| \le |x - 1|$ for all $x \ge 0$, the first inequality of the second case (71) holds. The second inequality holds by symmetry.

$\square$

Lemma B.5. *Let $L \in \mathbb{Z}^+$, let $(\{P_l\}, \{Q_l\}) \in \mathcal{P}_L \times \mathcal{P}_L$, and let $\Delta_l := |P_l - Q_l|$. Let $C \in (0, \infty)$ be such that $\frac{nI(\{P_l\}, \{Q_l\})}{L+1} \ge 4C$ and suppose $I(\{P_l\}, \{Q_l\}) \le 2\log 2$. Define $L_1 := \{l \in \{0, \dots, L\} : \frac{\Delta_l^2}{P_l \vee Q_l} \ge C\}$. Then*

$$\frac{1}{2} \sum_{l \in L_1} \frac{\Delta_l^2}{P_l \vee Q_l} \le I(\{P_l\}, \{Q_l\}) \le 4 \sum_{l \in L_1} \frac{\Delta_l^2}{P_l \vee Q_l}. \tag{72}$$

Proof. Observe that

$$I(\{P_l\}, \{Q_l\}) = -2\log\left(1 - \frac{\sum_{l=0}^{L}(\sqrt{P_l} - \sqrt{Q_l})^2}{2}\right).$$

Since $\log(1 - x) \le -x$ for all $x \in (-1, \infty)$, we have

$$I(\{P_l\}, \{Q_l\}) \ge \sum_{l=0}^{L}(\sqrt{P_l} - \sqrt{Q_l})^2 \ge \sum_{l=0}^{L} \frac{\Delta_l^2}{(\sqrt{P_l} + \sqrt{Q_l})^2} \ge \frac{1}{2} \sum_{l \in L_1} \frac{\Delta_l^2}{P_l \vee Q_l}.$$

On the other hand, we know that $\sum_{l=0}^{L} \sqrt{P_l Q_l} = e^{-\frac{1}{2}I(\{P_l\}, \{Q_l\})} \ge \frac{1}{2}$. By the fact that $-2x \le \log(1 - x)$ for all $x \in [0, 1/2]$, we have

$$I(\{P_l\}, \{Q_l\}) \le 2\sum_{l=0}^{L}(\sqrt{P_l} - \sqrt{Q_l})^2 \le 2\sum_{l=0}^{L} \frac{\Delta_l^2}{P_l \vee Q_l}$$

$$\le 2\sum_{l \in L_1} \frac{\Delta_l^2}{P_l \vee Q_l} + 2\frac{(L+1)C}{n} \le 2\sum_{l \in L_1} \frac{\Delta_l^2}{P_l \vee Q_l} + \frac{I(\{P_l\}, \{Q_l\})}{2}.$$

$\square$

The following lemma slightly expands upon Lemma 4 of Gao et al [17].

Lemma B.6. *Let $\sigma, \sigma' : [n] \to [K]$ be two clusters such that, for some $T \in [n]$, the minimum cluster size of σ is at least T. Suppose $l(\sigma, \sigma') < \frac{T}{2n}$. Then there is a unique $\xi \in S_K$ such that $l(\sigma, \sigma') = \frac{1}{n}d_H(\xi \circ \sigma, \sigma')$; furthermore, the unique permutation ξ is of the form*

$$\xi(k) = \underset{k' \in [K]}{\arg\max} |\{v \in [n] : \sigma(v) = k\} \cap \{v \in [n] : \sigma'(v) = k'\}|. \tag{73}$$

Proof. Suppose $\pi \in S_K$ satisfies $l(\sigma, \sigma') = \frac{1}{n}d_H(\pi \circ \sigma, \sigma') < \frac{T}{2n}$. Fix $k \in [K]$. Then

$$|\{u \in [n] : \sigma(u) = k\} \cap \{u \in [n] : \sigma'(u) \ne \pi(k)\}| \le d_H(\pi \circ \sigma, \sigma') < \frac{T}{2},$$

and

$$|\{u \in [n] : \sigma(u) = k\} \cap \{u \in [n] : \sigma'(u) = \pi(k)\}|$$
$$= |\{u \in [n] : \sigma(u) = k\}| - |\{u \in [n] : \sigma(u) = k\} \cap \{u \in [n] : \sigma'(u) \ne \pi(k)\}|$$
$$\ge T - d_H(\xi \circ \sigma, \sigma') \ge \frac{T}{2}.$$

It thus follows that $\pi(k)$ is the unique maximizer of $\arg\max_{k' \in [K]} |\{u \in [n] : \sigma(u) = k\} \cap \{u \in [n] : \sigma'(u) = k'\}|$. Since k was fixed arbitrarily, the lemma follows. $\square$

C   Proof of Proposition 6.2

Proof. Let μ be the Lebesgue measure on $[0,1]$. Let us arbitrarily fix $((P_0, p), (Q_0, q)) \in \mathcal{G}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t}$ and define $\tilde{g}$ as in Condition C2. Let $L \in \mathbb{N}$, and suppose $L \geq \tilde{c}_1^{-1} \vee \tilde{c}_2^{-1}$. Let $\{[a_l, b_l]\}_{l \in \{1, \ldots, L\}}$ be a uniformly spaced binning of $[0,1]$, and for each $l \in \{1, \ldots, L\}$, define

$$\tilde{P}_l := \int_{a_l}^{b_l} \tilde{p}(z) dz \quad \text{and} \quad \int_{a_l}^{b_l} \tilde{q}(z) dz. \tag{74}$$

Define $R := \{z \in [0,1] : \tilde{g}(z) \leq (2\tilde{C}L)^{1/r}\}$. By Markov's inequality, we have $\mu(R^c) \leq \frac{1}{2L}$. Since $\tilde{g}(z)$ is quasi-convex, R must be an interval. Thus, only bins $[0, 1/L]$ and $[1 - 1/L, 1]$ have non-empty intersection with R^c. Let $[a_l, b_l]$ be a bin such that $[a_l, b_l] \subset R$. Then

$$\frac{P_l}{Q_l} = \frac{(1 - P_0) \int_{a_l}^{b_l} \tilde{p}(z) \, dz}{(1 - Q_0) \int_{a_l}^{b_l} \tilde{q}(z)} \leq \tilde{C} \frac{\int_{a_l}^{b_l} \frac{\tilde{p}(z)}{\tilde{q}(z)} \tilde{q}(z) \, dz}{\int_{a_l}^{b_l} \tilde{q}(z) \, dz} \leq \tilde{C} \exp((2\tilde{C}L)^{1/r}).$$

Likewise, we can show that $\frac{P_l}{Q_l} \geq \tilde{C}^{-1} \exp(-(2\tilde{C}L)^{1/r})$. Now we consider $[0, 1/L]$ and suppose $[0, 1/L] \cap R^c \neq \emptyset$. Define $\tilde{P}_l' := \int_{\mathrm{bin}_l \cap R} \tilde{p}(z) \, dz$ and $\tilde{Q}_l' := \int_{\mathrm{bin}_l \cap R} \tilde{q}(z) \, dz$, and define $\tilde{P}_l'' := \int_{\mathrm{bin}_l \cap R^c} p(z) \, dz$ and $\tilde{Q}_l'' := \int_{\mathrm{bin}_l \cap R^c} q(z) \, dz$. We can use the same reasoning as above to show that $\exp(-(2\tilde{C}L)^{1/r}) \leq \frac{\tilde{P}_l'}{\tilde{Q}_l'} \leq \exp((2\tilde{C}L)^{1/r})$. Since $L \geq \tilde{c}_1^{-1}$, both $\tilde{p}(z)$ and $\tilde{q}(z)$ are non-decreasing in $[0, 1/L]$ by Assumption C5. Thus,

$$\tilde{P}_l' \geq \min_{z \in [0, 1/L] \cap R} \frac{\tilde{p}(z)}{2L} \geq \max_{z \in [0, 1/L] \cap R^c} \frac{\tilde{p}(z)}{2L} \geq \tilde{P}_l'',$$

where the first inequality follows because $\mu(R^c) \leq \frac{1}{2L}$. With the same reasoning, we know that $Q_l' \geq Q_l''$. Thus,

$$\frac{1}{2} \exp(-(2\tilde{C}L)^{1/r}) \leq \frac{\tilde{P}_l'}{2\tilde{Q}_l'} \leq \frac{\tilde{P}_l}{\tilde{Q}_l} \leq \frac{2\tilde{P}_l'}{\tilde{Q}_l'} \leq 2 \exp((2\tilde{C}L)^{1/r}).$$

Therefore, we have

$$\frac{1}{2\tilde{C}} \exp(-(2\tilde{C}L)^{1/r}) \leq \frac{(1 - P_0)\tilde{P}_l}{(1 - Q_0)\tilde{Q}_0} \leq 2\tilde{C} \exp((2\tilde{C}L)^{1/r}).$$

Using identical reasoning, we may obtain the same bounds for $\frac{P_l}{Q_l}$ for l corresponding to the $[1 - 1/L, 1]$ bin. This proves the first claim of the proposition. Moreover, if $\tilde{g}$ is non-decreasing, then $R^c \subseteq [1 - 1/L, 1]$. Thus, the condition that $\tilde{p}, \tilde{q}$ are non-increasing in $(1 - c_2, 1]$ yields the first claim of the proposition as observed in footnote 3.

For the second claim, define

$$\tilde{H} := \int_0^1 (\sqrt{\tilde{p}(z)} - \sqrt{\tilde{q}(z)})^2 \, dz, \quad \tilde{H}_L := \sum_{l=1}^{L} (\sqrt{\tilde{P}_l} - \sqrt{\tilde{Q}_l})^2,$$

$$H := (\sqrt{P_0} - \sqrt{Q_0})^2 + (\sqrt{1 - P_0} - \sqrt{1 - Q_0})^2 + \sqrt{(1 - P_0)(1 - Q_0)} \tilde{H},$$

$$H_L := (\sqrt{P_0} - \sqrt{Q_0})^2 + (\sqrt{1 - P_0} - \sqrt{1 - Q_0})^2 + \sqrt{(1 - P_0)(1 - Q_0)} \tilde{H}_L.$$

Let $\eta_L := \sup\{|1 - \frac{\tilde{H}_L}{\tilde{H}}| : ((P_0, \tilde{p}), (Q_0, \tilde{q})) \in \tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t}\}$; we know then by Proposition C.1 that $\lim_{L \to \infty} \eta_L = 0$. Fix $((P_0, \tilde{p}), (Q_0, \tilde{q})) \in \tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t}$, and let $\{P_l, Q_l\}$ be the corresponding discretized probabilities. Then

$$|H - H_L| = \sqrt{(1 - P_0)(1 - Q_0)} |\tilde{H}_L - \tilde{H}| \leq \sqrt{(1 - P_0)(1 - Q_0)} \tilde{H} \eta_L \leq H \eta_L.$$

Observe that we also have

$$I((P_0, \tilde{p}), (Q_0, \tilde{q})) = -2 \log\left\{ \sqrt{P_0 Q_0} + \sqrt{(1 - P_0)(1 - Q_0)} \int_0^1 \sqrt{\tilde{p}(z)\tilde{q}(z)} \, dz \right\} = -2 \log(1 - \tfrac{1}{2} H),$$

and also $I(\{P_l\}, \{Q_l\}) = -2 \log(1 - \tfrac{1}{2} H_L)$. Hence,

$$I(\{P_l\}, \{Q_l\}) - I((P_0, \tilde{p}), (Q_0, \tilde{q})) = -2 \log \frac{1 - \tfrac{1}{2} H_L}{1 - \tfrac{1}{2} H} = -2 \log\left(1 + \frac{1}{2} \frac{H - H_L}{1 - \tfrac{1}{2} H} \right)$$

$$\geq -\frac{H - H_L}{1 - \tfrac{1}{2} H} \geq -2(H - H_L) \geq -2\eta_L H \overset{(a)}{\geq} -2\eta_L I((P_0, \tilde{p}), (Q_0, \tilde{q})),$$

where (a) follows from Lemma H.1. On the other hand, we have by the Cauchy-Schwarz inequality that

$$\int_0^1 \sqrt{\tilde{p}(z)\tilde{q}(z)} \, dz \leq \left\{ \int_0^1 \tilde{p}(z) \, dz \right\}^{1/2} \left\{ \int_0^1 \tilde{q}(z) \, dz \right\}^{1/2} = \sqrt{\tilde{P}_l \tilde{Q}_l}.$$

Therefore, we have $I((P_0, \tilde{p}), (Q_0, \tilde{q})) \geq I(\{P_l\}, \{Q_l\})$. Since $((P_0, \tilde{p}), (Q_0, \tilde{q}))$ was chosen arbitrarily, the proposition follows. $\qquad \square$

Proposition C.1. *Let $\tilde{C} \in [1, \infty)$, $\tilde{c}_1, \tilde{c}_2 \in (0, 1/2)$, $r > 2$, and $t \in (0, 1)$. For any $((P_0, \tilde{p}), (Q_0, \tilde{q})) \in \tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t}$ and for any $L \in \mathbb{N}$, let $\tilde{P}_l, \tilde{Q}_l$ be defined as in equation (74) for $l \in \{1, \dots, L\}$. Then*

$$\lim_{L \to \infty} \sup_{\substack{((P_0, \tilde{p}), (Q_0, \tilde{q})) \\ \in \tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t}}} \left| 1 - \frac{\sum_{l=1}^L (\sqrt{\tilde{P}_l} - \sqrt{\tilde{Q}_l})^2}{\int_0^1 (\sqrt{\tilde{p}(z)} - \sqrt{\tilde{q}(z)})^2 \, dz} \right| = 0.$$

Proof. Let us arbitrarily fix $((P_0, \tilde{p}), (Q_0, \tilde{q})) \in \tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t}$. Let $\tilde{\alpha} := \int_0^1 (\sqrt{\tilde{p}(z)} - \sqrt{\tilde{q}(z)})^2 \, dz$, and let $\tilde{\gamma}(z) := \frac{\tilde{p}(z) - \tilde{q}(z)}{\tilde{\alpha}}$. Let $L \in \mathbb{N}$ and suppose that $L \geq 4^{-\frac{r(t+1)}{2 - rt}}$.

For each $l = 1, \dots, L$, we also define $\tilde{\gamma}_l := \int_{a_l}^{b_l} \tilde{\gamma}(z) \, dz$. Since $\tilde{p}$ and $\tilde{q}$ are continuous and bounded, we may define $z_l := \arg\max_{z \in [a_l, b_l]} \tilde{p}(z) + \tilde{q}(z)$, $z_l' := \arg\max_{z \in [a_l, b_l]} \tilde{p}(z)$, and $z_l'' := \arg\max_{z \in [a_l, b_l]} \tilde{q}(z)$. We also define the shorthand

$$D_l := \int_{a_l}^{b_l} \frac{\tilde{\gamma}(z)^2}{(\sqrt{\tilde{p}(z)} + \sqrt{\tilde{q}(z)})^2} \, dz, \quad D_l' := \frac{\tilde{\gamma}_l^2}{(\sqrt{\tilde{P}_l} + \sqrt{\tilde{Q}_l})^2}, \quad \text{and} \quad D_l'' := \frac{1}{L} \frac{\tilde{\gamma}(z_l)^2}{(\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')})^2}.$$

for each $l \in \{1, \dots, L\}$.

