Proceedings of Machine Learning Research vol 291:1–41, 2025 38th Annual Conference on Learning Theory

Low coordinate degree algorithms II: Categorical signals and generalized stochastic block models

Dmitriy Kunisky KUNISKY@JHU.EDU
Department of Applied Mathematics & Statistics, Johns Hopkins University

Editors: Nika Haghtalab and Ankur Moitra

Abstract

We study when low coordinate degree functions (LCDF)—linear combinations of functions depending on small subsets of entries of a vector—can test for the presence of categorical structure, including community structure and generalizations thereof, in high-dimensional data. This complements recent results studying the power of LCDF in testing for continuous structure like real-valued signals corrupted by additive noise. We study a general form of stochastic block model (SBM), where a population is assigned random labels and every p-tuple generates an observation according to an arbitrary probability measure associated to the p labels of its members. We show that the performance of LCDF admits a unified analysis for this class of models. As applications, we prove tight lower bounds against LCDF for broad families of previously studied graph and uniform hypergraph SBMs, always matching suitable generalizations of the Kesten-Stigum threshold. We also prove tight lower bounds for group synchronization and abelian group sumset problems under the "truth-or-Haar" noise model, and give an improved analysis of Gaussian multi-frequency group synchronization. In most of these models, for some parameter settings our lower bounds give new evidence for conjectural statistical-to-computational gaps. Finally, interpreting some of our findings, we propose a new analogy between categorical and continuous signals: a general SBM as above behaves qualitatively like a spiked p_*-tensor model of a certain order p_* depending on the parameters of the SBM.

Keywords: Low-degree polynomials, community detection, stochastic block model, subexponential time algorithms, average-case computational complexity

1. Introduction

The recent work of Kunisky (2024) proposed *low coordinate degree function (LCDF) algorithms* (reviewed in Section 3 below) as a fruitful alternative to the now widely-studied class of *low degree polynomial (LDP) algorithms* for hypothesis testing (Hopkins and Steurer, 2017; Barak et al., 2019; Hopkins et al., 2017; Hopkins, 2018; Kunisky et al., 2022). That work found that LCDF algorithms, though a more powerful and expressive class than LDP algorithms, are actually often easier to analyze. It gave a unified analysis of the performance of LCDF for detecting weak high-dimensional signals observed through entrywise independent noisy channels, including models such as spiked matrices and tensors under general noise models. In those results, the notion of a *weak* signal has a specific quantitative meaning: the signal takes continuous values, say in $\mathbb{R}$, and has typically to be close to zero. In that case, the difficulty of a testing problem is governed only by the way that the noisy channel corrupts infinitesimal signals, which is why it turns out that, in that setting, channels do not all behave differently but rather fall into large *universality classes*: only a single scalar summary statistic characterizes how difficult they are for LCDF algorithms.

Here we continue to explore LCDF algorithms and the greater level of generality at which they allow us to analyze such statistical questions, but we focus on the complementary setting of *categorical* signals. The motivation for doing this is another broad but very different class of models of interest in the high-dimensional statistics literature: often a signal of interest is not a continuous object at all, but rather a combinatorial one, like the assignment of the nodes of a random graph to latent communities that govern nodes' probabilities of being connected to one another or not. Sometimes, as discussed for certain stochastic block models by Kunisky (2024), it is possible to shoehorn such models into the setting of continuous signals (after all, the community assignments above are not continuous, but the *probabilities* of pairs being connected still are). But this is not always possible or convenient, and we will see that developing a separate framework tailored to combinatorial models reveals different universality phenomena and subtler parallels between testing for continuous and categorical signals.

2. Generalized Stochastic Block Models

Our results will treat a general form of *stochastic block model (SBM)*, a model of random graphs originating in the social science literature; see Moore (2017); Abbe (2017) for recent theoretical surveys. Our model is similar to the families of models considered in Heimlicher et al. (2012); Lelarge et al. (2015) under the name of *labeled SBMs*, and especially similar to the *generalized SBM* of Xu et al. (2014), though still differing slightly in the level of generality of different ingredients of the definition. The definition is so similar in spirit to the latter, though, that we keep the same name.

Definition 1 (Generalized stochastic block model) *Let $p \geq 2$, let $k, n \geq 1$, and let Ω be a measurable space. A* generalized stochastic block model (GSBM) *is specified by, for each $\boldsymbol{a} \in [k]^p$, a probability measure $\mu_{\boldsymbol{a}}$ on Ω. Write*

$$\mu_{\mathsf{avg}} := \frac{1}{k^p} \sum_{\boldsymbol{a} \sim \mathrm{Unif}([k]^p)} \mu_{\boldsymbol{a}}, \tag{1}$$

so that μ_{avg} is another probability measure on Ω. The GSBM then consists of the following two probability measures over $\boldsymbol{Y} \in \Omega^{\binom{[n]}{p}}$:

1. *Under $\mathbb{Q}$, draw $\boldsymbol{Y} \sim \mathbb{Q}$ with $Y_S \sim \mu_{\mathsf{avg}}$ independently for each $S \in \binom{[n]}{p}$.*
2. *Under $\mathbb{P}$, first draw $\boldsymbol{x} \sim \mathrm{Unif}([k]^n)$. Then, for each $S = \{s_1 < \cdots < s_p\} \in \binom{[n]}{p}$, draw $Y_S \sim \mu_{x_{s_1}, \dots, x_{s_p}}$ independently.*

We call p the order *and n the* size *of a GSBM.*

We will be interested in the problem of *detection* or *hypothesis testing* in such a model: given a draw from either $\mathbb{Q}$ or $\mathbb{P}$, can we with high probability determine which measure the observation was drawn from? Precise details about our notion of testing follow below in Section 3.

Remark 2 (Non-uniform label assignments) *For the sake of simplicity we restrict our attention to the special case $\boldsymbol{x} \sim \mathrm{Unif}([k]^n)$ in the definition of $\mathbb{P}$. For general distributions of $\boldsymbol{x}$, a similar result to Lemma 41 will hold, with the multinomial distribution featuring there replaced by a suitable random variable related to two independent draws of $\boldsymbol{x}$. However, analyzing and proving lower bounds for general distributions of $\boldsymbol{x}$ would require a careful generalization of the tools we develop in Appendix A.3 for the multinomial distribution, for which case they are already non-trivial innovations.*

The class of GSBMs models essentially arbitrary situations where some collection of objects are each assigned discrete information denoted by their label in $[k]$, and where each subset of p objects interacts in a way that depends only on their labels. We then observe the outcome of all of these interactions, through a family of noisy channels described by the collection of $(\mu_{\boldsymbol{a}})_{\boldsymbol{a}\in[k]^p}$, which say how any collection of p labels in $[k]$ gives rise to a random observation in Ω.

In conventional stochastic block models on graphs or hypergraphs, what we have called "labels" above are assignments to communities, and $\Omega = \{0, 1\}$, so that outcomes of interactions may be interpreted as Boolean variables and the entire collection of outcomes as (the adjacency matrix or tensor of) a graph (if $p = 2$) or hypergraph (if $p \geq 3$). We emphasize, though, that Ω need not equal $\{0, 1\}$ or even a discrete set at all, but can be absolutely arbitrary. We handle but do not give applications where Ω is continuous, but one non-Boolean situation we will consider in some of our applications is when Ω is a finite group, for instance.

We conclude the discussion of the setting with the assumptions we make on a GSBM in our main results.

Assumption 3 *We always make the following assumptions on a GSBM:*

1. *The GSBM is* non-trivial*: there exists $\boldsymbol{a} \in [k]^p$ such that $\mu_{\boldsymbol{a}} \neq \mu_{\mathsf{avg}}$.*
2. *The GSBM is L^2-*integrable*: for all $\boldsymbol{a} \in [k]^p$, the likelihood ratio $d\mu_{\boldsymbol{a}}/d\mu_{\mathsf{avg}}$ belongs to $L^2(\mu_{\mathsf{avg}})$.*[1]
3. *The GSBM is* weakly symmetric*: for all $\boldsymbol{a}, \boldsymbol{b} \in [k]^p$ and all permutations $\sigma \in \mathrm{Sym}([p])$,*

$$\mathop{\mathbb{E}}_{y\sim\mu_{\mathsf{avg}}}\left[\frac{d\mu_{(a_1,\dots,a_p)}}{d\mu_{\mathsf{avg}}}(y)\cdot\frac{d\mu_{(b_1,\dots,b_p)}}{d\mu_{\mathsf{avg}}}(y)\right] = \mathop{\mathbb{E}}_{y\sim\mu_{\mathsf{avg}}}\left[\frac{d\mu_{(a_{\sigma(1)},\dots,a_{\sigma(p)})}}{d\mu_{\mathsf{avg}}}(y)\cdot\frac{d\mu_{(b_{\sigma(1)},\dots,b_{\sigma(p)})}}{d\mu_{\mathsf{avg}}}(y)\right].$$

Note that these expectations are well-defined and finite by the integrability assumption.

Non-triviality excludes the case where $\mathbb{P} = \mathbb{Q}$, i.e. the two models we propose testing between are exactly the same probability measure, in which case of course testing is trivially impossible. Integrability ensures that various objects akin to χ^2 divergences that we will encounter are always finite. Weak symmetry is slightly subtler. One useful remark is that it is implied by the following stronger and more natural condition.

Definition 4 *We say that a GSBM is* strongly symmetric *if, for all $\boldsymbol{a} \in [k]^p$ and $\sigma \in \mathrm{Sym}([p])$ a permutation, $\mu_{(a_1,\dots,a_p)} = \mu_{(a_{\sigma(1)},\dots,a_{\sigma(p)})}$.*

Proposition 5 *A strongly symmetric GSBM is also weakly symmetric.*

Weak symmetry suffices for our calculations, and we will encounter at least one interesting model, synchronization over finite groups, where only weak symmetry is satisfied (in short, because there pairs of group elements interact such that from the pair g, h one receives a noisy version of the asymmetric group difference gh^{-1}). One may think of strong symmetry as asking that the observation channel $\mu_{\boldsymbol{a}}$ truly has no "direction" it imposes among its inputs $(a_1, \dots, a_p)$, while weak symmetry allows for such a directionality but only asks that a certain scalar measurement of the "relative similarity" of distributions of observations through the channel does not depend on it.

1. That the likelihood ratio exists, i.e., that $\mu_{\boldsymbol{a}}$ is absolutely continuous to μ_{avg}, follows from the definition (1) of μ_{avg}.

3. Low Coordinate Degree Algorithms

We will be interested in the following specific notion of success in hypothesis testing.