Let $\tau = \frac{2 + r}{r} \frac{1}{1 + t}$, and note that $0 < \tau < 1$, since $2 < rt$ by assumption. Also, $L^{\tau - 1} \leq \frac{1}{4}$, since we assumed that $L \geq 4^{-\frac{r(t+1)}{2 - rt}}$. Define $\mathcal{B} := \{l \in \{1, \dots, L\} : \sup_{z \in [a_l, b_l]} \tilde{h}(z) \leq L^\tau\}$ and $\mathcal{B}^c := \{1, \dots, L_n\} \setminus \mathcal{B}$. Then

$$\left| \sum_{l=1}^{L_n} D_l - \sum_{l=1}^{L_n} D_l' \right| \leq \sum_{l \in \mathcal{B}^c} |D_l - D_l'| + \sum_{l \in \mathcal{B}} |D_l - D_l'|. \tag{75}$$

We first bound the first term of inequality (75). Let $l \in \mathcal{B}^c$ and note that, by the Cauchy-Schwarz

inequality, we have

$$D_l' \le \frac{\tilde{\gamma}_l^2}{\tilde{P}_l + \tilde{Q}_l} \le \left\{ \int_{a_l}^{b_l} \frac{\tilde{\gamma}(z)}{\tilde{p}(z) + \tilde{q}(z)} \frac{\tilde{p}(z) + \tilde{q}(z)}{\tilde{P}_l + \tilde{Q}_l} \right\}^2 \tilde{P}_l + \tilde{Q}_l$$

$$\le \int_{a_l}^{b_l} \left(\frac{\tilde{\gamma}(z)}{\tilde{p}(z) + \tilde{q}(z)} \right)^2 (\tilde{p}(z) + \tilde{q}(z))\, dz$$

$$\le \left\{ \int_{a_l}^{b_l} \left(\frac{\tilde{\gamma}(z)}{\tilde{p}(z) + \tilde{q}(z)} \right)^r (\tilde{p}(z) + \tilde{q}(z))\, dz \right\}^{2/r} \left\{ \int_{a_l}^{b_l} \tilde{p}(z) + \tilde{q}(z)\, dz \right\}^{(r-2)/r}$$

$$\le \tilde{C}^r L^{-(r-2)/r}.$$

By similar reasoning, we have $D_l \le \tilde{C}^r L^{-(r-2)/r}$. Now, since $\tilde{h}(z)$ is quasi-convex, the set $\{z \in [0,1] : \tilde{h}(z) \le L^\tau\}$ is an interval and $\frac{1}{L}|\mathcal{B}^c| \le \mu(\{z \in [0,1] : \tilde{h}(z) > L^\tau\}) + \frac{4}{L}$. Thus, by Markov's inequality, we have

$$|\mathcal{B}^c| \le L\mu(\{z \in [0,1] : \tilde{h}(z) > L^\tau\}) + 4 \le \tilde{C}L^{1-\tau t} + 4 \le 8\tilde{C}L^{1-\tau t},$$

where the last inequality follows because $1 - \tau t = \frac{r - 2t}{(1+t)r} > 0$. Therefore,

$$\sum_{l \in \mathcal{B}^c} |D_l - D_l'| \le \tilde{C}^r L^{-(r-2)/r} |\mathcal{B}^c| \le 8\tilde{C}^{r+1} L^{-\frac{r-2}{r} + (1 - \tau t)} = 8\tilde{C}^{r+1} L^{\frac{2 - rt}{(1+t)r}}. \tag{76}$$

We now turn our attention to the second term of inequality (75). Let $l \in \mathcal{B}$, and let $z \in [a_l, b_l]$. By the Mean Value Theorem, there exists some $c_{\gamma,z}, c_{p,z}, c_{q,z} \in [a_l, b_l]$ such that

$$\frac{\tilde{\gamma}(z)^2}{(\sqrt{\tilde{p}(z)} + \sqrt{\tilde{q}(z)})^2} = \frac{(\tilde{\gamma}(z_l) + \tilde{\gamma}'(c_{\gamma,z})(z - z_l))^2}{(\sqrt{\tilde{p}(z_l') + \tilde{p}'(c_{p,z})(z - z_l')} + \sqrt{\tilde{q}(z_l'') + \tilde{q}'(c_{q,z})(z - z_l'')})^2}$$

$$= \frac{\left(\frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} + T_1 \right)^2}{\left(\sqrt{\tilde{p}(z_l')}(1 + T_2) + \sqrt{\tilde{q}(z_l'')}(1 + T_3) \right)^2} (\tilde{p}(z_l) + \tilde{q}(z_l))^2, \tag{77}$$

where we denote $T_1 := \frac{\tilde{\gamma}'(c_{\gamma,z})(z - z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)}$, $T_2 := \frac{\tilde{p}'(c_{p,z})}{\tilde{p}(z_l')}(z - z_l)$, and $T_3 := \frac{\tilde{q}'(c_{q,z})}{\tilde{q}(z_l'')}(z - z_l)$. Since $c_{\gamma,z} \in [a_l, b_l]$ and $l \in \mathcal{B}$, we have

$$|T_1| = \left| \frac{\tilde{\gamma}'(c_{\gamma,z})(z - z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} \right| \le \left| \frac{\tilde{\gamma}'(c_{\gamma,z})(z - z_l)}{\tilde{p}(c_{\gamma,z}) + \tilde{q}(c_{\gamma,z})} \right| \le |\tilde{h}(c_{\gamma,z})|L^{-1} \le L^{\tau - 1} \le \frac{1}{4}.$$

Likewise, we have $|T_2|, |T_3| \le L^{\tau-1} \le \frac{1}{4}$.

We now observe that

$$\left(\frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} + T_1 \right)^2 - \left(\frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} \right)^2 \le 2 \left| \frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} \right| |T_1| \tag{78}$$

$$\le \left(\frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} \right)^2 |T_1| + |T_1| \le \left(\frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} \right)^2 L^{\tau-1} + L^{\tau-1} \tag{79}$$

$$\left(\frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} + T_1 \right)^2 - \left(\frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} \right)^2 \ge -2 \left| \frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} \right| |T_1| \tag{80}$$

$$\ge -\left(\frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} \right)^2 |T_1| + |T_1| \ge -\left(\frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} \right)^2 L^{\tau-1} + L^{\tau-1}. \tag{81}$$

Furthermore,

$$\left(\sqrt{\tilde{p}(z_l')(1+T_2)} + \sqrt{\tilde{q}(z_l'')(1+T_3)}\right)^{-2} \geq \left(\left(\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')}\right)(1+|T_2|+|T_3|)\right)^{-2}$$

$$\geq \left(\frac{1}{\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')}}(1-|T_2|-|T_3|)\right)^2$$

$$\geq \left(\frac{1}{\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')}}\right)^2 (1-4L^{\tau-1}), \qquad (82)$$

and

$$\left(\sqrt{\tilde{p}(z_l')(1+T_2)} + \sqrt{\tilde{q}(z_l'')(1+T_3)}\right)^{-2} \leq \left(\left(\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')}\right)(1-|T_2|-|T_3|)\right)^{-2}$$

$$\overset{(a)}{\leq} \left(\frac{1}{\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')}}(1+2|T_2|+2|T_3|)\right)^2$$

$$\leq \left(\frac{1}{\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')}}\right)^2 (1+8L^{\tau-1}). \qquad (83)$$

where (a) follows because $\frac{1}{1-x} \leq 1+2x$ for all $x \in [0, 1/2]$. Combining inequalities (77), (79), (81), (82), and (83), we have

$$\left|\frac{\tilde{\gamma}(z)^2}{(\sqrt{\tilde{p}(z)} + \sqrt{\tilde{q}(z)})^2} - \frac{\tilde{\gamma}(z_l)^2}{(\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')})^2}\right|$$

$$\leq 10L^{\tau-1}\frac{\tilde{\gamma}(z_l)^2}{(\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')})^2} + 8L^{\tau-1}\frac{(\tilde{p}(z_l) + \tilde{q}(z_l))^2}{(\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')})^2}$$

$$\leq 10L^{\tau-1}\frac{\tilde{\gamma}(z_l)^2}{(\sqrt{\tilde{p}(z_l')} + \sqrt{\tilde{q}(z_l'')})^2} + 8L^{\tau-1},$$

where the last inequality holds because $\tilde{p}(z_l') \geq \tilde{p}(z_l)$ and $\tilde{q}(z_l'') \geq \tilde{q}(z_l)$ by the definitions of z_l' and z_l''. Hence, $|D_l - D_l''| \leq 10L^{\tau-2}D_l'' + 8L^{\tau-2}$, which, with the triangle inequality, implies that

$$|D_l - D_l''| \leq 20L^{\tau-2}D_l + 16L^{\tau-2}.$$

We now bound $|D_l' - D_l''|$ in the same manner. Note that

$$\tilde{\gamma}_l = \int_{a_l}^{b_l} \tilde{\gamma}(z_l) + \tilde{\gamma}'(c_{\gamma,z})(z - z_l)\, dz = \frac{1}{L}\left(\frac{\tilde{\gamma}(z_l)}{\tilde{p}(z_l) + \tilde{q}(z_l)} + T_1'\right)(\tilde{p}(z_l) + \tilde{q}(z_l)),$$

where $T_1' := \frac{1}{L}\int_{a_l}^{b_l} \frac{\tilde{\gamma}'(c_{\gamma,z})}{\tilde{p}(z_l) + \tilde{q}(z_l)}(z - z_l)\, dz$ satisfies $|T_1'| \leq L^{\tau-1}$. Likewise, we have $\tilde{P}_l = \frac{1}{L}\tilde{p}(z_l')(1+T_2')$ and $\tilde{Q}_l = \frac{1}{L}\tilde{q}(z_l'')(1+T_3')$, where $T_2' := \frac{1}{L}\int_{a_l}^{b_l}\frac{\tilde{p}'(c_{p,z})}{\tilde{p}(z_l')}(z - z_l)\, dz$ and $T_3' := \frac{1}{L}\int_{a_l}^{b_l}\frac{\tilde{q}'(c_{q,z})}{\tilde{q}(z_l'')}(z - z_l)\, dz$ both satisfy $|T_2'|, |T_3'| \leq L^{\tau-1}$. By similar reasoning as in the case of $|D_l - D_l''|$, we have

$$|D_l' - D_l''| \leq 10L^{\tau-2}D_l'' + 8L^{\tau-2} \leq 20L^{\tau-2}D_l + 16L^{\tau-2}.$$

Therefore,

$$\sum_{l \in \mathcal{B}} |D_l - D_l'| \leq \sum_{l \in \mathcal{B}} |D_l - D_l''| + |D_l' - D_l''|$$

$$\leq 20L^{\tau-1}\sum_{l \in \mathcal{B}} D_l + 16L^{\tau-1} \leq 36L^{\tau-1}. \qquad (84)$$

53

Combining inequalities (76) and (84), we obtain

$$\left| 1 - \frac{\sum_{l=1}^{L}(\sqrt{\tilde{P}_l} - \sqrt{\tilde{Q}_l})^2}{\int_0^1 (\sqrt{\tilde{p}(z)} - \sqrt{\tilde{q}(z)})^2 \, dz} \right| \le \sum_{l=1}^{L} |D_l - D_l'| \le 36 L^{\tau-1} + 8\tilde{C}^{r+1} L^{\frac{2-rt}{(1+t)r}} \le (36 + 8\tilde{C}^{r+1}) L^{\frac{2-rt}{(1+t)r}}.$$

The RHS goes to 0 as $L \to \infty$, since $rt > 2$ by assumption. Since the RHS does not depend on the particular choice of $((P_0, \tilde{p}), (Q_0, \tilde{q}))$, the proposition follows. $\square$

D Proof of Theorem 5.1

We provide the proof of Theorem 5.1, with proofs of supporting propositions in the succeeding subsection.

D.1 Main argument: Proof of Theorem 5.1

Fix $((P_0, p), (Q_0, q)) \in \mathcal{G}_{\Phi, C, c_1, c_2, r, t}$, and let $\tilde{p}, \tilde{q}$ be defined as in Proposition D.1. If the random network $A \in S^{n \times n}$ has the distribution $WSBM(\sigma_0, ((P_0, p), (Q_0, q)))$, then the transformed network $\tilde{A}$ has the distribution $WSBM(\sigma_0, ((P_0, \tilde{p}), (Q_0, \tilde{q})))$. By Proposition D.1, we know that there exists $\tilde{c}_1, \tilde{c}_2 \in (0, 1/2)$ such that $((P_0, \tilde{p}), (Q_0, \tilde{q})) \in \tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t}$.

Let A_{L_n} be the labeled network that is the result of discretizing $\tilde{A}$ with L_n bins, and let $\{P_l, Q_l\}_{l \in \{1, \dots, L_n\}}$ be defined as in equation (6). Then A_{L_n} follows the distribution $LSBM(\sigma_0, (\{P_l\}, \{Q_l\}))$. Let $\rho_n := 2\tilde{C} \exp((2\tilde{C}L_n)^{1/r})$; by Proposition 6.2, we have $(\{P_l\}, \{Q_l\}) \in \mathcal{G}_{L_n, \rho_n}$.

Note that $\frac{n I_n'}{(L_n + 1)\rho_n^2 \log rho_n} \to \infty$. Thus, by Proposition 6.1, there exists $\zeta_n' \to 0$ such that

$$\lim_{n \to \infty} \sup_{\substack{(\{P_l\}, \{Q_l\}) \in \mathcal{G}_{L_n, \rho_n} \\ I_n' \le I(\{P_l\}, \{Q_l\}) \le 2I_n}} \mathbb{P}_{(\{P_l\}, \{Q_l\})} \left(l(\hat{\sigma}(A_{L_n}), \sigma_0) > \exp\left(-(1 - \zeta_n') \frac{n}{\beta K} I(\{P_l\}, \{Q_l\}) \right) \right).$$

We now define

$$\zeta_n := 1 - (1 - \zeta_n') \sup_{\substack{((P_0, \tilde{p}), (Q_0, \tilde{q})) \\ \in \tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t}}} \frac{I((P_0, \tilde{p}), (Q_0, \tilde{q}))}{I(\{P_l\}, \{Q_l\})}.$$

By Proposition 6.2 and the fact that $L_n \to 0$, we have $\lim_{n \to \infty} \zeta_n = 0$. Since $I((P_0, p), (Q_0, p)) = I((P_0, \tilde{p}), (Q_0, \tilde{q}))$, we have

$$\sup_{\substack{(\{P_l\}, \{Q_l\}) \in \mathcal{G}_{L_n, \rho_n} \\ I_n' \le I(\{P_l\}, \{Q_l\}) \le 2I_n}} \mathbb{P}_{(\{P_l\}, \{Q_l\})} \left(l(\hat{\sigma}(A_{L_n}), \sigma_0) > \exp\left(-(1 - \zeta_n') \frac{n}{\beta K} I(\{P_l\}, \{Q_l\}) \right) \right)$$

$$\le \sup_{\substack{((P_0, \tilde{p}), (Q_0, \tilde{q})) \\ \in \tilde{\mathcal{G}}_{\tilde{C}, \tilde{c}_1, \tilde{c}_2, r, t} \\ I_n' \le I((P_0, \tilde{p}), (Q_0, \tilde{q})) \le I_n}} \mathbb{P}_{((P_0, \tilde{p}), (Q_0, \tilde{q}))} \left(l(\hat{\sigma}(\tilde{A}), \sigma_0) > \exp\left(-(1 - \zeta_n) \frac{n}{\beta K} I((P_0, \tilde{p}), (Q_0, \tilde{q})) \right) \right)$$

$$\le \sup_{\substack{((P_0, p), (Q_0, q)) \\ \in \tilde{\mathcal{G}}_{\Phi, C, c_1, c_2, r, t} \\ I_n' \le I((P_0, p), (Q_0, q)) \le I_n}} \mathbb{P}_{((P_0, p), (Q_0, q))} \left(l(\hat{\sigma}(A), \sigma_0) > \exp\left(-(1 - \zeta_n) \frac{n}{\beta K} I((P_0, p), (Q_0, q)) \right) \right).$$

The first claim of the theorem follows immediately. The second claim can be shown in exactly the same manner.

D.2 Transformation analysis

Proposition D.1. *Let Φ be a transformation function* (5.1), $C \in [1,\infty)$, $c_1, c_2 \in (0,\infty)$, $r > 2$, *and* $t \geq \frac{2}{r}$. *Let* $((P_0,p),(Q_0,q)) \in \mathcal{G}_{\Phi,C,c_1,c_2,r,t}$, *and let* $\tilde{p}(z) := \frac{p(\Phi^{-1}(z))}{\phi(\Phi^{-1}(z))}$ *and* $\tilde{q}(z) = \frac{q(\Phi^{-1}(z))}{\phi(\Phi^{-1}(z))}$ *for* $z \in [0,1]$. *Then, with* $C_\Phi := \sup_{x \in S} \left|\frac{\phi'(x)}{\phi(x)}\right|$, *with* $\tilde{C} = C(1 + C_\Phi)$, *with* $\tilde{c}_1 = \Phi(c_1)$, *and with* $\tilde{c}_2 = 1 - \Phi(c_2)$, *we have that* $((P_0,\tilde{p}),(Q_0,\tilde{q})) \in \tilde{\mathcal{G}}_{\tilde{C},\tilde{c}_1,\tilde{c}_2,r,t}$.

Proof. Let $((P_0,p),(Q_0,q)) \in \mathcal{G}_{\phi,C,c_1,c_2,r,t}$. We show that $((P_0,\tilde{p}),(Q_0,\tilde{q})) \in \tilde{\mathcal{G}}_{\tilde{C},\tilde{c}_1,\tilde{c}_2,r,t}$ by verifying conditions C0–C5 in the definition of $\tilde{\mathcal{G}}_{\tilde{C},\tilde{c}_1,\tilde{c}_2,r,t}$ in Section 6.2. Condition C0 follows trivially from A0. It is also trivial to verify condition C1 from the definitions of $\tilde{p}$ and $\tilde{q}$. For condition C2, define $\tilde{g}(z) := g(\Phi^{-1}(z))$; the integrability conditions holds from a change of variables $x := \Phi^{-1}(z)$. For condition C3, we first note that $\tilde{\alpha} = \alpha$ and $\tilde{\gamma}(z) = \gamma(\Phi^{-1}(z))$. The integrability condition follows by a change of variable again.

For condition C4, define $\tilde{h}(z) := (1 + C_\Phi)h(\Phi^{-1}(z))$. Then

$$\left|\frac{\tilde{\gamma}'(z)}{\tilde{q}(z) + \tilde{p}(z)}\right| \leq \left|\frac{1}{\alpha}\frac{p'(\Phi^{-1}(z)) - q'(\Phi^{-1}(z))}{p(\Phi^{-1}(z)) + q(\Phi^{-1}(z))}\right|\frac{1}{\phi(\Phi^{-1}(z))} \tag{85}$$

$$+ \left|\frac{1}{\alpha}\frac{p(\Phi^{-1}(z)) - q(\Phi^{-1}(z))}{q(\Phi^{-1}(z)) + p(\Phi^{-1}(z))}\right|\left|\frac{\phi'(\Phi^{-1}(z))}{\phi(\Phi^{-1}(z))}\right|\frac{1}{\phi(\Phi^{-1}(z))} \tag{86}$$

$$\leq \left|\frac{\gamma'(\Phi^{-1}(z))}{p(\Phi^{-1}(z)) + q(\Phi^{-1}(z))}\right|\frac{1}{\phi(\Phi^{-1}(z))} + \left|\frac{\gamma(\Phi^{-1}(z))}{p(\Phi^{-1}(z)) + q(\Phi^{-1}(z))}\right|\frac{C_\Phi}{\phi(\Phi^{-1}(z))} \tag{87}$$

$$\leq (1 + C_\Phi)h(\Phi^{-1}(z)) = \tilde{h}(z). \tag{88}$$

In the same manner, we can show that $\tilde{h}(z) \geq \left|\frac{\tilde{p}'(z)}{\tilde{p}(z)}\right|, \left|\frac{\tilde{q}'(z)}{\tilde{q}(z)}\right|$. The integrability condition again follows from a change of variable. Condition C5 follows directly from A5; footnote 3 follows from footnote 1. $\qquad\square$

E Proof of Proposition 5.1

Proof. Condition A0 is satisfied by assumption. Let $\theta_1, \theta_0 \in \Theta$. Observe that for any $x \in S$, we have

$$\left|\log\frac{p(x)}{q(x)}\right| \leq |f_{\theta_1}(x) - f_{\theta_0}(x)| \leq \|\theta_1 - \theta_0\|_2 \|\nabla f_{\bar{\theta}}(x)\|_2 \leq g^*(x).$$

Thus, condition **A2** holds where we let $g = g^*$.

To verify A3, we let $\alpha = \int_S (\sqrt{p(x)} - \sqrt{q(x)})^2\, dx$. By Lemma E.1, it holds that $\alpha \geq \frac{1}{2C^*}\|\theta_0 - \theta_1\|_2^2$. By the Mean Value Theorem, there exists $\rho : S \to [0,1]$ such that

$$\int_S \left(\frac{\gamma(x)}{p(x) + q(x)}\right)^r (p(x) + q(x))\, dx = \int_S \left(\frac{(f_{\theta_1}(x) - f_{\theta_0}(x))e^{\rho(x)f_{\theta_1}(x)+(1-\rho(x))f_{\theta_0}(x)}}{\alpha(p(x) + q(x))}\right)^r (p(x) + q(x))\, dx$$

$$\leq 2C^* \int_S \sup_{\theta \in \Theta} \|\nabla f_\theta(x)\|_2^r (p(x) + q(x))\, dx$$

$$\leq 2C^{*2} \int_S \sup_{\theta \in \Theta} \|\nabla f_\theta(x)\|_2^r \phi(x)\, dx$$

$$\leq 2C^{*2} \int_S g^*(x)^r \phi(x)\, dx \leq 2C^{*3}.$$

For A4, observe that

$$\left|\frac{\gamma(x)}{p(x)+q(x)}\right| = \left|\frac{e^{f_{\theta_1}(x)} - e^{f_{\theta_0}(x)}}{\alpha(p(x)+q(x))}\right| \leq \frac{|f_{\theta_1}(x) - f_{\theta_0}(x)|}{\alpha}$$
$$\leq 2C^* \sup_{\theta \in \Theta} \|\nabla f_\theta(x)\|_2 \leq 2C^* h^*(x)\phi(x),$$

and

$$\left|\frac{\gamma'(x)}{p(x)+q(x)}\right| = \left|\frac{f'_{\theta_1}(x)p(x) - f'_{\theta_0}(x)q(x)}{\alpha(p(x)+q(x))}\right|$$
$$\leq \frac{1}{\alpha}|f'_{\theta_1}(x) - f'_{\theta_0}(x)| + |f'_{\theta_0}(x)||\gamma(x)|$$
$$\leq 2C^* \left(\sup_{\theta \in \Theta} \|\nabla f'_\theta(x)\|_2 + |f'_{\theta_0}(x)| \sup_{\theta \in \Theta} \|\nabla f_\theta(x)\|_2\right)$$
$$\leq 4C^* h^*(x)^2 \phi(x).$$

It is straightforward to verify that $\phi(x)h^*(x)^2 \geq \left|\frac{p'(x)}{p(x)}\right|, \left|\frac{q'(x)}{q(x)}\right|$, so

$$\left|\frac{\gamma'(x)}{p(x)+q(x)}\right|, \left|\frac{\gamma(x)}{p(x)+q(x)}\right|, \left|\frac{p'(x)}{p(x)}\right|, \left|\frac{q'(x)}{q(x)}\right| \leq 2C^{*3}h^*(x)^2\phi(x).$$

We then define $h(x) := 2C^{*3}h^*(x)^2$. The integrability and quasi-convexity properties hold, implying A4. Condition A5 follows trivially; footnote 1 follows from footnote 2. $\qquad\square$

E.1 Supporting lemmas

Lemma E.1. *Let $\{f_\theta\}_{\theta \in \Theta}$ be as defined in Proposition 5.1. Let $\theta_0, \theta_1 \in \Theta$ and let $p(x) = \exp(f_{\theta_1}(x))$ and $q(x) = \exp(f_{\theta_0}(x))$. Then*

$$\frac{1}{2C^*}\|\theta_1 - \theta_0\|_2^2 \leq \int_S \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx \leq C^*\|\theta_0 - \theta_1\|_2^2.$$

Proof. By the Mean Value Theorem, there exists a function $\rho: S \to [0,1]$ such that

$$\int_S \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx = \int_S e^{f_{\theta_1}(x)}\left(1 - e^{\frac{1}{2}(f_{\theta_0}(x) - f_{\theta_1}(x))}\right)^2 dx$$
$$= \int_S (f_{\theta_1}(x) - f_{\theta_0}(x))^2 e^{\rho(x)f_{\theta_0}(x) + (1-\rho(x))f_{\theta_1}(x)} dx. \qquad (89)$$

For the upper bound, we note that, from (89),

$$\int_S \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx \leq \max\left\{\int_S (f_{\theta_1}(x) - f_{\theta_0}(x))^2 e^{f_{\theta_0}(x)} dx, \int_S (f_{\theta_1}(x) - f_{\theta_0}(x))^2 e^{f_{\theta_1}(x)} dx\right\}.$$

Applying the Mean Value Theorem again, there exists a function $\rho: S \to [0,1]$ such that

$$\int_S (f_{\theta_1}(x) - f_{\theta_0}(x))^2 e^{f_{\theta_0}(x)} dx \leq \|\theta_1 - \theta_0\|^2 \int_S \|\nabla f_{\rho(x)\theta_1 + (1-\rho(x))\theta_0}(x)\|^2 e^{f_{\theta_0}(x)} dx$$
$$\leq \|\theta_1 - \theta_0\|^2 \int_S g^*(x)^2 \phi(x) dx \leq C^*\|\theta_1 - \theta_0\|_2^2.$$

We can bound $\int_S (f_{\theta_1}(x) - f_{\theta_0}(x))^2 e^{f_{\theta_0}(x)} dx$ using the same argument.

For the lower bound, we first note that, from (89),

$$\int_S \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx \geq \min\left\{\int_S (f_{\theta_1}(x) - f_{\theta_0}(x))^2 e^{f_{\theta_0}(x)} dx, \int_S (f_{\theta_1}(x) - f_{\theta_0}(x))^2 e^{f_{\theta_1}(x)} dx\right\}.$$

Define a function $\Psi : \Theta \to [0, \infty)$ by $\theta \mapsto \int_S (f_\theta(x) - f_{\theta_0}(x))^2 e^{f_{\theta_0}(x)} dx$. Then the gradient of Ψ at θ_0 is 0. Thus, by Taylor's theorem, there exists $\bar{\theta} = \rho\theta_0 + (1 - \rho)\theta_1$ for some $\rho \in [0, 1]$ such that

$$\int_S (f_{\theta_1}(x) - f_{\theta_0}(x))^2 e^{f_{\theta_0}(x)} dx$$

$$= (\theta_1 - \theta_0)^{\mathrm{T}}\left\{\int_S 2(\nabla f_{\bar{\theta}}(x))(\nabla f_{\bar{\theta}}(x))^{\mathrm{T}} e^{f_{\theta_0}(x)} dx + \int_S 2(f_{\bar{\theta}}(x) - f_{\theta_0}(x))H(f_{\bar{\theta}})(x)e^{f_{\theta_0}(x)} dx\right\}(\theta_1 - \theta_0)$$

$$\geq \|\theta_1 - \theta_0\|_2^2 \inf_{\theta \in \Theta} \lambda_{\min}\left(\int_S 2(\nabla f_\theta(x))(\nabla f_\theta(x))^{\mathrm{T}} e^{f_{\theta_0}(x)} dx\right)$$

$$\quad - \|\theta_1 - \theta_0\|_2^3 \left\{\int_S \sup_{\theta \in \Theta} \|\nabla f_\theta(x)\|^2 e^{f_{\theta_0}(x)} dx\right\}^{1/2} \left\{\sup_{\theta \in \Theta} \int_S \lambda_{\max}\left(H(f_\theta)(x)\right)^2 e^{f_{\theta_0}(x)} dx\right\}^{1/2}$$

$$\geq \|\theta_1 - \theta_0\|_2^2\left(\frac{1}{C^*} - C^*\mathrm{diam}(\Theta)\right) \geq \frac{1}{2C^*}\|\theta_1 - \theta_0\|_2^2.$$

By the same reasoning, the same lower bound holds for $\int_S (f_{\theta_1}(x) - f_{\theta_0}(x))^2 e^{f_{\theta_0}(x)} dx$. The proposition thus follows. $\qquad\square$

E.2 Proofs of examples

Proposition E.1. *Let $k \geq 2$ be an integer and let $f : \mathbb{R} \to [-\infty, \infty)$ be a k-times continuously differentiable function such that $\int_{-\infty}^{\infty} e^{f(x)} dx = 1$. Suppose that*

(a) $\sup_{x \in \mathbb{R}} |f^{(k)}(x)| < \infty$, *and*

(b) there exist $c > 0$ and $M > 0$ such that $f'(x) > M$ for $x < -c$ and $f'(x) < -M$ for $x > c$.