Definition 6 (Strong separation) *Consider sequences of probability measures $\mathbb{P}_n, \mathbb{Q}_n$ over measurable spaces Ω_n. We say that functions $f_n : \Omega_n \to \mathbb{R}$ achieve* strong separation *if, as $n \to \infty$,*

$$\mathop{\mathbb{E}}_{\boldsymbol{Y}\sim\mathbb{P}_n} f_n(\boldsymbol{Y}) - \mathop{\mathbb{E}}_{\boldsymbol{Y}\sim\mathbb{Q}_n} f_n(\boldsymbol{Y}) = \omega\left(\sqrt{\mathop{\mathsf{Var}}_{\boldsymbol{Y}\sim\mathbb{Q}_n} f_n(\boldsymbol{Y})} + \sqrt{\mathop{\mathsf{Var}}_{\boldsymbol{Y}\sim\mathbb{P}_n} f_n(\boldsymbol{Y})}\right). \tag{2}$$

Strong separation is a notion of it being possible to hypothesis test consistently: given a strongly separating family of functions, one may choose a suitable threshold and obtain a sequence of hypothesis tests having Type I and II error probabilities both $o(1)$ as $n \to \infty$, by Chebyshev's inequality. The converse is also true, and strong separation by *some* family of functions is equivalent to the above notion of consistent testing, often called *strong detection*.

However, we will ask not whether there exist *any* functions achieving strong separation, but whether there exist *computationally tractable* such functions. As alluded to above, much prior work has focused on this question for low degree polynomial (LDP) functions. It should already be clear, however, that this will not work for the models we have proposed to study, since our observations will take values in arbitrary spaces. Instead, and as explored in Kunisky (2024) based on an early proposal by Hopkins (2018), we focus on the following more general class of functions that also have the benefit of being well-defined much more generally.[2]

Let Ω be a measurable space and $\mathbb{Q}$ be a product measure on Ω^N for some $N \geq 1$, as in Definition 1. We follow the definitions from Kunisky (2024). For $\boldsymbol{y} \in \Omega^N$ and $T \subseteq [N]$, we write $\boldsymbol{y}_T \in \Omega^T$ for the restriction of $\boldsymbol{y}$ to the coordinates in T.

Definition 7 (Coordinate degree) *For each $D \geq 0$, define a subspace of $L^2(\mathbb{Q})$,*

$$V_{\leq D} := \sum_{\substack{T\subseteq[N] \\ |T|\leq D}} \{f \in L^2(\mathbb{Q}) : f(\boldsymbol{y}) \text{ depends only on } \boldsymbol{y}_T\}.$$

For $f \in L^2(\mathbb{Q})$, define $\mathsf{cdeg}(f) := \min\{D : f \in V_{\leq D}\}$, and call this the coordinate degree *of f.*

We call $V_{\leq D}$ a space of low coordinate degree functions (LCDF), and this is the precise meaning of this term mentioned earlier.

For polynomials f, $\mathsf{cdeg}(f) \leq \deg(f)$. Thus, when we prove lower bounds against LCDF we also always learn lower bounds against LDP, in the style of the results cited above as well as recent works like Bandeira et al. (2020); Ding et al. (2023); Bandeira et al. (2021, 2022); Brennan et al. (2020); Rush et al. (2023); Kothari et al. (2023). But, LCDF have further advantages: they are sensible for arbitrary Ω, for instance Ω an object like a finite group whose elements have only algebraic meaning, and they include many more functions, such as LDP applied after arbitrary entrywise functions taking numerical values (say, low-degree polynomials of a representation applied to a vector of group elements).

2. These functions also played an important role in the recent work Brennan et al. (2020).

Remark 8 (Equivalence of LDP and LCDF for finite domains) *For finite sets Ω, coordinate degree is the same as polynomial degree when Ω is given the "one-hot" Boolean encoding, as discussed in Koehler and Mossel (2021); Kunisky (2024). As in our applications Ω is always finite, our lower bounds against LCDF are equivalently lower bounds against LDP under the one-hot encoding. However, we emphasize that our general results also allow for Ω infinite, in which case, even when Ω consists of numerical values so that LDP are well-defined, LCDF are of a given degree are a strictly larger function class than LDP of the same degree. Further, as we will see, working with LCDF leads to clearer and simpler calculations than directly working with LDP on the one-hot encoding.*

We now have all of the ingredients needed to describe the template that all of our main results will follow: we will show, for sequences of hypothesis testing problems of $\mathbb{Q}_n$ versus $\mathbb{P}_n$ arising from sequences of GSBMs for growing n, that sequences of LCDF $(f_n)_{n\geq 1}$ with $\mathsf{cdeg}(f_n) \leq D(n)$ cannot strongly separate $\mathbb{Q}_n$ and $\mathbb{P}_n$. In particular, as for LDP (see, e.g., the discussion in Kunisky et al. (2022)), we may think intuitively of LCDF of coordinate degree $D(n)$ as describing all functions taking roughly time $n^{D(n)} = \exp(\widetilde{\Theta}(D(n)))$ to compute. Thus, these kinds of lower bounds suggest, depending on the scaling of $D(n)$, that various orders of superpolynomial but subexponential time are required to solve testing problems.

4. Main Results

4.1. General Theory

At the center of the theory we develop around the performence of LCDF on GSBMs is the following object, which contains all of the important information about the difficulty of testing in a GSBM.

Definition 9 (Characteristic tensor) *For a GSBM specified by $(\mu_{\boldsymbol{a}})_{\boldsymbol{a}\in[k]^p}$, we define its* characteristic tensor *(or* matrix *if $p=2$) to be $\boldsymbol{T} = \boldsymbol{T}^{(p)} \in (\mathbb{R}^{[k]\times[k]})^{\otimes p}$ having entries*

$$T_{(a_1,b_1),\dots,(a_p,b_p)} = \frac{1}{p!}\mathop{\mathbb{E}}_{y\sim\mu_{\mathsf{avg}}}\left[\left(\frac{d\mu_{(a_1,\dots,a_p)}}{d\mu_{\mathsf{avg}}}(y)-1\right)\left(\frac{d\mu_{(b_1,\dots,b_p)}}{d\mu_{\mathsf{avg}}}(y)-1\right)\right].$$

To hint at the statistical relevance of $\boldsymbol{T}$, note that the entries $T_{(a_1,a_1),\dots,(a_p,a_p)}$ are proportional to the χ^2 divergence between $\mu_{\boldsymbol{a}}$ and μ_{avg}, a measurement of how atypical the observations under $\mathbb{P}$ are compared to those under $\mathbb{Q}$. The remaining entries contain a kind of "cross–χ^2 divergence" between different $\mu_{\boldsymbol{a}}$ relative to μ_{avg}. The reader interested in these matters is directed to the discussion in Kunisky (2024); for the sake of brevity, we will not revisit those details here.

Note that, for a weakly symmetric GSBM, $\boldsymbol{T}$ is a symmetric tensor, but this would not be the case if we viewed $\boldsymbol{T}$ instead as a tensor in $(\mathbb{R}^k)^{\otimes 2p}$, not flattening (a_i, b_i) into an element of $[k]^2$.

For the next definition and several to follow, a small amount of notation for working with tensors and their "contractions" will be useful.

Definition 10 (Partial contraction) *Let $\boldsymbol{T} \in (\mathbb{R}^N)^{\otimes p}$ be a symmetric tensor and $\boldsymbol{v}_1,\dots,\boldsymbol{v}_m \in \mathbb{R}^N$ for some $1 \leq m \leq p$. We write $\boldsymbol{T}[\boldsymbol{v}_1,\dots,\boldsymbol{v}_m,\cdot,\dots,\cdot] \in (\mathbb{R}^N)^{\otimes p-m}$ for the tensor with entries*

$$(\boldsymbol{T}[\boldsymbol{v}_1,\dots,\boldsymbol{v}_m,\cdot,\dots,\cdot])_{i_1,\dots,i_{p-m}} = \sum_{j_1,\dots,j_m=1}^{N} T_{j_1,\dots,j_m,i_1,\dots,i_{p-m}}(\boldsymbol{v}_1)_{j_1}\cdots(\boldsymbol{v}_m)_{j_m}.$$

The reader familiar with tensor network notation may view this as attaching $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$ to m of the "axes" of $\boldsymbol{T}$, while leaving the remaining $p - m$ axes free.

Definition 11 (Marginal characteristic tensors) *From the characteristic tensor $\boldsymbol{T}^{(p)}$ of a GSBM as above, we further define a sequence of tensors $\boldsymbol{T}^{(p-j)} \in (\mathbb{R}^{[k]\times[k]})^{\otimes(p-j)}$ by*

$$\boldsymbol{T}^{(p-j)} = \frac{1}{k^{2j}} \boldsymbol{T}^{(p)}[\underbrace{\mathbf{1}, \dots, \mathbf{1}}_{j \text{ times}}, \underbrace{\cdot, \dots, \cdot}_{p-j \text{ times}}],$$

where $\mathbf{1}$ is the vector all of whose entries are 1 (in this case of dimension k^2).

The reason for the name is that, as one may check, $\boldsymbol{T}^{(p-j)}$ is the characteristic tensor of another GSBM, now with $(p-j)$-way interactions, parametrized by $\mu_{\boldsymbol{a}}^{(p-j)}$ formed by *marginalizing* the $\mu_{\boldsymbol{a}}$ defining the original GSBM, in the sense that

$$\mu^{(p-j)}_{a_1,\dots,a_{p-j}} = \frac{1}{k^j} \sum_{a_{p-j+1},\dots,a_p=1}^{k} \mu_{a_1,\dots,a_p}.$$

Alternatively, one samples from $\mu^{(p-j)}_{\widetilde{\boldsymbol{a}}}$ by extending $\widetilde{\boldsymbol{a}} \in [k]^{p-j}$ from a $(p-j)$-tuple to a p-tuple by adding j entries from $[k]$ uniformly at random to form $\boldsymbol{a} \in [k]^p$, and then sampling from the corresponding $\mu_{\boldsymbol{a}}$.

The basic idea that our calculations will reflect is that, whenever one observes a GSBM, one observes many nearly-independent copies of its marginal GSBMs as well. For instance, the ordinary graph SBM is a GSBM with $p = 2$ describing community structure in a random graph: each $\mu_{(a,b)}$ is a Bernoulli measure describing how much communities a and b tend to interact. The case $j = 1$ of marginalization describes the degree distributions of the vertices in different communities: in the above interpretation, to sample from $\mu^{(1)}_{(a)}$, one chooses a random $b \in [k]$, i.e., a random community, and draws from $\mu_{(a,b)}$. So, $\mu^{(1)}_{(a)}$ is another Bernoulli measure, now just describing how much community a tends to connect to *all* other communities. And indeed, when we observe a random graph, we observe the connection patterns of all n vertices to the other vertices, which look approximately like n independent draws from the $j = 1$ marginalization.