For any $\mu \in \mathbb{R}$ and $\sigma > 0$, define

$$f_{\mu,\sigma}(x) := f\left(\frac{x - \mu}{\sigma}\right) - \log \sigma.$$

Then there exists $C^{**} \geq 1$, $c_1, c_2 \in \mathbb{R}$, $r > 4$, and $t \in (2/r, 1/2)$, such that the following holds: For some $C_\mu > 0$ and $c_\sigma > 1$, with $\Theta := [-C_\mu, C_\mu] \times [c_\sigma^{-1}, c_\sigma]$, the family $\{f_{\mu,\sigma}\}_{(\mu,\sigma)\in\Theta}$ satisfies conditions B1–B4 in Proposition 5.1 with respect to ϕ defined in equation (3) for $S = \mathbb{R}$, and C^{**}, c_1, c_2, r, and t.

Proof. We use the notation $a \lesssim_f b$ to denote that $a \leq C_f b$ for some $C_f > 0$ possibly dependent on f. Let $r > 4$ and $t \in (2/r, 1/2)$ be fixed arbitrarily. let $\Theta_0 = [-1, 1] \times [1/2, 2]$. To verify conditions B1–B4, we will show that

$$\sup_{x \in \mathbb{R}} e^{f_{\mu,\sigma}(x)} - \phi(x) < \infty, \tag{90}$$

$$\sup_{(\mu,\sigma)\in\Theta_0} \lambda_{\min}^{-1}\left(\int_{-\infty}^{\infty} 2\nabla f_{\mu,\sigma}(x)(\nabla f_{\mu,\sigma}(x))^{\top}\phi(x) dx\right) < \infty, \tag{91}$$

$$\sup_{(\mu,\sigma)\in\Theta_0} \int_{-\infty}^{\infty} \lambda_{\max}\left(H(f_{\mu,\sigma})\right)^2 \phi(x) dx < \infty, \tag{92}$$

that there exists a quasi-convex $g^* : \mathbb{R} \to [0, \infty)$ such that $g^*(x) \geq \sup_{(\mu,\sigma)\in\Theta_0} \|\nabla f_{\mu,\sigma}(x)\|_2$ and

$$\int_{-\infty}^{\infty} g^*(x)^r \phi(x)\, dx < \infty, \tag{93}$$

and that there exists a quasi-convex $h^* : \mathbb{R} \to [0, \infty)$ such that

$$h^*(x) \geq \frac{1}{\phi(x)} \max\left\{ \sup_{(\mu,\sigma)\in\Theta_0} \|\nabla f_{\mu,\sigma}(x)\|, \quad \sup_{(\mu,\sigma)\in\Theta_0} |\nabla f'_{\mu,\sigma}(x)|, \quad \sup_{(\mu,\sigma)\in\Theta_0} |f'_{\mu,\sigma}(x)| \right\}$$

and

$$\int_{-\infty}^{\infty} h^*(x)^{2t} \phi(x)\, dx < \infty. \tag{94}$$

We may then choose C^{**} as the maximum of bounds (90), (91), (92), (93), and (94), and choose Θ as any subset of Θ_0 satisfying $\mathrm{diam}(\Theta) \leq \frac{1}{2C^{**2}}$.

We make a few observations before proving the above statements. Note that for any $x > c$, we have

$$f(x) \leq \sup_{x\in[-c,c]} f(x) + \int_c^x f'(t)dt \lesssim_f 1 - \int_c^x M dt \lesssim_f 1 - x.$$

Similarly, for any $x < -c$, it can be shown that $f(x) \lesssim_f 1 + x$. Therefore, we conclude that $f(x) \lesssim_f 1 - |x|$. Thus, for any $\mu \in [-1, 1]$ and $\sigma \in [1/2, 2]$,

$$f\left(\frac{x-\mu}{\sigma}\right) \lesssim_f 1 - \left|\frac{x-\mu}{\sigma}\right| \lesssim_f 1 - \left|\frac{x}{\sigma}\right| + \frac{\mu}{\sigma} \lesssim_f 1 - |x|, \tag{95}$$

Since $f^{(k)}(x)$ is bounded, L'Hopital's rule implies that $|f'(x)| \lesssim_f |x|^{k-1}+1$ and $|f''(x)| \lesssim_f |x|^{k-2}+1$. Moreover,

$$\left|f'\left(\frac{x-\mu}{\sigma}\right)\right| \lesssim_f \left|\frac{x-\mu}{\sigma}\right|^{k-1} + 1 \lesssim_f \left|\frac{x}{\sigma}\right|^{k-1} + \left|\frac{\mu}{\sigma}\right|^{k-1} + 1 \lesssim_f |x|^{k-1} + 1. \tag{96}$$

To verify inequality (90), observe that for any $\theta \in \Theta$ and $x \in \mathbb{R}$, we have

$$\log \phi(x) - f_{\mu,\sigma}(x) = \log\frac{e}{8} - \sqrt{|x|+1} - f\left(\frac{x-\mu}{\sigma}\right) - \log\sigma$$

$$\geq -\sqrt{|x|+1} - f\left(\frac{x-\mu}{\sigma}\right) - \log\frac{1}{c_\sigma} + \log\frac{e}{8}$$

$$\gtrsim_f -\sqrt{|x|+1} + |x| - 1.$$

Inequality (90) follows so long as we choose C^{**} such that $\log C^{**} \gtrsim_f \sup_{x\in\mathbb{R}} -|x| + \sqrt{|x|+1} + 1$.

To verify inequality (91), we first claim that

$$\sup_{(\mu,\sigma)\in\Theta} \lambda_{\max}\left(\int_S 2\nabla f_{\mu,\sigma}(x)(\nabla f_{\mu,\sigma}(x))^\top \phi(x)\, dx\right) < \infty. \tag{97}$$

Since

$$\nabla f_{\mu,\sigma}(x) = \left[\begin{array}{c} -\frac{1}{\sigma} f'\left(\frac{x-\mu}{\sigma}\right) \\ -\left(\frac{x-\mu}{\sigma^2}\right) f'\left(\frac{x-\mu}{\sigma}\right) - \frac{1}{\sigma} \end{array} \right] = -\frac{1}{\sigma} f'\left(\frac{x-\mu}{\sigma}\right) \left[\begin{array}{c} 1 \\ \frac{x-\mu}{\sigma}+1 \end{array} \right],$$

we have, for any $(\mu, \sigma) \in \Theta$, that

$$\lambda_{\max}\left(\int_{-\infty}^{\infty} 2\nabla f_\theta(x)(\nabla f_\theta(x))^\top \phi(x)\, dx\right) \leq \int_{-\infty}^{\infty} 2\|\nabla f_\theta(x)\|^2 \phi(x)\, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma^2} f'\left(\frac{x-\mu}{\sigma}\right)^2 \left(\left(\frac{x-\mu}{\sigma}+1\right)^2 + 1\right)\phi(x)\, dx \lesssim_f \int_{-\infty}^{\infty} (|x|^{2k}+1)\phi(x)\, dx.$$

Since the RHS is finite and does not depend on (μ, σ), inequality (97) holds. To verify inequality (91), we need only show that

$$\inf_{(\mu,\sigma)\in\Theta} \det\left(\int_{-\infty}^{\infty} 2(\nabla f_{\mu,\sigma}(x))(\nabla f_{\mu,\sigma}(x))^\top \phi(x)\, dx\right) > 0. \tag{98}$$

To show the bound (98), fix $(\mu, \sigma) \in \Theta$ and let $s_{\mu,\sigma}(x) = \frac{1}{\sigma^2}f'\left(\frac{x-\mu}{\sigma}\right)^2 \phi(x)$. Note that $s_{\mu,\sigma}$ is positive and integrable. Since $|f'(x)| \geq M$ for all $|x| > c$, it holds that $\inf_{(\mu,\sigma)\in\Theta}\int_{-\infty}^{\infty} s_{\mu,\sigma}(x)\, dx > 0$. Thus, $s_{\mu,\sigma}$ may be normalized to a density $\bar{s}_{\mu,\sigma}$. Fix $(\mu, \sigma) \in \Theta$. We then have

$$\det\left(\int_{-\infty}^{\infty} 2\nabla f_{\mu,\sigma}(x))(\nabla f_{\mu,\sigma}(x))^\top \phi(x)\, dx\right) \geq \int_{-\infty}^{\infty}\left(\frac{x-\mu}{\sigma}+1\right)^2 s_{\mu,\sigma}(x)\, dx - \left(\int_{-\infty}^{\infty}\left(\frac{x-\mu}{\sigma}+1\right)s_{\mu,\sigma}(x)\, dx\right)^2$$

$$\geq \operatorname{Var}_{s_{\mu,\sigma}}\left(\frac{X-\mu}{\sigma}+1\right) \gtrsim \operatorname{Var}_{s_{\mu,\sigma}}(X).$$

Since $\inf_{(\mu,\sigma)\in\Theta}\int_{-\infty}^{\infty} s_{\mu,\sigma}(x)\, dx > 0$, we have $\inf_{(\mu,\sigma)\in\Theta}\operatorname{Var}_{s_{\mu,\sigma}}(X) > 0$, which proves inequality (98) and thus also inequality (91). For inequality (92), observe that

$$H(f_{\mu,\sigma})(x) = \frac{1}{\sigma^2}f'\left(\frac{x-\mu}{\sigma}\right)\begin{pmatrix} 0 & 1 \\ 1 & 2\frac{x-\mu}{\sigma}+1 \end{pmatrix} + \frac{1}{\sigma^2}f''\left(\frac{x-\mu}{\sigma}\right)\begin{pmatrix} 1 & \frac{x-\mu}{\sigma} \\ \frac{x-\mu}{\sigma} & \frac{x-\mu}{\sigma}+\left(\frac{x-\mu}{\sigma}\right)^2 \end{pmatrix}.$$

Therefore, we have

$$\int_{-\infty}^{\infty} \lambda_{\max}\left(H(f_{\mu,\sigma})(x)\right)^2 \phi(x)\, dx \leq \int_{-\infty}^{\infty} \|H(f_{\mu,\sigma})(x)\|_F^2 \phi(x)\, dx$$

$$\lesssim_f \int_{-\infty}^{\infty} (|x|^{2k}+1)\phi(x)\, dx.$$

Since the RHS is finite and does not depend on (μ, σ), inequality (92) follows. To verify (93), note that

$$\|\nabla f_{\mu,\sigma}(x)\|_2 = \left|\frac{1}{\sigma}f'\left(\frac{x-\mu}{\sigma}\right)\right|\sqrt{1 + \left(\frac{x-\mu}{\sigma}+1\right)^2} \lesssim_f |x|^k + 1.$$

Thus, there exists $C > 0$ such that if we set $g^*(x) := C(1+|x|^k)$, then $g^*(x) \geq \sup_{(\mu,\sigma)\in\Theta}\|\nabla f_{\mu,\sigma}(x)\|_2$. The function $g^*(x)$ is quasi-convex and $\int_{-\infty}^{\infty} g^*(x)^r\phi(x)dx < \infty$, since all the moments of ϕ are finite.

To verify inequality (94), observe that

$$f'_{\mu,\sigma}(x) = \frac{1}{\sigma}f'\left(\frac{x-\mu}{\sigma}\right), \quad \text{and} \quad \nabla f'_{\mu,\sigma}(x) = \left[\begin{array}{c} -\frac{1}{\sigma^2}f''\left(\frac{x-\mu}{\sigma}\right) \\ -\frac{1}{\sigma^2}f'\left(\frac{x-\mu}{\sigma}\right) - \frac{x-\mu}{\sigma^3}f''\left(\frac{x-\mu}{\sigma}\right) \end{array}\right].$$

Therefore, $|f'_{\mu,\sigma}(x)| \lesssim 1 + |x|^{k-1}$, and

$$\|\nabla f'_{\mu,\sigma}(x)\| \leq \frac{1}{\sigma^2}\left|f''\left(\frac{x-\mu}{\sigma}\right)\right| + \frac{1}{\sigma^2}\left|f'\left(\frac{x-\mu}{\sigma}\right)\right| + \frac{1}{\sigma^2}\left|f''\left(\frac{x-\mu}{\sigma}\right)\right|\left|\frac{x-\mu}{\sigma}\right|$$

$$\lesssim_f 1 + |x|^{k-1}.$$

Thus, there exists $C > 0$ such that if we set $h^*(x) := C(1 + |x|^k)$, then

$$h^*(x) \geq \max\Big\{ \sup_{(\mu,\sigma)\in\Theta} \|\nabla f_{\mu,\sigma}(x)\|, \ \sup_{(\mu,\sigma)\in\Theta} \|\nabla f'_{\mu,\sigma}(x)\|, \ \sup_{(\mu,\sigma)\in\Theta} |f'_{\mu,\sigma}(x)| \Big\}.$$

It is clear that h^* is quasi-convex and $\int_S h^*(x)^{2t} \phi(x)\, dx < \infty$.

Finally, to verify condition B5 of Proposition 5.1, observe that

$$(\log \phi)'(x) = \begin{cases} \frac{1}{2} \frac{1}{\sqrt{1-x}}, & \text{if } x < 0, \\ -\frac{1}{2} \frac{1}{\sqrt{1+x}}, & \text{if } x > 0. \end{cases}$$

In particular, $(\log \phi)'(x) \to 0$ as $|x| \to \infty$. Since $f'(x) \geq M$ for all $x \leq -c$ and $f'(x) \leq -M$ for all $x \geq c$, if $x \leq -\frac{c}{c_\sigma} - C_\mu$, then $\frac{x-\mu}{\sigma} \leq -c$ and

$$f'_{\mu,\sigma}(x) = \frac{1}{\sigma} f'\left(\frac{x-\mu}{\sigma} \right) \geq \frac{M}{c_\sigma}.$$

If $x \geq \frac{c}{c_\sigma} + C_\mu$, then $\frac{x-\mu}{\sigma} \geq c$ and

$$f'_{\mu,\sigma}(x) = \frac{1}{\sigma} f'\left(\frac{x-\mu}{\sigma} \right) \leq -\frac{M}{c_\sigma}.$$

Thus, there exist $c_1 < 0$ and $c_2 > 0$ such that B5 holds. $\qquad\square$

Proposition E.2. *Let $k \geq 2$ be an integer and let $f : [0,\infty) \to [-\infty,\infty)$ be a k-times continuously differentiable function such that $\int_0^\infty e^{f(x)}\, dx = 1$. Suppose*

(a) $\sup_{x\in\mathbb{R}} |f^{(k)}(x)| < \infty$, *and*

(b) *there exist $c > 0$ and $M > 0$ such that $f'(x) < -M$ for $x > c$.*

For any $\sigma > 0$, define

$$f_\sigma(x) := f\left(\frac{x}{\sigma} \right) - \log \sigma.$$

Then there exist $C^{**} \geq 1$, $c_1, c_2 \in (0,\infty)$, $r > 4$, and $t \in (2/r, 1/2)$ such that the following holds: For some $c_\sigma > 1$, with $\Theta := [c_\sigma^{-1}, c_\sigma]$, the family $\{f_\sigma\}_{(\sigma)\in\Theta}$ satisfies conditions B1–B4 in Proposition 5.1 with respect to ϕ defined in equation (3) for $S = [0,\infty)$, and C^{**}, c_1, c_2, r, and t.

The proof is almost identical to that of Proposition E.1. We note that the g^* function that arises from Proposition E.2 is non-decreasing on $[0,\infty)$. Hence, by footnote 2, we need only verify the right side condition of **B5**.

Proposition E.3. *For any $\alpha > 1$, $\beta > 0$, define $f_{\alpha,\beta}(x) = (\alpha - 1)\log x - \beta x + \alpha \log \beta - \log \Gamma(\alpha)$ for $x \in (0,\infty)$ and $f_{\alpha,\beta}(0) = -\infty$. Then there exist $C^{**} \geq 1$, $c_1, c_2 \in (0,\infty)$, $r > 4$, and $t \in (2/r, 1/2)$ such that the following holds: For some $C_\alpha > 0$ and $C_\beta > 1$, with $\Theta := [2 - C_\alpha, 2 + C_\alpha] \times [C_\beta^{-1}, C_\beta]$, the family $\{f_{\alpha,\beta}\}_{(\alpha,\beta)\in\Theta}$ satisfies conditions B1–B5 in Proposition 5.1 with respect to ϕ defined in equation (3) for $S = [0,\infty)$, and C^{**}, c_1, c_2, r, and t.*

Proof. As in the proof of Proposition E.1, we first define $\Theta_0 = [2,4] \times [1/2, 2]$ and show that the analogous versions of inequalities (90), (91), (92), (93), and (94) hold. The arguments are almost identical, so we only highlight some key points.

Verifying the analog of inequality (90) is straightforward. For inequalities (91) and (92), observe that

$$\nabla f_{\alpha,\beta}(x) = \begin{bmatrix} \log x + \log \beta - d_\alpha \log \Gamma(\alpha) \\ -x + \frac{\alpha}{\beta} \end{bmatrix},$$

$$H f_{\alpha,\beta}(x) = \begin{bmatrix} d_\alpha^2 \log \Gamma(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & -\frac{\alpha}{\beta^2} \end{bmatrix}.$$

The argument is then straightfoward using the same technique as in the proof of Proposition E.1.

Since

$$\|\nabla f_{\alpha,\beta}(x)\|_2 \le |\log x| + x + |\log \beta| + |d_\alpha \log \Gamma(\alpha)| + \frac{\alpha}{\beta}$$

$$\lesssim |\log x| + x + 1,$$

there exists $C' > 0$ such that if we let $g^*(x) = C'(|\log x| + x + 1)$, then $g^*(x) \ge \sup_{\theta \in \Theta_0} \|\nabla f_\theta(x)\|_2$ and

$$\int g^*(x)^r \phi(x) dx = \int_0^\infty C'(|\log x| + x + 1)^r \phi(x) dx$$

$$\lesssim \int_0^\infty |\log x|^r \phi(x) + \int_0^\infty x^r \phi(x) dx < \infty,$$

thus verifying inequality (93). Moreover,

$$f'_{\alpha,\beta}(x) = \frac{(\alpha - 1)}{x} - \beta, \quad \text{and} \quad \nabla f'_{\alpha,\beta}(x) = \begin{bmatrix} \frac{1}{x} \\ -1 \end{bmatrix}.$$

We thus conclude that $|f'_{\alpha,\beta}(x)| \lesssim 1 + x^{-1}$ and $\|\nabla f'_{\alpha,\beta}(x)\|_2 \lesssim 1 + x^{-1}$. Hence, there exists $C' > 0$ such that $h^*(x) := C'(1 + x^{-1} + x)$ satisfies

$$h^*(x) \ge \max\{ \sup_{(\alpha,\beta) \in \Theta_0} \|\nabla f_{\alpha,\beta}(x)\|_2, \sup_{(\alpha,\beta) \in \Theta_0} \|\nabla f'_{\alpha,\beta}(x)\|_2, \sup_{(\alpha,\beta) \in \Theta_0} |f'_{\alpha,\beta}(x)|\}.$$

Note that h^* is clearly quasi-convex and

$$\int_0^\infty h^*(x)^{2t} \phi(x) dx \lesssim 1 + \int_0^\infty x^{-2t} + x^{2t} \phi(x) dx.$$

Since $2t < 1$ by assumption, the integral converges.

Condition B5 follows since $\alpha > 1$ for all $(\alpha, \beta) \in \Theta_0$. $\qquad\square$

F A discussion of networks with discrete weights

Let P and Q be distributions on $\mathbb{N}$, and suppose

$$A \sim \begin{cases} P, & \text{if } \sigma_0(u) = \sigma_0(v), \\ Q, & \text{if } \sigma_0(u) \ne \sigma_0(v). \end{cases}$$

We propose a crude truncation scheme to adapt the algorithm described in Section 4 for this setting: for an input parameter $L \in \mathbb{N}$, we collect all edges with weight greater than L, and assign them a new weight of L. In other words, we create a new matrix A_L, where if $A_{uv} \in [L]$, then $A_{L,uv} = A_{u,v}$; and if $A_{uv} \in \{L+1, L+2, \ldots\}$, we set $A_{L,uv} = L$. In this way, A_L is a network with L labels. We then proceed with initialization (Algorithm 4.3) and refinement (Algorithm 4.5).

From Proposition 6.1, we know that the truncation scheme achieves a rate of $\exp(-\frac{n}{\beta K} I(P,Q)(1 + o(1))$ if the following two criteria are met. For convenience, let P' and Q' represent distributions on $[L]$ that result from truncating P and Q at L.

1. There exists $\rho_n > 1$ such that $\rho_n^{-1} \le \frac{P'_l}{Q'_l} \le \rho_n$ for all $l \in [L]$, and $\frac{nI'}{(L_n+1)\rho_n^2 \log \rho_n} \to \infty$.

2. $\lim_{L_n \to \infty} |I(P', Q')/I(P, Q) - 1| \to 0$.

To show that these criteria are not unrealistic, we argue informally that the P and Q were members of the Poisson family $\{\text{Poi}(\lambda)\}_{\lambda \in [1,2]}$, then they satisfy both criteria if L_n is chosen so that $L_n = o(\log\log(nI')$ and $L_n \to \infty$ as $n \to \infty$. To be precise, let $a \in (0, 1]$, let $(\lambda_1, \lambda_0) \in [1, 2]$, and for $l > 0$, let $P_l = a\frac{\lambda_1^l e^{-\lambda_1}}{l!}$ and $Q_l = a\frac{\lambda_0^l e^{-\lambda_0}}{l!}$. Let $P_0 = (1-a) + ae^{-\lambda_1}$ and $Q_0 = (1-a) + ae^{-\lambda_0}$. Note that we assume a common sparsity factor a for simplicity.

It is then straightforward to verify that $\rho_n \lesssim 2^{L_n}$ satisfies the first criterion. For the second criterion, we assume that $I(P, Q) \leq 1$ and that $a = 0$; the argument for when $a > 0$ is exactly the same. Observe that

$$\frac{I(P', Q') - I(P, Q)}{I(P, Q)} \lesssim \frac{1}{(\lambda_1 - \lambda_0)^2} \sum_{l=L}^{\infty} (\sqrt{P_l} - \sqrt{Q_l})^2$$

$$\lesssim \frac{1}{(\lambda_1 - \lambda_0)^2} \sum_{l=L}^{\infty} \left(\frac{\lambda_1^l e^{-\lambda_1}}{l!} - \frac{\lambda_0^l e^{-\lambda_0}}{l!} \right)^2 \frac{l!}{\lambda_0^l e^{-\lambda_0}}$$

$$\leq \frac{1}{(\lambda_1 - \lambda_0)^2} \sum_{l=L}^{\infty} \left(\left(\frac{\lambda_1}{\lambda_0}\right)^l e^{\lambda_0 - \lambda_1} - 1 \right)^2 \frac{\lambda_0^l e^{-\lambda_0}}{l!} \to 0,$$

as $L \to \infty$, as required. Our information-theoretic lower bound can be extended to this setting, as well. However, this truncation scheme may not be sensible for other types of discrete distributions; we leave a thorough investigation of networks with discrete weights for future work.

G   Appendix for Theorem 5.2

We begin by defining some notation. Let $\sigma, \sigma_0 : [n] \to [K]$ be two clusterings. Let

$$S_K[\sigma, \sigma_0] := \underset{\pi \in S_K}{\arg\min}\, d_H(\pi \circ \sigma, \sigma_0), \tag{99}$$

where $d_H(\cdot, \cdot)$ denotes the Hamming distance, and define

$$\mathcal{E}[\sigma, \sigma_0] := \Big\{ v : \pi(\sigma(v)) \neq \sigma_0(v), \text{ for some } \pi \in S_K[\sigma, \sigma_0] \Big\}. \tag{100}$$

When $S_K[\sigma, \sigma_0]$ is a singleton, the set $\mathcal{E}[\sigma, \sigma_0]$ contains all nodes misclustered by σ in relation to σ_0. When $S_K[\sigma, \sigma_0]$ contains multiple elements, we continue to call $\mathcal{E}[\sigma, \sigma_0]$ the set of *misclustered* nodes.

G.1   Proof of Theorem 5.2

Let σ_0 be defined as in the statement of Theorem 5.2. For any clustering σ, we define

$$\tilde{l}(\sigma, \sigma_0) := \frac{1}{n} \sum_{v=1}^{n} \mathbf{1}\{v \in \mathcal{E}[\sigma, \sigma_0]\},$$

where $\mathcal{E}[\sigma, \sigma_0]$ is defined in equation (100). In particular, note that if $|S_K[\hat{\sigma}(A), \sigma_0]| = 1$, we have $\tilde{l} = l$. We claim the following:

Claim 1: There exists a sequence of real numbers $\zeta_n \to 0$ and $c \in (0, \infty)$ such that, for any permutation equivariant estimator $\hat{\sigma}$,

$$\inf_{\substack{((P_0, p), (Q_0, q)) \in \mathcal{G}^* \\ I'_n \leq I((P_0, p), (Q_0, q)) \leq C}} \mathbb{E}_{\substack{(P_0, p) \\ (Q_0, q)}} \Big[\tilde{l}(\hat{\sigma}(A), \sigma_0) \Big] \exp\left((1 + \zeta_n)\frac{n}{\beta K} I((P_0, p), (Q_0, q)) \right) \geq c. \tag{101}$$

Claim 2: For any $c > 0$, there exists $c' > 0$ such that, for any permutation equivariant estimator $\hat{\sigma}$,

$$\inf_{\substack{((P_0,p),(Q_0,q))\in\mathcal{G}^* \\ I((P_0,p),(Q_0,q))\leq c/n}} \mathbb{E}_{\substack{(P_0,p) \\ (Q_0,q)}} \left[\tilde{l}(\hat{\sigma}(A),\sigma_0)\right] \geq c'. \tag{102}$$

We first prove the theorem from the claims. Let $((P_0,p),(Q_0,p)) \in \mathcal{G}^*$, and let $\hat{\sigma}$ be an arbitrary clustering algorithm. We omit the $((P_0,p),(Q_0,q))$ subscript on the expectation and probability from this point on for simplicity of presentation. Let $\hat{\sigma}$ be an arbitrary permutation equivariant estimator. If $\mathbb{P}\left(l(\hat{\sigma}(A),\sigma_0) \geq \frac{1}{2\beta K}\right) \geq \frac{1}{2}\mathbb{E}\tilde{l}(\hat{\sigma}(A),\sigma_0)$, then by Markov's inequality, we have

$$\mathbb{E}l(\hat{\sigma}(A),\sigma_0) \geq \frac{1}{2\beta K}\mathbb{P}\left(l(\hat{\sigma}(A),\sigma_0) \geq \frac{1}{2\beta K}\right) \geq \frac{1}{4\beta K}\mathbb{E}\tilde{l}(\hat{\sigma}(A),\sigma_0).$$

On the other hand, if $\mathbb{P}\left(l(\hat{\sigma}(A),\sigma_0) \geq \frac{1}{2\beta K}\right) < \frac{1}{2}\mathbb{E}\tilde{l}(\hat{\sigma}(A),\sigma_0)$, we have

$$\mathbb{E}l(\hat{\sigma}(A),\sigma_0) \geq \mathbb{E}\left[l(\hat{\sigma}(A),\sigma_0) \,\Big|\, l(\hat{\sigma}(A),\sigma_0) < \frac{1}{2\beta K}\right]\mathbb{P}\left(l(\hat{\sigma}(A),\sigma_0) < \frac{1}{2\beta K}\right)$$

$$\overset{(a)}{=} \mathbb{E}\left[\tilde{l}(\hat{\sigma}(A),\sigma_0) \,\Big|\, l(\hat{\sigma}(A),\sigma_0) < \frac{1}{2\beta K}\right]\mathbb{P}\left(l(\hat{\sigma}(A),\sigma_0) < \frac{1}{2\beta K}\right)$$

$$= \mathbb{E}\tilde{l}(\hat{\sigma}(A),\sigma_0) - \mathbb{E}\left[\tilde{l}(\hat{\sigma}(A),\sigma_0) \,\Big|\, l(\hat{\sigma}(A),\sigma_0) \geq \frac{1}{2\beta K}\right]\mathbb{P}\left(l(\hat{\sigma}(A),\sigma_0) \geq \frac{1}{2\beta K}\right)$$

$$\geq \mathbb{E}\tilde{l}(\hat{\sigma}(A),\sigma_0) - \frac{1}{2}\mathbb{E}\tilde{l}(\hat{\sigma}(A),\sigma_0) = \frac{1}{2}\mathbb{E}\tilde{l}(\hat{\sigma}(A),\sigma_0),$$

where (a) holds by Lemma B.6. Hence, we have that, in all cases, $\mathbb{E}l(\hat{\sigma}(A),\sigma_0) \geq \frac{1}{4\beta K}\mathbb{E}\tilde{l}(\hat{\sigma}(A),\sigma_0)$.