More generally, the correct intuition is that from one observation of a GSBM we get information that looks approximately like $\Theta(n^j)$ draws of the marginalization to a model of order $p - j$. This large number of draws will carry a large amount of signal for hypothesis testing, *unless* $\boldsymbol{T}^{(p-j)} = \mathbf{0}$, in which case the marginalized model is what we have called *trivial*—in the above SBM example, this happens if the average degrees of members of any particular community are the same, so just looking at degrees cannot be helpful for testing regardless of how many "effective draws" from this marginalized model we receive. By making this precise, we will see that the difficulty of detection in a GSBM is governed by the greatest amount of marginalization we can perform before reaching a vanishing characteristic tensor, to which we give the following name:

Definition 12 (Marginal order) *The* marginal order *of a non-trivial GSBM is the smallest p_* for which the marginal characteristic tensor $\boldsymbol{T}^{(p_*)} \neq \mathbf{0}$ (equivalently, for which the marginal model of order p_*, in the above sense, is non-trivial).*

By construction, $\boldsymbol{T}^{(0)}$ is the scalar zero, and thus we always have $p_* > 0$. On the other hand if $\boldsymbol{T}^{(p)} = \boldsymbol{0}$ then we must have $\mu_{(a_1,\dots,a_p)} = \mu_{\mathsf{avg}}$ for all $\boldsymbol{a}$, in which case $\mathbb{P} = \mathbb{Q}$ and the original model is trivial. So, under our assumption that a GSBM is non-trivial, we always have $p_* \leq p$ (in particular p_* is always well-defined, hence our assumption of non-triviality in the definition), and the range of possible values of the marginal order in this case is

$$1 \leq p_* \leq p.$$

We are now ready to state our main abstract results, which give simple conditions for lower bounds against LCDF for sequences of GSBMs in terms of the characteristic tensors of a model and its marginalizations.

Definition 13 (Injective norm) *For a symmetric tensor $\boldsymbol{X} \in (\mathbb{R}^N)^{\otimes p}$, its* injective norm *is*

$$\|\boldsymbol{X}\|_{\mathsf{inj}} := \max_{\substack{\boldsymbol{v} \in \mathbb{R}^N \\ \|\boldsymbol{v}\|=1}} |\langle \boldsymbol{X}, \boldsymbol{v}^{\otimes p} \rangle| = \max_{\substack{\boldsymbol{v} \in \mathbb{R}^N \\ \|\boldsymbol{v}\|=1}} \left| \sum_{i_1,\dots,i_p=1}^{d} X_{i_1,\dots,i_d} v_{i_1} \cdots v_{i_d} \right|.$$

Remark 14 *It seems likely from examining our technical calculations that the "right" quantity that should appear in these results is not the injective norm but the tensor analog of the maximum eigenvalue,*

$$\max_{\substack{\boldsymbol{v} \in \mathbb{R}^N \\ \|\boldsymbol{v}\|=1}} \langle \boldsymbol{X}, \boldsymbol{v}^{\otimes p} \rangle.$$

However, there appear to be technical obstructions to achieving this, and we have not found any examples where our current methods give loose estimates while a version improved in this way would be significantly sharper.

The shared setup for the two results below will be as follows. Let $p \geq 2$, $k \geq 1$, and Ω a measurable space. Consider a sequence of non-trivial GSBMs, with p, k, Ω fixed but with size $n \geq 1$ increasing, and $\mu_{\boldsymbol{a}} = \mu_{n,\boldsymbol{a}}$ possibly depending on n, and suppose these lead to characteristic tensors and corresponding marginalizations $\boldsymbol{T}_n^{(j)}$. Suppose every GSBM in this sequence, for sufficiently large n, has marginal order at least $p_* \in [p]$.[3] Let $\mathbb{Q} = \mathbb{Q}_n$ and $\mathbb{P} = \mathbb{P}_n$ be the sequence of probability measures over $\Omega^{\binom{[n]}{p}}$ corresponding to these GSBMs.

Theorem 15 (General lower bound for general marginal order) *Suppose that $p_* \geq 2$ in the above setting. There is a constant $c = c_{p,k}$ depending only on p and k such that, if for all sufficiently large n we have $D(n) \leq cn$ and*

$$\max_{p_* \leq j \leq p} \|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}} \leq c n^{-\frac{2p-p_*}{2}} D(n)^{-\frac{p_*-2}{2}}, \tag{3}$$

then no sequence of functions of coordinate degree at most $D(n)$ can strongly separate $\mathbb{Q}_n$ from $\mathbb{P}_n$.

3. Results that we state for lower marginal order always also apply to higher marginal order—higher marginal order is a stronger assumption, asking that more marginalized characteristic tensors be zero.

Theorem 16 (General lower bound for marginal order 2) *Suppose that $p_* = 2$ in the above setting. In this case $\boldsymbol{T}^{(p_*)} = \boldsymbol{T}^{(2)}$ is a symmetric matrix and $\|\boldsymbol{T}^{(2)}\|_{\mathsf{inj}} = \|\boldsymbol{T}^{(2)}\|$ is equivalently the operator norm. Suppose that, for constants $C > 0$ and $\varepsilon \in (0,1)$, for all sufficiently large n,*

$$\|\boldsymbol{T}_n^{(2)}\| \leq (1-\varepsilon)\frac{k^2}{p(p-1)}\frac{1}{n^{p-1}},$$
$$\|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}} \leq C\frac{1}{n^{p-1}} \text{ for all } 3 \leq j \leq p.$$

Then, no sequence of functions of coordinate degree at most $D(n) = O(n / \log n)$ can strongly separate $\mathbb{Q}_n$ from $\mathbb{P}_n$.

A few comments are in order. First, we do not explore here the issue of whether our lower bounds against LCDF are optimal. To do this, when the assumptions above are *not* satisfied, one would produce an LCDF that does strongly separate $\mathbb{Q}_n$ from $\mathbb{P}_n$. We do expect it to be possible to produce nearly optimal such LCDF *upper bounds*, and below in discussing our concrete applications we will reference works describing efficiently computable statistics (though not always ones that can clearly be approximated by LCDF) that achieve strong separation or that solve a related estimation task for many of the models we consider. Generally, one would expect such an LCDF to be obtained as a maximizer in the optimization problem defining the coordinate advantage. One drawback of our proof techniques is that they leave the description of such a maximizer (a coordinate degree version of the *low degree likelihood ratio* from the LDP theory) implicit. We will not violate this implicitness here, in order to illustrate the clean treatment it allows of lower bounds, and leave proving matching upper bounds for GSBM or for some of the concrete models discussed below to future work.

Second, at a high level, we may view these as somewhat akin to universality results like those studied in Kunisky (2024) for testing for continuous signals: the results state that, while the entire collection of channel measures $(\mu_{\boldsymbol{a}})$ is potentially a complicated object, our lower bounds depend only on the finite-dimensional characteristic tensor $\boldsymbol{T}$. We will see below some examples of different models with the same characteristic tensor, for which our results are then identical. Since these lower bounds are often tight (though our results do not imply that), it seems that categorical testing problems might have similar universality phenomena to continuous ones, though further investigation would be needed to establish that rigorously.

Next, the second result should be viewed as an elaboration of the first in the special case $p_* = 2$: in that case, the condition of the first result asks that $\|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}} \leq c/n^{p-1}$ for an unspecified constant c. The second result instead gives a condition involving a precise constant $k^2/p(p-1)$ in the bound on $\|\boldsymbol{T}_n^{(2)}\|$ (and clarifies that the other injective norms may be bounded much more crudely), which we will see often leads to tight analyses giving lower bounds precisely complementary to the best known algorithmic results.

When $p_* \geq 3$, then the condition on the injective norms in the first result depends on the choice of the degree $D(n)$. Thus in this case our lower bound will include a smooth tradeoff between the amount of signal required for testing to succeed and the available computational budget as measured by $D(n)$. As mentioned before, we may think of LCDF of coordinate degree $D(n)$ as taking roughly time $\exp(\widetilde{\Theta}(D(n)))$ to compute. When we consider $D(n)$ polynomial in n, say $D(n) \sim n^\delta$ for small δ, then, since the other term on the right-hand side of (3) is also polynomial in n, our lower bound will allow for a regime of *subexponential time* algorithms, where increasing

δ allows for testing for substantially weaker signals (as measured by the norms of the characteristic tensors). In contrast, together with prior algorithmic results, we will see that Theorem 16 for $p_* = 2$ leads to much sharper threshold phenomena: outside of a very narrow window of parameters, either a polynomial-time algorithm can solve a testing problem, or coordinate degree close to n, i.e., nearly exponential time according to the previous heuristic, is required.

This state of affairs exactly mirrors the difference in behaviors between spiked matrix and spiked tensor models. For testing between a p-ary tensor of i.i.d. Gaussian noise and one with a rank-one tensor $\lambda \boldsymbol{x}^{\otimes p}$ added, there is precisely the same kind of tradeoff between degree and signal-to-noise ratio λ as in Theorem 15. Indeed, replacing the left-hand side of (3) with λ^2 and setting $p = p_*$, one obtains exactly the statement of an optimal lower bound against LDP for the p-ary spiked tensor model![4] The reader may compare Theorem 15 with Theorem 3.3 of Kunisky et al. (2022). In contrast, for $p = 2$, this becomes a spiked matrix model, which exhibits precisely the sharper computational threshold described in our Theorem 16; cf. Theorem 3.9 of Kunisky et al. (2022).

While here we only probe this intriguing analogy from the point of view of *lower* bounds, similar parallels have also been observed, at least in specific situations, algorithmically as well. For instance, in Appendix C.3.3 we will treat the example of detecting planted solutions in random XOR-SAT formulas viewed as a GSBM, and in this case Wein et al. (2019) devised a unified spectral algorithm for both this task and an analogous tensor PCA problem. We hope that these observations will bring attention to what seems to be a high-level principle concerning these models:

"A GSBM of marginal order p_* behaves qualitatively like a spiked p_*-tensor model."

Finally, let us remark on the proof techniques, deferring full proofs to the appendices. The techniques of Kunisky (2024) imply that the performance of the best LCDF for hypothesis testing (in a certain smoothed L^2 sense) is governed by a suitable *advantage*, an expectation involving the random *overlap* of two draws of the assignment of labels in $[k]$ to the population. In GSBMs, the overlap depends on, when we draw independent collections of labels $x_1^{(1)}, \dots, x_n^{(1)}, x_1^{(2)}, \dots, x_n^{(2)} \sim \mathrm{Unif}([k])$, how many indices $i \in [n]$ have $(x_i^{(1)}, x_i^{(2)}) = (a, b)$ for each $(a, b) \in [k]^2$. Probabilistically, the advantage is then just an expectation involving a finite-dimensional multinomial random vector. We are able to avoid the high-dimensional moment expansions often used in prior work on LDP (the approach originally suggested by Hopkins (2018)) and instead to apply new non-asymptotic bounds on the tails and moments of certain functions of multinomial vectors (in particular, of the Pearson χ^2 statistics built from them) to understand the advantage precisely. See Appendix A.3 for these technical results, which may be of independent interest.

4.2. Applications: Graph Stochastic Block Models

We begin with applications to the familiar graph-valued SBMs discussed above. The following is a general form of such a model, with some special assumptions particular to our toolkit.