We now focus on proving the claims. Let us arbitrarily fix a permutation equivariant algorithm $\hat{\sigma}$, fix $((P_0,p),(Q_0,q)) \in \mathcal{G}^*$, and suppose $I((P_0,p),(Q_0,q)) \leq 2\log 2$. Without loss of generality, suppose cluster 1 has size $\frac{n}{\beta K} + 1$ and cluster 2 has size $\frac{n}{\beta K}$. Let $C_k := \{u \in [n] : \sigma_0(u) = k\}$ denote the kth cluster. We also suppose without loss of generality that $C_1 = \{1, 2, ..., \frac{n}{\beta K} + 1\}$ and $C_2 = \{\frac{n}{\beta K} + 2, ..., 2\frac{n}{\beta K} + 1\}$.

Define $\sigma_0^1 := \sigma_0$ and define $\sigma_0^2 : [n] \to [K]$ such that $\sigma_0^2(v) = \sigma_0(v)$ for all $v \neq 1$, and $\sigma_0^2(1) = 2$. Define σ^* as a random cluster assignment with the distribution

$$\sigma^* := \begin{cases} \sigma_0^1, & \text{with probability } \frac{1}{2}, \\ \sigma_0^2, & \text{with probability } \frac{1}{2}. \end{cases} \tag{103}$$

We note that $\sigma^*(u) = \sigma_0(u)$ for all $u \neq 1$ and $\sigma^*(1)$ is either 1 or 2, each with $\frac{1}{2}$ probability.

Let $\mathbb{P}_\Phi$ denote a probability measure on $\{\sigma_0^1, \sigma_0^2\} \times \left(\mathbb{R}^{\frac{n(n-1)}{2}}, \mathcal{B}(\mathbb{R}^{\frac{n(n-1)}{2}})\right)$—where $\{\sigma_0^1, \sigma_0^2\}$ is by default equipped by the discrete σ-algebra—defined by

$$\mathbb{P}_\Phi(\sigma^*, \mathbf{A}) = \mathbb{P}(\sigma^*)\mathbb{P}_{SBM}(\mathbf{A} \,|\, \sigma^*),$$

for a Borel-measurable set $\mathbf{A} \in \mathcal{B}(\mathbb{R}^{\frac{n(n-1)}{2}})$, where $\mathbb{P}_{SBM}(\cdot \,|\, \sigma^*)$ is the probability measure $\left(\mathbb{R}^{\frac{n(n-1)}{2}}, \mathcal{B}(\mathbb{R}^{\frac{n(n}{2}}\right.$ defined by the weighted SBM with respect to $((P_0,p),(Q_0,q))$, treating σ^* as the true cluster assignment. Let Ψ denote an alternative probability measure on $\{\sigma_0^1, \sigma_0^2\} \times (\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$ defined by

$$\mathbb{P}_\Psi(\sigma^*, \mathbf{A}) = P(\sigma^*)\mathbb{P}_\Psi(\mathbf{A} \,|\, \sigma^*),$$

where $\mathbb{P}_\Psi(\mathbf{A} \,|\, \sigma^*)$ is a product of $\frac{n(n-1)}{2}$ distributions over $\mathbb{R}$, defined as follows: if A is a random upper-triangular matrix taking value in $\mathbb{R}^{\frac{n(n-1)}{2}}$, whose distribution is $\mathbb{P}_\Psi(\cdot \,|\, \sigma^*)$, then

1. for $(u,v) \in [n]^2$ where $u < u$, if $u \neq 1$ and $v \neq 1$, then A_{uv} is distributed as in $\mathbb{P}_{SBM}(\cdot \,|\, \sigma^*)$;

2. if $u = 1$ and $v \notin C_1 \cup C_2$, then A_{uv} is distributed as in $P_{SBM}(\cdot \,|\, \sigma^*)$;

3. if $u = 1$ and $v \in C_1 \cup C_2$, then A_{uv} is distributed as Y, where Y is the distribution on $\mathbb{R}$ that minimizes $D((P_0, p), (Q_0, q))$ in Lemma G.2; i.e., $Y_0 \propto (P_0 Q_0)^{1/2}$ and $(1 - Y_0)y(x) \propto \sqrt{(1 - P_0)p(x)(1 - Q_0)q(x)}$.

Note that $P_\Psi(\cdot \,|\, \sigma^*)$ does not actually depend on whether $\sigma^* = \sigma_0^1$ or σ_0^2. Since Y is absolutely continuous with respect to P and Q, the distribution $\mathbb{P}_\Psi$ is absolutely continuous with respect to $\mathbb{P}_\Phi$. Thus, we may define a Borel measurable function $\mathcal{Q} : \{\sigma_0^1, \sigma_0^2\} \times \mathbb{R}^{\frac{n(n-1)}{2}} \to [0, \infty)$ as

$$\mathcal{Q}(\sigma^*, A) := \log \frac{d\mathbb{P}_\Psi}{d\mathbb{P}_\Phi}(\sigma^*, A) = \log \frac{d\mathbb{P}_{SBM}(A \,|\, \sigma^*)}{d\mathbb{P}_\Psi(A \,|\, \sigma^*)}.$$

For $\sigma^* \in \{\sigma_0^1, \sigma_0^2\}$ and $A \in \mathbb{R}^{n \times n}$, we denote $P(A_{u,1}) := P_0$ if $A_{u,1} = 0$, and $P(A_{u,1}) := (1 - P_0)p(A_{u,1})$ if $A_{u,1} \neq 0$, and similarly for Q and Y. Then, from the fact that $\mathbb{P}_{SBM}(\cdot \,|\, \sigma^*)$ and $\mathbb{P}_\Psi(\cdot \,|\, \sigma^*)$ are both product measures, we have

$$\mathcal{Q}(\sigma^*, A) = \sum_{\substack{u \in C_{\sigma^*(1)} \\ u \neq 1}} \log \frac{Y(A_{1,u})}{P(A_{1,u})} + \sum_{u \in C_1 \cup C_2 \setminus C_{\sigma^*(1)}} \log \frac{Y(A_{1,u})}{Q(A_{1,u})}. \tag{104}$$

We now define the event

$$E = \left\{ \{1\} \notin \mathcal{E}[\hat{\sigma}(A), \sigma^*] \text{ and } \tilde{l}(\hat{\sigma}(A), \sigma^*) \leq \frac{1}{4\beta K} \right\}.$$

Let $t := -\log(20\beta K \mathbb{E}_\Phi \tilde{l}(\hat{\sigma}(A), \sigma^*))$. Then

$$\mathbb{P}_\Psi\big(\mathcal{Q}(\sigma^*, A) \leq t\big) = \mathbb{P}_\Psi\big(\mathcal{Q}(\sigma^*, A) \leq t, E^c\big) + \mathbb{P}_\Psi\big(\mathcal{Q}(\sigma^*, A) \leq t, E\big). \tag{105}$$

We bound the first term of inequality (105) as follows:

$$
\begin{aligned}
\mathbb{P}_\Psi(\mathcal{Q}(\sigma^*, A) \leq t, E^c) &= \int_{\{\mathcal{Q} \leq t\} \cap E^c} d\mathbb{P}_\Psi = \int_{\{\mathcal{Q} \leq t\} \cap E^c} \exp(\mathcal{Q}) d\mathbb{P}_\Phi \\
&\leq e^t \mathbb{P}_\Phi\big(\mathcal{Q}(\sigma^*, A) \leq t, E^c\big) \leq e^t \mathbb{P}_\Phi(E^c) \\
&\leq e^t \left(\mathbb{P}_\Phi(\{1\} \in \mathcal{E}[\hat{\sigma}(A), \sigma^*]) + \mathbb{P}_\Phi\left(\tilde{l}(\hat{\sigma}(A), \sigma^*) \geq \frac{1}{4\beta K} \right) \right).
\end{aligned} \tag{106}
$$

Furthermore,

$$
\begin{aligned}
\mathbb{E}_\Phi \tilde{l}(\hat{\sigma}(A), \sigma^*) &= \frac{1}{n} \sum_{v=1}^n P_\Phi(v \in \mathcal{E}[\hat{\sigma}(A), \sigma^*]) \geq \frac{1}{n} \sum_{v \in C_{\sigma^*(1)}} P_\Phi(v \in \mathcal{E}[\hat{\sigma}(A), \sigma^*]) \\
&\overset{(a)}{=} \frac{|C_{\sigma^*(1)}|}{n} P_\Phi(1 \in \mathcal{E}[\hat{\sigma}(A), \sigma^*]) \geq \frac{1}{\beta K} P_\Phi(1 \in \mathcal{E}[\hat{\sigma}(A), \sigma^*]),
\end{aligned} \tag{107}
$$

where (a) follows from Corollary G.1, and

$$
\begin{aligned}
\mathbb{E}_\Phi \tilde{l}(\hat{\sigma}(A), \sigma^*) &\geq \mathbb{E}_\Phi\left[\tilde{l}(\hat{\sigma}(A), \sigma^*) \,\Big|\, \tilde{l}(\hat{\sigma}(A), \sigma^*) \geq \frac{1}{4\beta K} \right] P_\Phi\left(\tilde{l}(\hat{\sigma}(A), \sigma^*) \geq \frac{1}{4\beta K} \right) \\
&\geq \frac{1}{4\beta K} P_\Phi\left(\tilde{l}(\hat{\sigma}(A), \sigma^*) \geq \frac{1}{4\beta K} \right).
\end{aligned} \tag{108}
$$

Hence, combining inequalities (106), (113), (114), we obtain

$$P_\Psi(\mathcal{Q} \leq t, E^c) \leq e^t 5\beta K \mathbb{E}_\Phi\big[\tilde{l}(\hat{\sigma}(A), \sigma^*) \big].$$

64

We now turn to the second term in equation (105). We have

$$P_\Psi(E) = \frac{1}{2} P_\Psi \left(1 \notin \mathcal{E}[\hat{\sigma}(A), \sigma_0^1] \text{ and } \tilde{l}(\hat{\sigma}(A), \sigma_0^1) \leq \frac{1}{4\beta K} \right)$$

$$+ \frac{1}{2} P_\Psi \left(1 \notin \mathcal{E}[\hat{\sigma}(A), \sigma_0^2] \text{ and } \tilde{l}(\hat{\sigma}(A), \sigma_0^2) \leq \frac{1}{4\beta K} \right). \quad (109)$$

In the event that $\tilde{l}(\hat{\sigma}(A), \sigma_0^1) \leq \frac{1}{4\beta K}$, we know, by the fact that $l(\hat{\sigma}(A), \sigma_0^1) \leq \tilde{l}(\hat{\sigma}(A), \sigma_0^1)$ and Lemma B.6, that $S_K[\hat{\sigma}(A), \sigma_0^1]$ contains only one element, which we denote by π. Since $d_H(\sigma_0^1, \sigma_0^2) = 1$, we have $\frac{1}{n} d_H(\pi \circ \hat{\sigma}(A), \sigma_0^2) \leq \frac{1}{4\beta K} + \frac{1}{n} \leq \frac{1}{2\beta K}$, so we may apply Lemma B.6 again to conclude that π is the only element of $S_K[\hat{\sigma}(A), \sigma_0^2]$, as well. However, since $\sigma_0^1(1) = 1$ and $\sigma_0^2(1) = 2$, we know that $(\pi \circ \hat{\sigma}(A))(1)$ cannot be equal to both $\sigma_0^1(1)$ and $\sigma_0^2(1)$. Hence, we cannot simultaneously have $1 \notin \mathcal{E}[\hat{\sigma}(A), \sigma_0^1]$ and $1 \notin \mathcal{E}[\hat{\sigma}(A), \sigma_0^2]$. Thus, the two events in equation (109) are disjoint, so $P_\Psi(\mathcal{Q} \leq t, E) \leq P_\Psi(E) \leq \frac{1}{2}$. Plugging back into equation (105), we conclude that

$$P_\Psi(\mathcal{Q} \leq t) \leq e^t 5\beta K \mathbb{E}_\Phi \tilde{l}(\hat{\sigma}(A), \sigma^*) + \frac{1}{2} \overset{(a)}{\leq} \frac{3}{4},$$

where (a) follows from the definition that $t = -\log(20\beta K \mathbb{E}_\Phi \tilde{l}(\hat{\sigma}(A), \sigma^*))$. By Chebyshev's inequality, we also have

$$P_\Psi \left(\mathcal{Q} \leq \mathbb{E}_\Psi \mathcal{Q} + \sqrt{5 V_\Psi(\mathcal{Q})} \right) \geq 4/5,$$

where $V_\Psi(Q) := \mathrm{Var}_\Psi \big(\mathcal{Q}(\sigma^*, A) \big)$. Hence, $\log \frac{1}{20\beta K \mathbb{E}_\Phi \tilde{l}(\hat{\sigma}(A), \sigma_0)} \leq \mathbb{E}_\Psi \mathcal{Q} + \sqrt{5 V_\Psi(\mathcal{Q})}$, or equivalently,

$$\mathbb{E}_\Phi \tilde{l}(\hat{\sigma}(A), \sigma^*) \geq \frac{1}{20\beta K} \exp \left(- (\mathbb{E}_\Psi \mathcal{Q} + \sqrt{5 V_\Psi(\mathcal{Q})}) \right). \quad (110)$$

We now compute $\mathbb{E}_\Psi \mathcal{Q}$ and $V_\Psi(\mathcal{Q})$. Note that

$$\mathbb{E}_\Psi \mathcal{Q} = \frac{1}{2} \mathbb{E}_\Psi[\mathcal{Q} \,|\, \sigma^* = \sigma_0^1] + \frac{1}{2} \mathbb{E}_\Psi[\mathcal{Q} \,|\, \sigma^* = \sigma_0^2].$$

By Lemma G.2, we have

$$\mathbb{E}_\Psi[\mathcal{Q} \,|\, \sigma^* = \sigma_0^1] = \mathbb{E}_\Psi \left[\sum_{u: u \neq 1, \, \sigma_0^1(u)=1} \log \frac{Y(A_{1,u})}{P(A_{1,u})} + \sum_{u: \sigma_0^1(u)=2} \log \frac{Y(A_{1,u})}{Q(A_{1,u})} \right]$$

$$= \frac{n}{\beta K} \int \log \frac{dY}{dP} dY + \frac{n}{\beta K} \int \log \frac{dY}{dQ} dY$$

$$= \frac{n}{\beta K} 2D((P_0, p), (Q_0, q)) = \frac{n}{\beta K} I((P_0, p), (Q_0, q)).$$

Similarly, we have $\mathbb{E}_\Psi[\mathcal{Q} \,|\, \sigma^* = \sigma_0^2] = \frac{n}{\beta K} I((P_0, p), (Q_0, q))$, so

$$\mathbb{E}_\Psi \mathcal{Q} = \frac{n}{\beta k} I((P_0, p), (Q_0, q)). \quad (111)$$

By Lemma G.3, we also have

$$\sqrt{5 V_\Psi(\mathcal{Q})} \leq 20 \sqrt{\frac{n}{\beta K} I((P_0, p), (Q_0, q))}. \quad (112)$$

Let us define $\zeta_n := 30\big(\frac{n}{\beta K}I((P_0,p),(Q_0,q))\big)^{-1/2}$. Combining inequalities (110), (111), and (112), we then have

$$\mathbb{E}_\Phi \tilde{l}(\hat{\sigma}(A),\sigma^*) \geq \frac{1}{20\beta K}\exp\left(-\frac{n}{\beta K}I((P_0,p),(Q_0,q)) - 30\left(\frac{n}{\beta K}I((P_0,p),(Q_0,q))\right)^{1/2}\right) \quad (113)$$

$$\geq \frac{1}{20\beta K}\exp\left(-(1-\zeta_n)\frac{n}{\beta K}I((P_0,p),(Q_0,q))\right). \quad (114)$$

Now suppose $I((P_0,p),(Q_0,q)) \geq I'_n$. Since $\zeta_n \leq 30\big(\frac{n}{\beta K}I'_n\big)^{-1/2}$, we have $\zeta_n \to 0$. On the other hand, fix $c > 0$ and suppose $I((P_0,p),(Q_0,q)) \leq \frac{c}{n}$. Then from inequality (113), we have

$$\mathbb{E}_\Phi \tilde{l}(\hat{\sigma}(A),\sigma^*) \geq \frac{1}{20\beta K}\exp\left(-\frac{c}{\beta K} - 30\sqrt{\frac{c}{\beta K}}\right) := c'.$$

To finish the proof of the two claims listed at the beginning, we need only show that the same lower bound holds for $\mathbb{E}\tilde{l}(\hat{\sigma}(A),\sigma_0)$.

Define two probability measures P_1 and P_2 on $(A,\hat{\sigma}(A))$, as follows:

$$P_1(A,\hat{\sigma}(A)) = P_{SBM}(A\,|\,\sigma_0^1)P_{alg}(\hat{\sigma}(A)\,|\,A),$$
$$P_2(A,\hat{\sigma}(A)) = P_{SBM}(A\,|\,\sigma_0^2)P_{alg}(\hat{\sigma}(A)\,|\,A).$$

Note that $\mathbb{E}_\Phi[\tilde{l}(\hat{\sigma}(A),\sigma^*)\,|\,\sigma^* = \sigma_0^1] = \mathbb{E}_1\tilde{l}(\hat{\sigma}(A),\sigma_0^1)$ and $\mathbb{E}_\Phi[\tilde{l}(\hat{\sigma}(A),\sigma^*)\,|\,\sigma^* = \sigma_0^2] = \mathbb{E}_2\tilde{l}(\hat{\sigma}(A),\sigma_0^2)$, where $\mathbb{E}_1$ and $\mathbb{E}_2$ are expectations taken with respect to P_1 and P_2, respectively. We claim that $\mathbb{E}_1\tilde{l}(\hat{\sigma}(A),\sigma_0^1) = \mathbb{E}_2\tilde{l}(\hat{\sigma}(A),\sigma_0^2)$, in which case $\mathbb{E}_\Phi l(\hat{\sigma}(A),\sigma^*) = \mathbb{E}_1\tilde{l}(\hat{\sigma}(A),\sigma_0^1) = \mathbb{E}\tilde{l}(\hat{\sigma}(A),\sigma_0)$ and the claims follow.

Define a permutation $\pi \in S_n$ that swaps $\{2,\ldots,\frac{n}{\beta K}+1\}$ with $\{\frac{n}{\beta K}+2,\ldots,2\frac{n}{\beta K}+1\}$ and satisfies $\pi(u) = u$ for $u = 1$ and $u \geq 2\frac{n}{\beta K}+2$. Clearly, $\sigma_0^2 = \tau \circ \sigma_0^1 \circ \pi^{-1}$, where $\tau \in S_K$ swaps cluster labels 1 and 2. Now let A be fixed and let $\rho \in S_K$ be arbitrary. We have

$$d_H(\rho \circ \hat{\sigma}(A),\sigma_0^1) = d_H(\rho \circ \hat{\sigma}(A),\tau^{-1}\circ\sigma_0^2\circ\pi)$$
$$= d_H(\rho\circ\hat{\sigma}(A)\circ\pi^{-1},\tau^{-1}\circ\sigma_0^2)$$
$$\stackrel{(a)}{=} d_H(\rho\circ\xi^{-1}\circ\hat{\sigma}(\pi A),\tau^{-1}\circ\sigma_0^2)$$
$$= d_H(\tau\circ\rho\circ\xi^{-1}\circ\hat{\sigma}(\pi A),\sigma_0^2),$$

where (a) follows because $\hat{\sigma}(\cdot)$ is permutation equivariant, which implies that there exists $\xi \in S_K$ such that $\hat{\sigma}(A)\circ\pi^{-1} = \xi^{-1}\circ\hat{\sigma}(\pi A)$. Thus, $\rho \mapsto \tau\circ\rho\circ\xi^{-1}$ is a bijection between $S_K[\hat{\sigma}(A),\sigma_0^1]$ and $S_K[\hat{\sigma}(\pi A),\sigma_0^2]$. Furthermore, if v satisfies $(\rho\circ\hat{\sigma}(A))(v) \neq \sigma_0^1(v)$, then letting $u = \pi(v)$, we equivalently have

$$(\rho\circ\hat{\sigma}(A))(v) \neq (\tau^{-1}\circ\sigma_0^2\circ\pi)(v) \iff (\rho\circ\hat{\sigma}(A)\circ\pi^{-1})(u) \neq (\tau^{-1}\circ\sigma_0^2)(u),$$

so $(\tau\circ\rho\circ\xi^{-1}\circ\hat{\sigma}(\pi A))(u) \neq \sigma_0^2(u)$. Thus, $v \in \mathcal{E}[\hat{\sigma}(A),\sigma_0^1]$ if and only if $\pi(v) \in \mathcal{E}[\hat{\sigma}(\pi A),\sigma_0^2]$. Finally, we conclude that

$$\mathbb{E}_1\tilde{l}(\hat{\sigma}(A),\sigma_0^1) = \frac{1}{n}\sum_{v=1}^n P_1(v \in \mathcal{E}[\hat{\sigma}(A),\sigma_0^1]) = \frac{1}{n}\sum_{v=1}^n P_1(\pi(v) \in \mathcal{E}[\hat{\sigma}(\pi A),\sigma_0^2])$$
$$\stackrel{(a)}{=} \frac{1}{n}\sum_{v=1}^n P_2(\pi(v) \in \mathcal{E}[\hat{\sigma}(A),\sigma_0^2]) = \mathbb{E}_2\tilde{l}(\hat{\sigma}(A),\sigma_0^2),$$

where (a) follows because $[\pi A]_{ij} = A_{\pi^{-1}(i),\pi^{-1}(j)}$, implying that if A is distributed according to $P_{SBM}(A\,|\,\sigma_0^1)$, then πA is distributed according to $P_{SBM}(A\,|\,\sigma_0^2)$. This concludes the proof.

G.2 Properties of permutation equivariant estimators

The following lemma establishes a symmetry property used to prove Theorem 5.2:

Lemma G.1. *Let the true clustering σ_0 be arbitrary. Suppose the weight matrix A is drawn from an arbitrary probability measure and $\hat{\sigma}$ is any permutation equivariant estimator. Let u and v be two nodes such that there exists $\pi \in S_n$ satisfying*

(1) $\pi(u) = v$,

(2) π is measure-preserving; i.e., $A \stackrel{d}{=} \pi A$, and

(3) π preserves the true clustering; i.e., there exists $\tau \in S_K$ such that $\tau \circ \sigma_0 \circ \pi^{-1} = \sigma_0$.

Then

$$P(u \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]) = P(v \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]).$$

Proof. Since $\hat{\sigma}(A) \stackrel{d}{=} \hat{\sigma}(\pi A)$, we have

$$P\Big(v \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]\Big) = P\Big(v \in \mathcal{E}[\hat{\sigma}(\pi(A)), \sigma_0]\Big).$$

We claim that $u \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]$ if and only if $v \in \mathcal{E}[\hat{\sigma}(\pi A), \sigma_0]$, implying the desired result:

$$P\Big(u \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]\Big) = P\Big(v \in \mathcal{E}[\hat{\sigma}(\pi A), \sigma_0]\Big) = P\Big(v \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]\Big).$$

Consider a fixed matrix A, and let $\tau \in S_K$ satisfy $\tau \circ \sigma_0 \circ \pi^{-1} = \sigma_0$. Let $\xi \in S_K$ be the permutation such that $\hat{\sigma}(\pi A) = \xi \circ \hat{\sigma}(A) \circ \pi^{-1}$. For any $\rho \in S_K$, we have

$$d_H(\rho \circ \hat{\sigma}(A), \ \sigma_0) = d_H(\tau \circ \rho \circ \xi^{-1} \circ \xi \circ \hat{\sigma}(A) \circ \pi^{-1}, \ \tau \circ \sigma_0 \circ \pi^{-1})$$
$$= d_H(\tau \circ \rho \circ \xi^{-1} \circ \hat{\sigma}(\pi A), \ \sigma_0).$$

Therefore, $\rho \in S_K[\hat{\sigma}(A), \sigma_0]$ if and only if $\tau \circ \rho \circ \xi^{-1} \in S_K[\hat{\sigma}(\pi(A)), \sigma_0]$. In particular, if $v \in \mathcal{E}[\hat{\sigma}(\pi A), \sigma_0]$, we have $\tau \circ \rho \circ \xi^{-1} \circ \hat{\sigma}(\pi A)(v) \neq \sigma_0(v)$ for some $\rho \in S_K[\hat{\sigma}(A), \sigma_0]$. Then

$$\hat{\sigma}(A)(u) = \hat{\sigma}(A)(\pi^{-1}(v)) = \xi^{-1} \circ \xi \circ \hat{\sigma}(A) \circ \pi^{-1}(v) = \xi^{-1} \circ \hat{\sigma}(\pi A)(v)$$
$$\neq \rho^{-1} \circ \tau^{-1} \circ \sigma_0(v) = \rho^{-1} \circ \tau^{-1} \circ \sigma_0(\pi(u)) = \rho^{-1}(\sigma_0(u)).$$

Thus, $u \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]$. Similar reasoning shows that if $u \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]$, then $v \in \mathcal{E}[\hat{\sigma}(\pi A), \sigma_0]$. $\qquad\square$

Corollary G.1. *Let the true clustering σ_0 be arbitrary. Suppose the weight matrix A is drawn from a weighted SBM and $\hat{\sigma}$ is any permutation equivariant estimator. Let u and v be two nodes lying in equal-sized clusters. Then*

$$P\Big(u \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]\Big) = P\Big(v \in \mathcal{E}[\hat{\sigma}(A), \sigma_0]\Big).$$

Proof. By Lemma G.1, it suffices to construct a permutation $\pi \in S_n$ satisfying conditions (1)–(3). First suppose u and v lie in the same cluster. It is easy to see that the conditions are satisfied when π is the permutation that swaps u and v and τ is the identity. If u and v lie in different clusters, suppose without loss of generality that u is in cluster 1 and v is in cluster 2, where clusters 1 and 2 have the same size. Let π be the permutation that exchanges all nodes in cluster 1 with all nodes in cluster 2. The conditions are satisfied when τ is the permutation that transposes cluster labels 1 and 2. $\qquad\square$

G.3 Properties of Renyi divergence

We first state a lemma that provides an alternative characterization of the Renyi divergence. The proof of this lemma uses the same technique as Anantharam [7].

Lemma G.2. *Let P and Q be two probability measures on $S \subset \mathbb{R}$ that are absolutely continuous with respect to each other, with point masses P_0 and Q_0 at zero. Then $I(P, Q) = 2D$, where*

$$D := \inf_{Y \in \mathcal{P}} \max \left\{ \int \log \frac{dY}{dP} dY, \int \log \frac{dY}{dQ} dY \right\},$$

and $\mathcal{P}$ denotes the set of probability distributions absolutely continuous with respect to both P and Q. Furthermore, the infimum in D is attained by a distribution Y^* whose singular (with respect to the Lebesgue measure) part is $Y_0^* = \frac{1}{Z}(P_0 Q_0)^{1/2}$ and continuous part is $(1 - Y_0^*) y^*(x) = \frac{1}{Z} \int_S \sqrt{(1 - P_0)(1 - Q_0) p(x) q(x)} \, dx$, where $Z = (P_0 Q_0)^{1/2} + \int_S \sqrt{(1 - P_0)(1 - Q_0) p(x) q(x)} \, dx$.

Proof. First note that D is finite by choosing $Y = P$. We claim that

$$D = \inf_{Y \in \mathcal{P}} \left\{ \int \log \frac{dY}{dP} dY : \int \log \frac{dP}{dQ} dY = 0 \right\}. \tag{115}$$

This holds because for any $Y \in \mathcal{P}$ such that $\int \log \frac{dP}{dQ} dY \neq 0$, we have $\int \log \frac{dY}{dP} dY \neq \int \log \frac{dY}{dQ} dY$. Suppose without loss of generality that the first quantity is larger. Then it is possible to take $\tilde{Y} = (1 - \epsilon) Y + \epsilon P$ for ϵ small enough such that $\max \left\{ \int \log \frac{d\tilde{Y}}{dP} d\tilde{Y}, \int \log \frac{d\tilde{Y}}{dQ} d\tilde{Y} \right\}$ strictly decreases, so the infimum in the definition of D could not have been achieved.