Definition 17 (Stochastic block model) *Suppose $k \geq 2$ and $\boldsymbol{Q} \in [0, 1]^{k \times k}_{\mathrm{sym}}$ is a symmetric matrix satisfying $\boldsymbol{Q}\mathbf{1} = \lambda \mathbf{1}$, so that $\lambda = \lambda_1(\boldsymbol{Q})$ is the largest (Perron-Frobenius) eigenvalue of $\boldsymbol{Q}$. Define*

$$q := \frac{1}{k^2} \sum_{a,b=1}^{k} Q_{ab} = \frac{1}{k^2} \mathbf{1}^\top \boldsymbol{Q} \mathbf{1} = \frac{\lambda}{k}.$$

4. For $\boldsymbol{x}$ suitably normalized.

We define the associated stochastic block model (SBM) *consisting, for each* $n \geq 1$*, of two probability measures over graphs on* n *vertices or, equivalently, their adjacency matrices,* $\boldsymbol{Y} \in \{0,1\}^{n\times n}_{\mathrm{sym}}$*:*

1. *Under* $\mathbb{Q}_n$*, draw* $Y_{ij} \sim \mathsf{Ber}(\frac{1}{n} \cdot q)$ *independently for each* $1 \leq i < j \leq n$*.*
2. *Under* $\mathbb{P}_n$*, first draw* $\boldsymbol{x} \in [k]^n$ *uniformly at random. Then, draw* $Y_{ij} \sim \mathsf{Ber}(\frac{1}{n} \cdot Q_{x_i x_j})$ *independently for each* $1 \leq i < j \leq n$*.*

We call $\boldsymbol{Q}$ *the* interaction matrix *of such a model.*

The condition that $\mathbf{1}$ is an eigenvector of $\boldsymbol{Q}$ will ensure, in the language we have introduced above, that this model has marginal order at least 2.

Theorem 18 (Lower bound for stochastic block model) *Suppose that* $\boldsymbol{Q}$ *is an interaction matrix as in Definition 17. If*

$$\max_{j \in \{2,\dots,k\}} |\lambda_j(\boldsymbol{Q})|^2 < k\lambda_1(\boldsymbol{Q}), \tag{4}$$

then no sequence of functions of coordinate degree $O(n \,/\, \log n)$ *can strongly separate* $\mathbb{Q}_n$ *from* $\mathbb{P}_n$ *in the SBM associated to* $\boldsymbol{Q}$*.*

The threshold (4) is a general form of the *Kesten-Stigum (KS) threshold* for SBMs studied extensively in previous literature and shown to be the threshold where numerous algorithms cease to succeed at testing Decelle et al. (2011a,b); Bordenave et al. (2015); Abbe and Sandon (2018); see also discussion in Abbe (2017). Theorem 18 thus gives a new complementary lower bound to all of these results. It is also known (and discussed in the references) that, once $k \geq 4$, there are inefficient algorithms that can test below this threshold. In these cases, Theorem 18 also gives new evidence for a *statistical-to-computational gap*, a parameter regime where the testing problem is possible to solve but only by an inefficient computation.

Example 1 *Consider the* symmetric SBM*, which corresponds to the choice*

$$Q_{ab} = \left\{ \begin{array}{ll} \alpha & \text{if } a = b, \\ \beta & \text{if } a \neq b \end{array} \right\},$$

for some $\alpha, \beta \geq 0$*. The condition (4) then reduces to the condition* $\frac{(\alpha-\beta)^2}{k(\alpha+(k-1)\beta)} < 1$*, a widely-studied special case of the KS threshold. In this case, the result of Theorem 18 was obtained by Bandeira et al. (2021), but the general case (to the best of our knowledge) has not been treated.*

The proofs of these results proceed just by computing characteristic tensors (matrices in this case) in terms of the interaction matrix and applying the previous general results. The same is true in all applications below, and we do not comment further on these proofs.

4.3. Applications: Hypergraph Stochastic Block Models

We next consider a version of the above results for SBMs defined over hypergraphs rather than graphs. These results will actually strictly generalize the preceding ones, but we state and prove them separately for the sake of exposition to the reader more familiar with the graph SBM literature.

Definition 19 (Hypergraph stochastic block model) *Suppose $k \geq 2$, $p \geq 3$, and $\boldsymbol{Q} \in ([0,1]^k)^{\otimes p}$ is a symmetric tensor satisfying*

$$\boldsymbol{Q}[\mathbf{1}, \dots, \mathbf{1}, \cdot\,] = \lambda \mathbf{1}.$$

Define

$$q := \frac{1}{k^p} \sum_{a_1, \dots, a_p = 1}^{k} Q_{a_1 \cdots a_p} = \frac{1}{k^p} \langle \boldsymbol{Q}, \mathbf{1}^{\otimes p} \rangle = \frac{\lambda}{k^{p-1}}.$$

Associated to such a tensor, we define the hypergraph stochastic block model (HSBM) *consisting, for each $n \geq 1$, of two probability measures over p-uniform hypergraphs on n vertices or, equivalently, their symmetric adjacency tensors, $\boldsymbol{Y} \in (\{0,1\}^n)^{\otimes p}$:*

1. *Under $\mathbb{Q}_n$, draw $Y_{i_1 \cdots i_p} \sim \mathsf{Ber}(q \,/\, \binom{n}{p-1})$ independently for each $1 \leq i_1 < \cdots < i_p \leq n$.*
2. *Under $\mathbb{P}_n$, first draw $\boldsymbol{x} \in [k]^n$ uniformly. Then, draw $Y_{i_1 \cdots i_p} \sim \mathsf{Ber}(Q_{x_{i_1} \cdots x_{i_p}} \,/\, \binom{n}{p-1})$ independently for each $1 \leq i_1 < \cdots < i_p \leq n$.*

We call $\boldsymbol{Q}$ the interaction tensor *of such a model.*

Theorem 20 (Lower bound for hypergraph stochastic block model) *Suppose that $\boldsymbol{Q}$ is an interaction tensor as in Definition 19. If*

$$\max_{j \in \{2, \dots, k\}} |\lambda_j(\boldsymbol{Q}[\mathbf{1}, \dots, \mathbf{1}, \cdot, \cdot])|^2 < \frac{k^{p-1}}{p-1} \lambda_1(\boldsymbol{Q}[\mathbf{1}, \dots, \mathbf{1}, \cdot, \cdot]), \tag{5}$$

then no sequence of functions of coordinate degree $O(n \,/\, \log n)$ can strongly separate $\mathbb{Q}_n$ from $\mathbb{P}_n$ in the HSBM associated to $\boldsymbol{Q}$.

The threshold in (5) is again a generalized form of the KS threshold. The hypergraph versions of these thresholds are newer, but conjectures and algorithmic results concerning such thresholds have appeared in Angelini et al. (2015); Pal and Zhu (2021); Chodrow et al. (2023); Stephan and Zhu (2024). Evidence of statistical-to-computational gaps for certain choices of parameters has also accumulated Angelini et al. (2015); Gu and Polyanskiy (2023); Gu and Pandey (2024). As in the graph case, Theorem 20 gives new evidence both for the optimality of algorithms reaching this generalized KS threshold, and for the existence of statistical-to-computational gaps when inefficient algorithms are known to succeed beyond the KS threshold.

Example 2 *As another concrete example, consider the* symmetric HSBM, *which has*

$$Q_{a_1 \cdots a_p} = \left\{ \begin{array}{ll} \alpha & \text{if } a_1 = \cdots = a_p, \\ \beta & \text{otherwise} \end{array} \right\}.$$

In this case, the condition (5) becomes $\frac{(p-1)(\alpha-\beta)^2}{k^{p-1}(\alpha+(k^{p-1}-1)\beta)} < 1$. To the best of our knowledge, even in this special case lower bounds against LDP or LCDF were not known previously for $p \geq 3$.

We do not give details here, but we discuss an example of a model with higher marginal order and therefore with a different profile of tradeoff between LCDF degree and signal-to-noise ratio, namely a model implementing random XOR-SAT formulas in a GSBM, in Appendix C.3.3. To summarize briefly, in this model there is only a "soft" computational threshold, where one may pay slightly more in runtime cost to achieve detection for slightly weaker signals, throughout a large regime of polynomial and subexponential runtimes. In particular, it does not make sense to ask for a KS-like computational threshold that is specified up to precise multiplicative constants. Still, our lower bounds match the performance and runtime of the families of algorithms proposed for this problem by Raghavendra et al. (2017); Wein et al. (2019).

4.4. Applications: Group Problems

In these next results, we will always consider discrete models associated to a group G, where we write $k = |G| > 1$ (which will be the same as the k in a GSBM). That is because our GSBMs will always be labelled by elements of G, and indeed our observations will also always be elements of G, so we will have $k = \ell = |G|$. Both problems we study will have $p = 2$, i.e., binary interactions.

Definition 21 (Truth-or-Haar synchronization) *Given a group G as above and a constant $\gamma > 0$, the* truth-or-Haar synchronization model *associated to G consists, for each $n \geq 1$, of the following pairs of probability measures over $\boldsymbol{Y} \in G^{\binom{n}{2}}$:*

1. *Under $\mathbb{Q}_n$, draw $Y_{ij} \sim \mathrm{Unif}(G)$ independently for each $1 \leq i < j \leq n$.*
2. *Under $\mathbb{P}_n$, first draw $x_1, \dots, x_n \sim \mathrm{Unif}(G)$ uniformly at random. Then, draw Y_{ij} independently for each $1 \leq i < j \leq n$ by setting it to $g_i g_j^{-1}$ with probability $\frac{\gamma}{\sqrt{n}}$ and setting it to a draw from $\mathrm{Unif}(G)$ with probability $1 - \frac{\gamma}{\sqrt{n}}$.*

Theorem 22 (Lower bound for synchronization) *For any finite group G, if $\gamma < 1$, then no sequence of functions of coordinate degree $O(n \,/\, \log n)$ can strongly separate $\mathbb{Q}_n$ from $\mathbb{P}_n$ in the truth-or-Haar group synchronization model associated to G.*

This lends new support to a prediction of Singer (2011) and, together with the results of Perry et al. (2016), gives evidence for a statistical-to-computational gap in these models once $|G| \geq 11$, in which case there is an inefficient algorithm that succeeds at testing for some values of $\gamma < 1$.

The following model is a curious variant of the above which, while its algebraic structure is quite different, at least for one class of G shares precisely the same analysis.

Definition 23 (Truth-or-Haar noisy sumset) *Given a group G as above and a constant $\gamma > 0$, the* truth-or-Haar noisy sumset model *associated to G is identical to the synchronization model, but with $g_i g_j^{-1}$ replaced by $g_i g_j$ (in the definition of $\mathbb{P}_n$).*

Theorem 24 (Lower bound for noisy sumset) *For any finite abelian group G, if $\gamma < 1$, then no sequence of functions of coordinate degree $O(n \,/\, \log n)$ can strongly separate $\mathbb{Q}_n$ from $\mathbb{P}_n$ in the truth-or-Haar noisy sumset model associated to G.*

The requirement that G be abelian is due to the requirement of our methods that a GSBM be weakly symmetric, which otherwise would not be satisfied. It is an interesting question whether the same result will hold for arbitrary G; it seems that it should by analogy with the synchronization problem, but the analysis would require different and more problem-specific calculations.

In Appendix C.4.3, we will discuss how our technical results can also be used to sharpen the recent results of Kireeva et al. (2024) on a related synchronization problem with a quite different noise model of adding Gaussian noise to a certain matrix associated to the synchronization task. We improve the allowable degree in this lower bound from the original $D(n) \sim n^{1/3}$ to $D(n) \sim n$.

4.5. Channel Calculus

Finally, in the spirit of analogous results derived in Kunisky (2024), we give two general principles describing how our lower bounds against LCDF transform under certain deformations of a GSBM.

Definition 25 (Channel resampling) *Consider a GSBM as in Definition 1, associated to a family of measures* $(\mu_{\boldsymbol{a}})_{\boldsymbol{a}\in[k]^p}$ *whose average is* μ_{avg}. *The* η-resampling *of this GSBM is that associated instead to the family* $((1-\eta)\mu_{\boldsymbol{a}} + \eta\mu_{\mathsf{avg}})_{\boldsymbol{a}\in[k]^p}$. *In words,* $\mathbb{Q}$ *of the* η*-resampled GSBM remains the same, while in order to sample from* $\mathbb{P}$*, we first sample from the original GSBM, and then for each observation independently with probability* η *independently replace it with a draw from* μ_{avg}.

Theorem 26 (Resampling and characteristic tensors) *If* $\boldsymbol{T}^{(j)}$ *is a given marginalization of the characteristic tensor of an GSBM, then the same marginalization of the characteristic tensor of the* η*-resampling of that GSBM is* $(1-\eta)^2\,\boldsymbol{T}^{(j)}$.

In the same vein, and directly akin to one of the results in Kunisky (2024) for continuous signals, we may also understand the effect of the following operation on a channel.

Definition 27 (Channel censorship) *Consider a GSBM as in Definition 1, associated to a family of measures* $(\mu_{\boldsymbol{a}})_{\boldsymbol{a}\in[k]^p}$*, whose average is* μ_{avg}*, over a measurable space* Ω*. Let "* $\bullet$*" denote a new symbol that does not belong to* Ω*. The* η-censorship *of this GSBM is that associated instead to the family* $((1-\eta)\mu_{\boldsymbol{a}} + \eta\delta_\bullet)_{\boldsymbol{a}\in[k]^p}$*, defined on the measurable space* $\Omega \sqcup \{\bullet\}$*, where* $\delta_\bullet$ *denotes the Dirac mass on this new element. In words, to sample from either* $\mathbb{Q}$ *or* $\mathbb{P}$ *under the* η*-censored GSBM, we first sample from the corresponding distribution in the original GSBM, and then for each observation independently with probability* η *replace it with* $\bullet$.

Remarkably, we find that censorship and resampling have almost the same effect on the characteristic tensor, and thus the aforementioned effective SNR, of a GSBM. As we would expect, resampling at a given rate reduces the SNR more than censorship does, but given any rate of resampling, there is a (greater) rate of censorship that produces a model that, from the point of view of our analysis, is equivalent to the resampled one.

Theorem 28 (Censorship and characteristic tensors) *If* $\boldsymbol{T}^{(j)}$ *is the characteristic tensor of a given order of marginalization of a GSBM, then the same marginalization of the characteristic tensor of the* η*-censorship of that GSBM is* $(1-\eta)\,\boldsymbol{T}^{(j)}$.

Both results follow from very simple calculations with characteristic tensors. Even so, in light of our general results above, these both can be useful for working with specific models. For instance, these results together with our analysis of SBMs allow us to immediately obtain various lower bounds for the *censored SBMs* studied by Abbe et al. (2014); Saade et al. (2015).

Acknowledgments

Thanks to Afonso Bandeira and Anastasia Kireeva for helpful discussions about group synchronization, to Yuzhou Gu for discussions, references, and suggestions about hypergraph stochastic block models and their Kesten-Stigum thresholds, and to Yuval Ishai for suggesting (a different version of) the noisy group sumset problem. Part of this work was done while the author was employed by the Department of Computer Science of Yale University. Inspiration for parts of this work was also drawn from discussions at the Workshop on Computational Complexity of Statistical Inference at the Banff International Research Station in February 2024.

References

Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Emmanuel Abbe and Colin Sandon. Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics*, 71(7):1334–1406, 2018.

Emmanuel Abbe, Afonso S Bandeira, Annina Bracher, and Amit 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.

Maria Chiara Angelini, Francesco Caltagirone, Florent Krzakala, and Lenka Zdeborová. Spectral detection on sparse hypergraphs. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 66–73. IEEE, 2015.

Afonso S Bandeira, Dmitriy Kunisky, and Alexander S Wein. Computational hardness of certifying bounds on constrained PCA problems. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151, pages 78:1–78:29, 2020. doi: 10.4230/LIPIcs.ITCS.2020.78.

Afonso S Bandeira, Jess Banks, Dmitriy Kunisky, Cristopher Moore, and Alexander S Wein. Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. In *34th Annual Conference on Learning Theory (COLT 2021)*, pages 410–473. PMLR, 2021.

Afonso S Bandeira, Ahmed El Alaoui, Samuel B Hopkins, Tselil Schramm, Alexander S Wein, and Ilias Zadik. The Franz-Parisi criterion and computational trade-offs in high dimensional statistics. *arXiv preprint arXiv:2205.09727*, 2022.

Jess Banks, Cristopher Moore, Roman Vershynin, Nicolas Verzelen, and Jiaming Xu. Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Transactions on Information Theory*, 64(7):4872–4894, 2018.

Boaz Barak, Samuel B Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357. IEEE, 2015.

Matthew Brennan, Guy Bresler, Samuel B Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low-degree tests are almost equivalent. *arXiv preprint arXiv:2009.06107*, 2020.

Philip Chodrow, Nicole Eikmeier, and Jamie Haddock. Nonbacktracking spectral clustering of nonuniform hypergraphs. *SIAM Journal on Mathematics of Data Science*, 5(2):251–279, 2023.

Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, Alexander S Wein, and Ilias Zadik. Statistical and computational phase transitions in group testing. In *Conference on Learning Theory*, pages 4764–4781. PMLR, 2022.

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011a.

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6): 065701, 2011b.

Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Subexponential-time algorithms for sparse PCA. *Foundations of Computational Mathematics*, pages 1–50, 2023.

IJ Good, TN Gover, and GJ Mitchell. Exact distributions for χ^2 and for the likelihood-ratio statistic for the equiprobable multinomial distribution. *Journal of the American Statistical Association*, 65(329):267–283, 1970.

Yuzhou Gu and Aaradhya Pandey. Community detection in the hypergraph stochastic block model and reconstruction on hypertrees. *arXiv preprint arXiv:2402.06856*, 2024.

Yuzhou Gu and Yury Polyanskiy. Weak recovery threshold for the hypergraph stochastic block model. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 885–920. PMLR, 2023.

Simon Heimlicher, Marc Lelarge, and Laurent Massoulié. Community detection in the labelled stochastic block model. *arXiv preprint arXiv:1209.2910*, 2012.

Golde I Holtzman and IJ Good. The Poisson and chi-squared approximations as compared with the true upper-tail probability of Pearson's chi-squared for equiprobable multinomials. *Journal of Statistical Planning and Inference*, 13:283–295, 1986.

Samuel B Hopkins. *Statistical inference and the sum of squares method*. PhD thesis, Cornell University, 2018.

Samuel B Hopkins and David Steurer. Efficient Bayesian estimation from few samples: community detection and related problems. In *58th Annual Symposium on Foundations of Computer Science (FOCS 2017)*, pages 379–390. IEEE, 2017.

Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *58th Annual Symposium on Foundations of Computer Science (FOCS 2017)*, pages 720–731, 2017.

Anastasia Kireeva, Afonso S Bandeira, and Dmitriy Kunisky. Computational lower bounds for multi-frequency group synchronization. *arXiv preprint arXiv:2406.03424*, 2024.

Frederic Koehler and Elchanan Mossel. Reconstruction on trees and low-degree polynomials. *arXiv preprint arXiv:2109.06915*, 2021.

Pravesh K Kothari, Santosh S Vempala, Alexander S Wein, and Jeff Xu. Is planted coloring easier than planted clique? In *36th Annual Conference on Learning Theory (COLT 2023)*. PMLR, 2023.

Dmitriy Kunisky. Hypothesis testing with low-degree polynomials in the Morris class of exponential families. In *34th Annual Conference on Learning Theory (COLT 2021)*, pages 2822–2848. PMLR, 2021a.

Dmitriy Kunisky. *Spectral Barriers in Certification Problems*. PhD thesis, New York University, 2021b.

Dmitriy Kunisky. Low coordinate degree algorithms I: Universality of computational thresholds for hypothesis testing. *arXiv preprint arXiv:2403.07862*, 2024.

Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In Paula Cerejeiras and Michael Reissig, editors, *Mathematical Analysis, its Applications and Computation*, pages 1–50, Cham, 2022. Springer International Publishing.

Marc Lelarge, Laurent Massoulié, and Jiaming Xu. Reconstruction in the labelled stochastic block model. *IEEE Transactions on Network Science and Engineering*, 2(4):152–163, 2015.

T Lewis, IW Saunders, and M Westcott. The moments of the Pearson chi-squared statistic and the minimum expected value in two-way tables. *Biometrika*, 71(3):515–522, 1984.

Diego Martinez-Taboada and Aaditya Ramdas. Empirical Bernstein in smooth Banach spaces. *arXiv preprint arXiv:2409.06060*, 2024.

Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors. In *Advances in Neural Information Processing Systems*, pages 217–225, 2015.

Cristopher Moore. The computer science and physics of community detection: landscapes, phase transitions, and hardness. *arXiv preprint arXiv:1702.00467*, 2017.

Soumik Pal and Yizhe Zhu. Community detection in the sparse hypergraph stochastic block model. *Random Structures & Algorithms*, 59(3):407–463, 2021.

Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of PCA for spiked random matrices and synchronization. *arXiv preprint arXiv:1609.05573*, 2016.

Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random CSPs below the spectral threshold. In *49th Annual ACM SIGACT Symposium on Theory of Computing (STOC 2017)*, pages 121–131. ACM, 2017.

Cynthia Rush, Fiona Skerman, Alexander S Wein, and Dana Yang. Is it easier to count communities than find them? In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, 2023.

Alaa Saade, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Spectral detection in the censored block model. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1184–1188. IEEE, 2015.

Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1):20–36, 2011.

Ludovic Stephan and Yizhe Zhu. Sparse random hypergraphs: Non-backtracking spectra and community detection. *Information and Inference: A Journal of the IMA*, 13(1):iaae004, 2024.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*. Cambridge University Press, 2018.

Alexander S Wein, Ahmed El Alaoui, and Cristopher Moore. The Kikuchi hierarchy and tensor PCA. *arXiv preprint arXiv:1904.03858*, 2019.