Now let $Y \in \mathcal{P}$ be such that $\int \log \frac{dP}{dQ} dY = 0$. Since $Y \ll Y^*$, we have

$$\int \log \frac{dY}{dP} dY = \int \log \frac{dY}{dY^*} \frac{dY^*}{dP} dY \geq \int \log \frac{dY}{dY^*} dY + \int \log \frac{dY^*}{dP} dY$$

$$= \frac{1}{2} \int \log \frac{dP}{dQ} dY - \log Z = - \log Z = -\frac{I(P, Q)}{2}.$$

We finish the proof by verifying that $\int \log \frac{dP}{dQ} dY^* = 0$, and that in the derivation above, the inequality holds with inequality when $Y = Y^*$. $\qquad \square$

G.4 Bounding the variance

Lemma G.3. *Let C^* and $\mathcal{G}^*$ be defined as in Theorem 5.2, let $((P_0, p), (Q_0, p)) \in \mathcal{G}^*$, and let σ_0 be a clustering satisfying the hypothesis of Theorem 5.2. Let σ^* be defined as (103), and $\mathcal{Q}$ be defined as in equation (104). Then*

$$V_\Psi(\mathcal{Q}) \leq 63 C^{*2} \frac{n}{\beta K} I((P_0, p), (Q_0, q)).$$

Proof. Observe that since $\mathbb{E}_\Psi[\mathcal{Q} \mid \sigma^* = \sigma_0^1] = \mathbb{E}_\Psi[\mathcal{Q} \mid \sigma^* = \sigma_0^2]$, we have

$$V_\Psi(\mathcal{Q}) = \text{Var}(\mathbb{E}_\Psi[\mathcal{Q} \mid \sigma^*]) + E[\text{Var}_\Psi(\mathcal{Q} \mid \sigma^*)] = E[\text{Var}_\Psi(\mathcal{Q} \mid \sigma^*)]$$

$$= \frac{1}{2} \text{Var}_\Psi(\mathcal{Q} \mid \sigma^* = \sigma_0^1) + \frac{1}{2} \text{Var}_\Psi(\mathcal{Q} \mid \sigma^* = \sigma_0^2). \tag{116}$$

Let us consider the first term of equation (116). Under the distribution $\mathbb{P}_\Psi$, the random variables $A_{1,u}$ for any $u \in C_1 \cup C_2$ are independent and identically distributed according to Y. Let u' be an

arbitrary node in $C_1 \cup C_2$ such that $u' \neq 1$. We thus have

$$\mathrm{Var}_{\Psi}(\mathcal{Q} \mid \sigma^* = \sigma_0^1) = \sum_{\substack{u:\, u \neq 1, \\ \sigma_0^1(u)=1}} \mathrm{Var}_{\Psi}\left(\log \frac{Y(A_{1,u})}{P(A_{1,u})}\right) + \sum_{u:\, \sigma_0^1(u)=2} \mathrm{Var}_{\Psi}\left(\log \frac{Y(A_{1,u})}{Q(A_{1,u})}\right)$$

$$\leq \frac{n}{\beta K}\mathbb{E}_{\Psi}\left[\left(\log \frac{Y(A_{1,u'})}{P(A_{1,u'})}\right)^2\right] + \frac{n}{\beta K}\mathbb{E}_{\Psi}\left[\left(\log \frac{Y(A_{1,u'})}{Q(A_{1,u'})}\right)^2\right].$$

By the same argument, we can show that the second term of equation (116) is upper-bounded by the same quantity, so

$$V_{\Psi}(\mathcal{Q}) \leq \frac{n}{\beta K}\frac{n}{\beta K}\mathbb{E}_{\Psi}\left[\left(\log \frac{Y(A_{1,u'})}{P(A_{1,u'})}\right)^2\right] + \frac{n}{\beta K}\mathbb{E}_{\Psi}\left[\left(\log \frac{Y(A_{1,u'})}{Q(A_{1,u'})}\right)^2\right]$$

$$\leq \frac{n}{\beta K}\left\{\int\left(\log \frac{dY}{dP}\right)^2 dY + \int\left(\log \frac{dY}{dQ}\right)^2 dY\right\}. \tag{117}$$

By the definition of Y, we have

$$\int\left(\log \frac{dY}{dP}\right)^2 dY = Y_0\left(\log \frac{Y_0}{P_0}\right)^2 + (1-Y_0)\int y(x)\left(\log \frac{(1-Y_0)y(x)}{(1-P_0)p(x)}\right)^2 dx. \tag{118}$$

For simplicity of presentation, we use the shorthand

$$I := I((P_0, p), (Q_0, q)) \quad \text{and}$$

$$H := (\sqrt{P_0} - \sqrt{Q_0})^2 + (\sqrt{1-P_0} - \sqrt{1-Q_0})^2 + \sqrt{(1-P_0)(1-Q_0)}\int(\sqrt{p(x)} - \sqrt{q(x)})^2 dx,$$

from this point on. Let $Z = \sqrt{P_0 Q_0} + \sqrt{(1-P_0)(1-Q_0)}\int\sqrt{p(x)q(x)}dx$. It is then clear that $I = -2\log Z$ and $Z \geq \frac{1}{2}$, since $I \leq 2\log 2$ by assumption. We bound the first term of equation (118). Note that, since $(x+y)^2 \leq 2x^2 + 2y^2$, we have

$$Y_0\left(\log \frac{Y_0}{P_0}\right)^2 \leq \frac{1}{2}Y_0\left(\log \frac{Q_0}{P_0}\right)^2 + \frac{1}{2}I^2. \tag{119}$$

Suppose $P_0 \leq Q_0$. Then by Lemma H.1, we have $0 \leq \log \frac{Q_0}{P_0} \leq \frac{Q_0 - P_0}{P_0}$. Thus, we have

$$Y_0\left(\log \frac{Y_0}{P_0}\right)^2 \leq C^{*2}\frac{(Q_0 - P_0)^2}{P_0 \vee Q_0} + \frac{1}{2}I^2$$

$$\leq 4C^{*2}(\sqrt{Q_0} - \sqrt{P_0})^2 + \frac{1}{2}I^2 \leq 4C^{*2}H + \frac{1}{2}I^2 \overset{(a)}{\leq} 4C^{*2}I + \frac{1}{2}I^2,$$

using inequality (119) and the fact that $((P_0, p), (Q_0, p)) \in \mathcal{G}^*$. Here, (a) follows from the fact that $I = -2\log(1 - \frac{1}{2}H)$ and Lemma H.1. Now suppose $P_0 \geq Q_0$. By Lemma H.1 again, we have $0 \geq \log \frac{Q_0}{P_0} \geq -(1 + \frac{1}{2}C^*)\left(\frac{P_0 - Q_0}{P_0}\right) \leq -C^*\left(\frac{P_0 - Q_0}{P_0}\right)$. Therefore,

$$Y_0\left(\log \frac{Y_0}{P_0}\right)^2 \leq C^{*2}\frac{(Q_0 - P_0)^2}{P_0 \vee Q_0} + \frac{1}{2}I^2 \leq 4C^{*2}I + \frac{1}{2}I^2.$$

Thus, we conclude that

$$Y_0\left(\log \frac{Y_0}{P_0}\right)^2 \leq 4C^{*2}I + \frac{1}{2}I^2. \tag{120}$$

Now we turn our attention to the second term in equation (118). We have

$$(1 - Y_0) \int y(x) \left(\log \frac{1 - Y_0}{1 - P_0} \frac{y(x)}{p(x)} \right)^2 dx$$

$$\leq (1 - Y_0) \int y(x) \left\{ \frac{1}{2} \left| \log \frac{1 - Q_0}{1 - P_0} \right| + \frac{1}{2} \left| \log \frac{q(x)}{p(x)} \right| + \frac{I}{2} \right\}^2 dx$$

$$\leq \frac{9}{4}(1 - Y_0) \int y(x) \left\{ \left| \log \frac{1 - Q_0}{1 - P_0} \right|^2 + \left| \log \frac{q(x)}{p(x)} \right|^2 + I^2 \right\} dx, \tag{121}$$

where we have used the fact that $(x + y + z)^2 \leq 9x^2 + 9y^2 + 9z^2$ in the last inequality. Define

$$\mathcal{A} := (1 - Y_0) \left| \log \frac{1 - Q_0}{1 - P_0} \right|^2 \int y(x)dx, \qquad \mathcal{B} := (1 - Y_0) \int y(x) \left| \log \frac{q(x)}{p(x)} \right|^2 dx,$$

$$\mathcal{C} := (1 - Y_0)I^2 \int y(x)dx.$$

We bound each term separately, beginning with $\mathcal{A}$. Note that

$$\int y(x)dx \leq 2 \int \sqrt{(1 - P_0)(1 - Q_0)p(x)q(x)}dx \leq 2. \tag{122}$$

Note also that we can use the same reasoning that we applied to derive inequality (120) to show that $(1 - Y_0) \left| \log \frac{1 - Q_0}{1 - P_0} \right|^2 \leq 4C^{*2}I$. Therefore, we have $\mathcal{A} \leq 8C^{*2}I$. Moving on to $\mathcal{B}$, we have

$$\mathcal{B} \leq 2\sqrt{(1 - P_0)(1 - Q_0)} \int \sqrt{p(x)q(x)} \left| \log \frac{q(x)}{p(x)} \right|^2 dx$$

$$\leq \sqrt{(1 - P_0)(1 - Q_0)} \int (p(x) + q(x)) \left| \log \frac{q(x)}{p(x)} \right|^2 dx$$

$$\overset{(a)}{\leq} C^* \sqrt{(1 - P_0)(1 - Q_0)} \int (\sqrt{p(x)} - \sqrt{q(x)})^2 dx \leq C^*I,$$

where (a) follows by the hypothesis of the proposition. Finally, from inequality (122), we have $\mathcal{C} \leq 8(1 - Y_0)I^2 \leq (16 \log 2)I$. Substituting back into inequality (121), we therefore obtain

$$(1 - Y_0) \int y(x) \left(\log \frac{1 - Y_0}{1 - P_0} \frac{y(x)}{p(x)} \right)^2 dx \leq \frac{9}{4}(8C^{*2} + C^* + 16 \log 2)I.$$

Substituting inequality (120) and the above inequality back into inequality (118), we obtain the desired bound on the first term of inequality (117). The second term of inequality (117) can be bounded in exactly the same manner. Thus, we have overall bound

$$V_\Psi(\mathcal{Q}) \leq 2(22C^{*2} + 3C^* + 13)\frac{nI}{\beta K} \leq 63C^{*2}\frac{nI}{\beta K},$$

where we have used the fact that $C^* \geq 1$. $\qquad \square$

H Additional useful lemmas

Lemma H.1. *For all $x \in (-\infty, 1)$, we have $\log(1 - x) \leq -x$. Moreover, fix $t \in [0, 1)$. Then for all $x \in [-t, t]$, we have*

$$-x\left(1 + \frac{x}{2(1 - t)^2}\right) \leq \log(1 - x) \leq -x.$$

Proof. Let $x \in [-t, t]$. Then $\log(1 - x) \leq -x$ by concavity of the logarithm. On the other hand, by Taylor's theorem, there exists x' such that $|x'| \leq |x|$ and

$$\log(1 - x) = -x - \frac{1}{2(1 - x')^2}x^2 \geq -x\left(1 + \frac{x}{2(1 - t)^2}\right).$$

$\square$

Lemma H.2. *Let P and Q be two distributions absolutely continuous with respect to each other, and let $H(P, Q) := \int\left(\left(\frac{dP}{dQ}\right)^{1/2} - 1\right)dQ$. Then $I(P, Q) \geq H(P, Q)$. If in addition $I(P, Q) \leq 2\log 2$, then $H(P, Q) \leq 1$ and $I(P, Q) \leq (1 + H(P, Q))H(P, Q)$.*

Proof. The conclusion follows from the fact that $I(P, Q) = -2\log(1 - H(P, Q)/2)$ and Lemma H.1.

$\square$

References

[1] E. Abbe. Community detection and stochastic block models: Recent developments. *arXiv preprint arXiv:1703.10146*, 2017.

[2] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer. Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *IEEE Transactions on Network Science and Engineering*, 1(1):10–22, 2014.

[3] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62:1, 2014.

[4] E. Abbe and C. Sandon. Recovering communities in the general stochastic block model without knowing the parameters. In *Advances in Neural Information Processing Systems*, pages 676–684, 2015.

[5] Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. pages 670–688, 2015.

[6] C. Aicher, A. Z. Jacobs, and A. Clauset. Learning latent block structure in weighted networks. *Journal of Complex Networks*, 2014.

[7] Venkat Anantharam. A variational characterization of rényi divergences. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 893–897. IEEE, 2017.

[8] S. Balakrishnan, M. Xu, A. Krishnamurthy, and A. Singh. Noise thresholds for spectral clustering. In *Advances in Neural Information Processing Systems*, pages 954–962, 2011.

[9] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752, 2004.

[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[11] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.

[12] P. Chin, A. Rao, and V. Vu. Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Proceedings of The 28th Conference on Learning Theory*, pages 391–423, 2015.

[13] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84:066106, Dec 2011.

[14] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.

[15] S. E. Fienberg, M. M. Meyer, and S. S. Wasserman. Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, 80(389):51–67, 1985.

[16] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Community detection in degree-corrected block models. *arXiv preprint arXiv:1607.06993*, 2016.

[17] Chao Gao, Zongming Ma, Anderson Y. Zhang, and Harrison H. Zhou. Achieving optimal misclassification proportion in stochastic block model. *The Journal of Machine Learning Research*, 18, 2017.

[18] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi. A survey of statistical network models. *Found. Trends Mach. Learn.*, 2(2):129–233, February 2010.

[19] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62:5, 2014.

[20] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *arXiv preprint arXiv:1502.07738*, 2015.

[21] B. Hajek, Y. Wu, and J. Xu. Submatrix localization via message passing. *arXiv preprint arXiv:1510.09219*, 2015.

[22] B. Hajek, Y. Wu, and J. Xu. Information limits for recovering a hidden community. *IEEE Transactions on Information Theory*, 2017.

[23] E Hartuv and R Shamir. A clustering algorithm based on graph connectivity. *Information Processing Letters*, 76(4–6):175–181, 2000.

[24] S. Heimlicher, M. Lelarge, and L. Massoulié. Community detection in the labelled stochastic block model. *arXiv preprint arXiv:1209.2910*, 2012.

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

[26] M. O. Jackson. *Social and Economic Networks*. Princeton University Press, 2010.

[27] V. Jog and P. Loh. Information-theoretic bounds for exact recovery in weighted stochastic block models using the Renyi divergence. *arXiv preprint arXiv:1509.06418*, 2015.

[28] J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

[29] M. Lelarge, L. Massoulié, and J. Xu. Reconstruction in the labeled stochastic block model. In *Information Theory Workshop (ITW), 2013 IEEE*, pages 1–5. IEEE, 2013.

[30] L. Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 694–703. ACM, 2014.

[31] F. McSherry. Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science*, pages 529–537. IEEE, 2001.

[32] E. Mossel, J. Neeman, and A. Sly. Stochastic Block Models and Reconstruction. *arXiv preprint arXiv:1202.1499*, 2012.

[33] E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for binary symmetric block models. *arXiv preprint arXiv:1407.1591*, 2014.

[34] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, pages 1–44, 2013.

[35] M. Newman, A.-L. Barabasi, and D. J. Watts. *The Structure and Dynamics of Networks: (Princeton Studies in Complexity)*. Princeton University Press, Princeton, NJ, USA, 2006.

[36] M. E. J. Newman. Analysis of weighted networks. *Physical Review E*, 70(5):056131, 2004.

[37] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[38] J. K. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.

[39] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010. Computational Models of the Brain.

[40] D.S. Sade. Sociometrics of Macaca mulatta: I. Linkages and cliques in grooming matrices. *Folia Primatologica*, 18(3–4):196–223, 1972.

[41] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, August 2000.

[42] S. Yun and A. Proutiere. Optimal cluster recovery in the labeled stochastic block model. In *Advances in Neural Information Processing Systems*, pages 965–973, 2016.

[43] Anderson Y Zhang and Harrison H Zhou. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016.

[44] B. Zhang and S. Horvath. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, 4(1):1128, 2005.

[45] Y. Zhao, E. Levina, and J. Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266–2292, 2012.