Justin Whitehouse, Zhiwei Steven Wu, and Aaditya Ramdas. Time-uniform self-normalized concentration for vector-valued processes. *arXiv preprint arXiv:2310.09100*, 2023.

Jiaming Xu, Laurent Massoulié, and Marc Lelarge. Edge label inference in generalized stochastic block models: from spectral theory to impossibility results. In *Conference on Learning Theory*, pages 903–920. PMLR, 2014.

Appendix A. Preliminaries

A.1. Notation

The asymptotic notations $o(\cdot), O(\cdot), \omega(\cdot), \Omega(\cdot), \Theta(\cdot)$ have their usual meanings and always refer to the limit $n \to \infty$. When these symbols have tildes on top (like $\widetilde{\Theta}$), then polylogarithmic factors in n are suppressed.

For a symmetric $k \times k$ matrix $\boldsymbol{M}$, $\lambda_1(\boldsymbol{M}) \geq \cdots \geq \lambda_k(\boldsymbol{M})$ are the ordered eigenvalues of $\boldsymbol{M}$.

$\mathsf{Ber}(p)$ denotes the Bernoulli probability measure of a random variable equal to 1 with probability p and 0 with probability $1-p$. $\mathrm{Unif}(S)$ denotes the uniform probability measure over a finite set S. For two probability measures μ, ρ and $\eta \in [0,1]$, $\eta\mu + (1-\eta)\rho$ denotes the corresponding mixture probability measure.

A.2. Sharp Vector Bernstein Inequality

We start with a seemingly unrelated topic, a specific version of a Bernstein-type concentration inequality over vectors. The point of this is that we will need to understand the concentration and moments of quantities like $\|\boldsymbol{z} - \mathbb{E}\boldsymbol{z}\|^2$, where $\boldsymbol{z} \in \mathbb{N}^k$ is the vector of numbers of elements of $[n]$ assigned each label from $[k]$ in a GSBM. The prior work of Kireeva et al. (2024) which arrives at a similar issue handled this by direct combinatorial means. We will be able to improve and simplify that analysis by instead understanding this via such a Bernstein inequality.

The kind of inequality that we need is known, if somewhat implicitly, as we will discuss below. However, we give a self-contained proof because a rather fine detail of this particular result will be important for us. In general, like the scalar Bernstein inequality, vector Bernstein inequalities state that $\|\sum_{i=1}^n \boldsymbol{v}_i\|$ for i.i.d. centered random vectors $\boldsymbol{v}_i$ has subgaussian tails up to a certain size of fluctuation and exponential tails beyond that size. It will be crucial in several of our applications (both sharpening the results of Kireeva et al. (2024) and in our analysis of models with marginal order 2 and thus having sharp computational thresholds) that we have sharp control over the variance proxy in this subgaussian bound on "somewhat large" deviations. As we discuss after the proof, this result achieves that, while similar results obtained by other means do not.

Lemma 29 *Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in \mathbb{R}^d$ be independent random vectors such that:*

1. *$\mathbb{E}\boldsymbol{v}_i = 0$ for all i,*
2. *$\|\operatorname{Cov}(\boldsymbol{v}_i)\| \leq \sigma^2$ for all i, and*
3. *$\|\boldsymbol{v}_i\| \leq M$ for all i, almost surely.*

Then, for any $\varepsilon \in (0,1)$, we have the tail bound

$$\mathbb{P}\left[\left\|\sum_{i=1}^n \boldsymbol{v}_i\right\| \geq t\right] \leq \left(1 + \frac{2}{\varepsilon}\right)^d \exp\left(-\frac{t^2}{2\frac{\sigma^2}{(1-\varepsilon)^2}n + \frac{2}{3}\frac{M}{1-\varepsilon}t}\right).$$

Proof Let $\mathcal{X}$ be an ε-net of the unit sphere in $\mathbb{R}^d$. As is well-known, one may take $|\mathcal{X}| \leq (1 + \frac{2}{\varepsilon})^d$ (see, e.g., (Vershynin, 2018, Corollary 4.2.13)). For any $\boldsymbol{v} \in \mathbb{R}^d$, we then have $\max_{\boldsymbol{x} \in \mathcal{X}} \langle \boldsymbol{v}, \boldsymbol{x} \rangle \leq \|\boldsymbol{v}\|$, as well as, for the $\boldsymbol{y} \in \mathcal{X}$ for which $\|\boldsymbol{y} - \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}\| \leq \varepsilon$, that

$$\begin{aligned}
\max_{\boldsymbol{x} \in \mathcal{X}} \langle \boldsymbol{v}, \boldsymbol{x} \rangle &\geq \langle \boldsymbol{y}, \boldsymbol{v} \rangle \\
&= \|\boldsymbol{v}\| + \left\langle \boldsymbol{y} - \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}, \boldsymbol{v} \right\rangle \\
&\geq \|\boldsymbol{v}\| - \varepsilon\|\boldsymbol{v}\|,
\end{aligned}$$

and thus in summary we have

$$\max_{\boldsymbol{x} \in \mathcal{X}} \langle \boldsymbol{v}, \boldsymbol{x} \rangle \leq \|\boldsymbol{v}\| \leq \frac{1}{1-\varepsilon} \max_{\boldsymbol{x} \in \mathcal{X}} \langle \boldsymbol{v}, \boldsymbol{x} \rangle.$$

Taking a union bound, we then have

$$\begin{aligned}
\mathbb{P}\left[\left\|\sum_{i=1}^n \boldsymbol{v}_i\right\| \geq t\right] &\leq \mathbb{P}\left[\max_{\boldsymbol{x} \in \mathcal{X}} \left\langle \sum_{i=1}^n \boldsymbol{v}_i, \boldsymbol{x} \right\rangle \geq (1-\varepsilon)t\right] \\
&\leq \sum_{\boldsymbol{x} \in \mathcal{X}} \mathbb{P}\left[\sum_{i=1}^n \langle \boldsymbol{v}_i, \boldsymbol{x} \rangle \geq (1-\varepsilon)t\right]
\end{aligned}$$

where the $\langle \boldsymbol{v}_i, \boldsymbol{x} \rangle$ are independent centered scalar random variables, which are bounded by $|\langle \boldsymbol{v}_i, \boldsymbol{x} \rangle| \leq \|\boldsymbol{v}_i\| \leq M$ almost surely, and which have variance $\operatorname{Var}[\langle \boldsymbol{v}_i, \boldsymbol{x} \rangle] = \boldsymbol{x}^\top \operatorname{Cov}(\boldsymbol{v}_i)\boldsymbol{x} \leq \|\operatorname{Cov}(\boldsymbol{v}_i)\| \leq \sigma^2$. Thus, the scalar Bernstein inequality applies, giving

$$\leq |\mathcal{X}| \exp\left(-\frac{t^2}{2\frac{\sigma^2}{(1-\varepsilon)^2}n + \frac{2}{3}\frac{M}{1-\varepsilon}t}\right),$$

and plugging in the bound for $|\mathcal{X}|$ completes the proof. ∎

One simple other way to obtain such a result is via symmetrization together with the Khintchine-Kahane inequality to show that $\|\sum_{i=1}^n \boldsymbol{v}_i\|$ has suitably decaying moments to show a similar inequality. Another approach is to construct and analyze the Doob martingale associated to this function of the independent random variables $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$, using the Azuma-Hoeffding inequality or similar tools. Both of these approaches, though, end up with a bound where $\sigma^2 = \|\operatorname{Cov}(\boldsymbol{v}_i)\|$ is replaced by $\operatorname{Tr}(\operatorname{Cov}(\boldsymbol{v}_i))$, which can be larger by up to a factor of d. This would be unacceptable for our purposes, so we need specifically the above version. Such results have appeared in the literature before; e.g., they are mentioned in Whitehouse et al. (2023); Martinez-Taboada and Ramdas (2024), but we hope to draw attention to this argument in the above simple and straightforward setting.

A.3. Tails and Moments of Pearson's χ^2 Statistic

We now give the application of the above inequality to certain random variables constructed from multinomial vectors. It turns out, though this connection will not play an important role for us, that these equivalently have the distribution of *Pearson's χ^2 statistics* from classical statistics theory.

Definition 30 *We denote by $\mathsf{Mult}(n, d)$ the multinomial distribution, the law of the random vector $(z_1, \dots, z_d) \in \mathbb{Z}^d_{\geq 0}$ giving the number of balls in each of d bins after throwing n balls into a bin independently and uniformly at random.*

Definition 31 *For $\boldsymbol{z} \sim \mathsf{Mult}(n, d)$, we denote by $\chi^2_{\mathsf{Pear}}(n, d)$ the* Pearson's χ^2 distribution*, the law of the random variable $\frac{d}{n}\sum_{i=1}^d (z_i - \frac{n}{d})^2 = \frac{d}{n}\|\boldsymbol{z} - \mathbb{E}\boldsymbol{z}\|^2$.*

It is a classical fact following from basic vector-valued central limit theorems that, for fixed d and $n \to \infty$, we have the convergence in distribution

$$\sqrt{\frac{d}{n}}\left(\boldsymbol{z} - \frac{n}{d}\mathbf{1}\right) \xrightarrow{\text{(law)}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{I}_d - \frac{1}{d}\mathbf{1}_d\mathbf{1}_d^\top\right),$$

where the covariance is the orthogonal projection to the orthogonal complement of the $\mathbf{1}_d$ direction. In particular, $X \sim \chi^2_{\mathsf{Pear}}$ then converges in distribution to $\chi^2(d-1)$, the χ^2 law with $d-1$ degrees of freedom.

Our goal here will be to show that, even non-asymptotically, the tails and moments of χ^2_{Pear} resemble those of $\chi^2(d-1)$. While the arguments are simple once we are aware of the sharp vector Bernstein inequality of Lemma 29, to the best of our knowledge these kinds of results have not appeared in the statistics literature before, even though the accuracy of the approximation $\chi^2_{\mathsf{Pear}} \approx \chi^2(d-1)$ has been discussed at length (Good et al., 1970; Lewis et al., 1984; Holtzman and Good, 1986).

Lemma 32 (Right tail of χ^2_{Pear}) *For any $n, d \geq 1$, $t \geq 0$, and $\varepsilon \in (0, 1)$,*

$$\mathop{\mathbb{P}}_{X \sim \chi^2_{\mathsf{Pear}}(n,d)}[X \geq t] \leq \left(1 + \frac{2}{\varepsilon}\right)^d \exp\left(-\frac{\frac{1}{2}(1-\varepsilon)^2 t}{1 + \frac{1-\varepsilon}{3}\sqrt{\frac{d-1}{n}}\sqrt{t}}\right).$$

We note that when d is fixed, ε is small, and $t = o(n)$, then we approximately recover the tail behavior of a χ^2 distribution, whose density is proportional to $\exp(-x/2)$ up to polynomial factors in x.

Proof Write $\boldsymbol{e}_1, \dots, \boldsymbol{e}_d \in \mathbb{R}^d$ for the standard basis of $\mathbb{R}^d$, and let $i_1, \dots, i_n \sim \mathrm{Unif}([d])$. Letting $\boldsymbol{z} := \sum_{a=1}^n \boldsymbol{e}_{i_a}$, we have that the law of $\boldsymbol{z}$ is $\mathsf{Mult}(n, d)$. This realizes the multinomial distribution as a sum of i.i.d. random vectors, to which we will apply our vector Bernstein inequality.

In particular, writing $\boldsymbol{v}_a := \boldsymbol{e}_{i_a} - \frac{1}{d}\boldsymbol{1}_d = \boldsymbol{e}_{i_a} - \mathbb{E}[\boldsymbol{e}_{i_a}]$, we have that $\frac{d}{n}\|\sum_{a=1}^n \boldsymbol{v}_a\|^2$ has the law $\chi^2_{\mathsf{Pear}}(n, d)$. We also have $\mathbb{E}\boldsymbol{v}_a = \boldsymbol{0}$, $\mathrm{Cov}(\boldsymbol{v}_a) = \frac{1}{d}\boldsymbol{I}_d - \frac{1}{d^2}\boldsymbol{1}_d\boldsymbol{1}_d^\top$ whereby $\|\mathrm{Cov}(\boldsymbol{v}_a)\| = \frac{1}{d}$, and $\|\boldsymbol{v}_a\|^2 = 1 - \frac{1}{d} = \frac{d-1}{d}$ almost surely.

Plugging this information into Lemma 29, we find that, for any $\varepsilon \in (0, 1)$,

$$\begin{aligned}
\mathop{\mathbb{P}}_{X \sim \chi^2_{\mathsf{Pear}}(n,d)}[\sqrt{X} \geq t] &= \mathbb{P}\left[\left\|\sum_{a=1}^n \boldsymbol{v}_a\right\| \geq \sqrt{\frac{n}{d}}t\right] \\
&\leq \left(1 + \frac{2}{\varepsilon}\right)^d \exp\left(-\frac{\frac{n}{d}t^2}{2\frac{1}{(1-\varepsilon)^2}\frac{n}{d} + \frac{2}{3}\frac{1}{1-\varepsilon}\sqrt{\frac{n}{d}}\sqrt{\frac{d-1}{d}}t}\right),
\end{aligned}$$

and rearranging gives the result. ∎

Corollary 33 (Moments of χ^2_{Pear}) *For any $n, d \geq 1$ integers, $r \geq 1$, and $\delta, \varepsilon \in (0, 1)$,*

$$\mathop{\mathbb{E}}_{X \sim \chi^2_{\mathsf{Pear}}(n,d)}[X^r] \leq 2r\left(1 + \frac{2}{\varepsilon}\right)^d \left[\left(\frac{1 + \sqrt{\delta d}}{(1-\varepsilon)^2}\right)^r 2^r\Gamma(r) + \left(\frac{\frac{4}{\delta} + 4d}{(1-\varepsilon)^2}\right)^r \frac{\Gamma(2r)}{n^r}\right].$$

Proof We start by integrating the tail bound:

$$\begin{aligned}
\mathop{\mathbb{E}}_{X \sim \chi^2_{\mathsf{Pear}}(n,d)}[X^r] &= \int_0^\infty \mathbb{P}[X^r > t]\,dt \\
&= \int_0^\infty \mathbb{P}[\sqrt{X} > t^{\frac{1}{2r}}]\,dt \\
&\leq \left(1 + \frac{2}{\varepsilon}\right)^d \int_0^\infty \exp\left(-\frac{\frac{1}{2}(1-\varepsilon)^2 t^{\frac{1}{r}}}{1 + \frac{1-\varepsilon}{\sqrt{2}}\sqrt{\frac{d}{n}}t^{\frac{1}{2r}}}\right)\,dt \\
&= \left(1 + \frac{2}{\varepsilon}\right)^d \cdot r\left(\frac{2}{(1-\varepsilon)^2}\right)^r \int_0^\infty \exp\left(-\frac{s}{1 + \sqrt{\frac{d}{n}s}}\right) s^{r-1}\,ds.
\end{aligned}$$

For the remaining integral, let us fix $\delta > 0$ as in the statement, and break the integral up into two parts at δn:

$$\begin{aligned}
&\int_0^\infty \exp\left(-\frac{s}{1 + \sqrt{\frac{d}{n}s}}\right) s^{r-1}\,ds \\
&= \int_0^{\delta n} \exp\left(-\frac{s}{1 + \sqrt{\frac{d}{n}s}}\right) s^{r-1}\,ds + \int_{\delta n}^\infty \exp\left(-\frac{s}{1 + \sqrt{\frac{d}{n}s}}\right) s^{r-1}\,ds \\
&\leq \int_0^\infty \exp\left(-\frac{s}{1 + \sqrt{\delta d}}\right) s^{r-1}\,ds + \int_0^\infty \exp\left(-\frac{\sqrt{s}}{\sqrt{\frac{1}{\delta n}} + \sqrt{\frac{d}{n}}}\right) s^{r-1}\,ds
\end{aligned}$$

and both remaining integrals may be viewed as evaluations of the Γ function, the second one after a further substitution $s = r^2$, giving

$$= (1 + \sqrt{\delta d})^r\, \Gamma(r) + 2\left(\frac{1}{\sqrt{\delta}} + \sqrt{d}\right)^{2r} \frac{\Gamma(2r)}{n^r}.$$

The result as stated is obtained by a few elementary bounds to simplify this expression. ■

The formula above is complicated but the point is simple: when δ, ε are both small and $r = o(n)$, then the prefactors and the second term are negligible, and the expression becomes $O_d(r2^r\Gamma(r))$, which is same scale as the rth moment of $\chi^2(d-1)$. We express this in the following corollary before giving the proof.

Corollary 34 (Simplified moment bound) *For all $\varepsilon > 0$, there exist constants $C, \gamma > 0$ such that, for all $n \geq 1$ and $1 \leq r \leq \gamma n$, we have*

$$\mathop{\mathbb{E}}_{X \sim \chi^2_{\mathsf{Pear}}(n,d)}[X^r] \leq r^{3/2} C^d \left(\frac{(2+\varepsilon)r}{e}\right)^r.$$

Proof Choosing δ, ε in the previous Corollary appropriately and using Stirling's approximation $\Gamma(r) \lesssim \sqrt{r}(r/e)^r$, we find that there exists C such that, for all $n \geq 1$ an integer and $r \geq 1$ real, we have

$$\mathop{\mathbb{E}}_{X \sim \chi^2_{\mathsf{Pear}}(n,d)}[X^r] \leq r^{3/2} C^d \left[\left(\frac{(2+\varepsilon)r}{e}\right) + \left(\frac{Cr^2}{n}\right)^r\right].$$

But now, choosing $\gamma \leq \frac{2}{Ce}$, if $r \leq \gamma n$, then $r/n \leq \gamma$, so the second term above is smaller than the first, and the result follows (taking C slightly larger to absorb a factor of 2). ■

A.4. Tools for Overlap Form of Low Degree Advantages

Next, we provide some general tools for working with expressions that often appear in the style of analysis of LCDF and LDP algorithms that seeks to reduce the task to questions about a single "overlap" random variable involving the similarity of two draws of a hidden signal. This theme has been explored in previous work including Bandeira et al. (2020); Kunisky et al. (2022); Bandeira et al. (2021); Kunisky (2021a, 2024), following similar observations for computations of χ^2 divergences in Montanari et al. (2015); Perry et al. (2016); Banks et al. (2018). We aim here to distill an elementary but important manipulation that often appears in these proofs, to let it be used more flexibly in our context and related ones in the future.

Definition 35 (Truncated exponential) *For $D \geq 0$ an integer and $t \in \mathbb{R}$, we define*

$$\exp^{\leq D}(t) := \sum_{d=0}^{D} \frac{t^d}{d!}.$$

Proposition 36 (Basic properties) *For any $t \in \mathbb{R}$ and $D \geq 0$ an integer,*

$$\begin{aligned} \exp^{\leq D}(t) &\leq \exp^{\leq D}(|t|) \\ &\leq 2 \cdot \frac{(2D \vee |t|)^D}{D!}. \end{aligned} \tag{6}$$

Further, $\exp^{\leq D}(t)$ is an increasing function on $t \in \mathbb{R}_{\geq 0}$.

All of these observations but the bound (6) are immediate, and that bound is proved in Corollary 5.2.3 of Kunisky (2021b).

The techniques of the works cited above lead to expressions of the form $\mathbb{E}\exp^{\leq D(n)}(R_n)$ for a sequence of scalar random variables $R_n \geq 0$ and a sequence of growing degrees $D(n)$. Indeed, often R_n even converge in distribution to a limiting random variable, but the challenge is to understand the "race" between this convergence and the convergence of the function $\exp^{\leq D(n)}(t)$ to $\exp(t)$. The following gives a general treatment of precisely this situation, with no reference to low degree analysis. We will use the following non-asymptotic form of Stirling's approximation.

Proposition 37 *For any $d \geq 1$, $d! \geq (d/e)^d$.*

Lemma 38 *Let $R_n \geq 0$ be a sequence of real bounded random variables and $D(n) \in \mathbb{N}$. Suppose there exists a sequence $A(n) \in \mathbb{R}_{\geq 0}$ such that the following conditions hold:*

1. *For all sufficiently large n,*
$$A(n) \geq D(n)\left(2 \vee \log\left(\frac{\|R_n\|_\infty}{D(n)}\right)\right).$$
(Recall that $\|R_n\|_\infty$ is the smallest $C > 0$ such that $R_n \leq C$ almost surely.)
2. *For bounded $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that $\int_0^\infty f(t)\,dt < \infty$, for sufficiently large n, for all $t \in [0, A(n)]$,*
$$\mathbb{P}[R_n \geq t] \leq f(t)\exp(-t).$$

Then, we have
$$\limsup_{n\to\infty} \mathbb{E}\exp^{\leq D(n)}(R_n) < \infty.$$

Proof We decompose

$$\mathbb{E}\exp^{\leq D(n)}(R_n) = \underbrace{\mathbb{E}\mathbb{1}\{R_n > A(n)\}\exp^{\leq D(n)}(R_n)}_{=:E_{\mathsf{large}}} + \underbrace{\mathbb{E}\mathbb{1}\{R_n \leq A(n)\}\exp^{\leq D(n)}(R_n)}_{=:E_{\mathsf{small}}}.$$

For the "large deviations" term E_{large}, we use that $\exp^{\leq D(n)}(r)$ is monotone in r and bound using our assumption on $\|R_n\|$ and Proposition 36

$$\begin{aligned} E_{\mathsf{large}} &\leq \mathbb{P}[R_n > A(n)]\exp^{\leq D(n)}(\|R_n\|_\infty) \\ &\leq 2f(A(n))\exp(-A(n))\frac{(2D(n) \vee \|R_n\|_\infty)^{D(n)}}{D(n)!} \end{aligned}$$

and using Proposition 37 we have

$$\begin{aligned} &\leq 2f(A(n))\exp(-A(n))\left(2e \vee \frac{\|R_n\|_\infty}{D(n)}\right)^{D(n)} \\ &\leq 2f(A(n))\exp\left(-A(n) + D(n)\log\left(2e \vee \frac{\|R_n\|_\infty}{D(n)}\right)\right), \end{aligned}$$

and thus E_{large} is bounded since by our assumptions f is bounded and the expression in the exponential is negative.

For the "small deviations" term E_{small}, we use that $\exp^{\leq D}(r) \leq \exp(r)$, so

$$\begin{aligned}
E_{\mathsf{small}} &\leq \mathbb{E}\mathbb{1}\{R_n \leq A(n)\}\exp(R_n) \\
&= \int_0^\infty \mathbb{P}[\mathbb{1}\{R_n \leq A(n)\}\exp(R_n) \geq u]\, du \\
&= \int_0^\infty \mathbb{P}[\mathbb{1}\{R_n \leq A(n)\}\exp(R_n) \geq \exp(t)] \cdot \exp(t)\, dt \\
&= \int_0^{A(n)} \mathbb{P}[R_n \geq t] \cdot \exp(t)\, dt \\
&\leq \int_0^{A(n)} f(t)\, dt \\
&\leq \int_0^\infty f(t)\, dt,
\end{aligned}$$

which is again a finite constant, so E_{small} is bounded. Combining the two gives the result. ■

The prototypical applications in, e.g., Perry et al. (2016); Kunisky et al. (2022) to spiked matrix models, have used this as follows: for a constant $\lambda > 0$, R_n is the law of $\frac{\lambda^2}{n}\langle \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\rangle^2$ for, say, $\boldsymbol{x}^{(i)} \sim \mathrm{Unif}(\{\pm 1\}^n)$ independently (or, more generally, vectors with i.i.d. bounded entries having mean zero and unit variance) and $0 < \lambda < 1$. Then, $\|R_n\|_\infty = O(n)$ since $R_n \leq \frac{\lambda^2}{n}\|\boldsymbol{x}^{(1)}\|^2\|\boldsymbol{x}^{(2)}\|^2$, and the conditions of the Lemma hold with $f(t) = C\exp(-\delta t)$ for large $C > 0$ and small $\delta > 0$ (depending on λ, which is where it is important that $\lambda < 1$) and $A(n) = \varepsilon n$ for small $\varepsilon > 0$. The tail bound on R_n in Condition 2 of the Lemma is what those works call a "local Chernoff bound" for the inner product $\langle \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\rangle$. This result then gives low-degree lower bounds for, say, any $D(n) = o(n/\log n)$.

We are just observing here that all of these details of their analysis are incidental, and the above is the distilled analytic content of the argument. This is a simple observation, but, given the general bounds developed recently in Kunisky (2024) leading to more unusual forms of the random variables R_n, it seems that this will be a useful tool, and it will be helpful in our setting already. For instance, we will encounter a situation where R_n decomposes as a sum of (dependent) random variables, but we may establish the conditions of Lemma 38 simply by a carefully tuned union bound.

Appendix B. Main Lower Bounds

B.1. Coordinate Advantage Bound

We will use the following object to show that strong separation is impossible, following a style of argument that has appeared recently in Coja-Oghlan et al. (2022); Bandeira et al. (2022) using such quantities to reason about separation criteria.

Definition 39 (Coordinate advantage) *For $\mathbb{P}, \mathbb{Q}$ as in Definition 1 and $D \geq 1$, we define*

$$\mathsf{CAdv}_{\leq D}(\mathbb{P}, \mathbb{Q}) := \left\{ \begin{array}{ll} \textit{maximize} & \mathbb{E}_{\boldsymbol{y}\sim\mathbb{P}}\, f(\boldsymbol{y}) \\ \textit{subject to} & \mathbb{E}_{\boldsymbol{y}\sim\mathbb{Q}}\, f(\boldsymbol{y})^2 \leq 1, \\ & \mathsf{cdeg}(f) \leq D \end{array} \right\}. \tag{7}$$

The following is then immediate from the definition of strong separation.

Proposition 40 *For a sequence of pairs of probability measures $\mathbb{Q}_n$, $\mathbb{P}_n$ and $D(n) \in \mathbb{N}$, if $\mathsf{CAdv}_{\leq D(n)}(\mathbb{P}_n, \mathbb{Q}_n) = O(1)$ as $n \to \infty$, then there is no sequence of functions of coordinate degree at most $D(n)$ that strongly separate $\mathbb{Q}_n$ from $\mathbb{P}_n$.*

We follow the approach of Kunisky (2024) in deriving tractable bounds on the coordinate advantage for GSBMs. Consider first a single GSBM rather than a sequence, with parameters p, k, n as in Definition 1 and with a resulting pair of probability measures $\mathbb{Q}$ and $\mathbb{P}$. We begin by deriving the following remarkably simple bound on the coordinate advantage involving the characterstic tensor and a multinomial random vector.

Lemma 41 *In the above setting, let $\boldsymbol{T}$ be the characteristic tensor of the GSBM. Then,*

$$\mathsf{CAdv}_{\leq D}(\mathbb{Q}, \mathbb{P})^2 \leq \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathsf{Mult}(n, k^2)} \exp^{\leq D}\left(\langle \boldsymbol{T}, \boldsymbol{z}^{\otimes p} \rangle\right).$$

Note that this Lemma achieves a dramatic dimensionality reduction, similar to the "overlap formulas" discussed earlier: while our original problem had dimension growing with n, $\boldsymbol{T}$ is a tensor of size not depending on n, so we have isolated the role of n to its participation in the multinomial distribution of the vector $\boldsymbol{z}$ (which, like $\boldsymbol{T}$, has fixed dimension).

Proof Following Kunisky (2024), we define the *channel overlap*, for $\boldsymbol{a} = (a_1, \dots, a_p), \boldsymbol{b} = (b_1, \dots, b_p) \in [k]^p$, to be the quantity:

$$R(\boldsymbol{a}, \boldsymbol{b}) := \mathop{\mathbb{E}}_{y \sim \mu_{\mathsf{avg}}}\left[\left(\frac{d\mu_{\boldsymbol{a}}}{d\mu_{\mathsf{avg}}}(y) - 1\right)\left(\frac{d\mu_{\boldsymbol{b}}}{d\mu_{\mathsf{avg}}}(y) - 1\right)\right].$$

Recall that these are also the entries of $\boldsymbol{T}$. By our assumption of weak symmetry, R is unchanged by a simultaneous permutation of both of its inputs: for any $\sigma \in \mathrm{Sym}([p])$,

$$R((a_{\sigma(1)}, \dots, a_{\sigma(p)}), (b_{\sigma(1)}, \dots, b_{\sigma(p)})) = R((a_1, \dots, a_p), (b_1, \dots, b_p)).$$

From this we build a "total overlap," which is just the sum of the channel overlap over all observations. That is, for $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in [k]^n$, we set

$$\begin{aligned} R(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) &:= \sum_{1 < i_1 < \cdots < i_p \leq n} R((x^{(1)}_{i_1}, \dots, x^{(1)}_{i_p}), (x^{(2)}_{i_1}, \dots, x^{(2)}_{i_p})) \\ &= \sum_{1 < i_1 < \cdots < i_p \leq n} \mathop{\mathbb{E}}_{y \sim \mu_{\mathsf{avg}}}\left[\left(\frac{d\mu_{(x^{(1)}_{i_1}, \dots, x^{(1)}_{i_p})}}{d\mu_{\mathsf{avg}}}(y) - 1\right)\left(\frac{d\mu_{(x^{(2)}_{i_1}, \dots, x^{(2)}_{i_p})}}{d\mu_{\mathsf{avg}}}(y) - 1\right)\right] \\ &= \frac{1}{p!} \sum_{\substack{\boldsymbol{i} \in [n]^p \\ \text{all entries distinct}}} \mathop{\mathbb{E}}_{y \sim \mu_{\mathsf{avg}}}\left[\left(\frac{d\mu_{(x^{(1)}_{i_1}, \dots, x^{(1)}_{i_p})}}{d\mu_{\mathsf{avg}}}(y) - 1\right)\left(\frac{d\mu_{(x^{(2)}_{i_1}, \dots, x^{(2)}_{i_p})}}{d\mu_{\mathsf{avg}}}(y) - 1\right)\right], \end{aligned}$$

where the last step follows by our observation from weak symmetry above.

By (Kunisky, 2024, Theorem 3.5), the low coordinate degree advantage is bounded in terms of this total overlap as:

$$\mathsf{CAdv}_{\leq D}(\mathbb{Q}, \mathbb{P})^2 \leq \mathop{\mathbb{E}}_{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \sim \mathrm{Unif}([k]^p)} \exp^{\leq D}(R(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)})).$$

Consider next the modified overlap allowing for all tuples of indices:

$$R'(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) := \frac{1}{p!} \sum_{\boldsymbol{i} \in [n]^p} \mathop{\mathbb{E}}_{y \sim \mu_{\mathsf{avg}}} \left[\left(\frac{d\mu_{(x^{(1)}_{i_1}, \dots, x^{(1)}_{i_p})}}{d\mu_{\mathsf{avg}}}(y) - 1 \right) \left(\frac{d\mu_{(x^{(2)}_{i_1}, \dots, x^{(2)}_{i_p})}}{d\mu_{\mathsf{avg}}}(y) - 1 \right) \right].$$

We claim that our bound only increases upon replacing R with R':

$$\mathop{\mathbb{E}}_{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \sim \mathrm{Unif}([k]^p)} \exp^{\leq D}(R(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)})) \leq \mathop{\mathbb{E}}_{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \sim \mathrm{Unif}([k]^p)} \exp^{\leq D}(R'(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)})). \tag{8}$$

Indeed, we may write the two overlaps as

$$R(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = \frac{1}{p!} \sum_{\substack{\boldsymbol{i} \in [n]^p \\ \text{all entries distinct}}} \left\langle \bar{L}_{(x^{(1)}_{i_1}, \dots, x^{(1)}_{i_p})}, \bar{L}_{(x^{(2)}_{i_1}, \dots, x^{(2)}_{i_p})} \right\rangle,$$

$$R'(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = \frac{1}{p!} \sum_{\boldsymbol{i} \in [n]^p} \left\langle \bar{L}_{(x^{(1)}_{i_1}, \dots, x^{(1)}_{i_p})}, \bar{L}_{(x^{(2)}_{i_1}, \dots, x^{(2)}_{i_p})} \right\rangle,$$

where $\bar{L}_{\boldsymbol{a}} = \frac{d\mu_{\boldsymbol{a}}}{d\mu_{\mathsf{avg}}} - 1$ is the centered likelihood ratio and the inner products are in $L^2(\mu_{\mathsf{avg}})$. These likelihood ratios satisfy

$$\mathop{\mathbb{E}}_{\boldsymbol{x}^{(j)} \sim \mathrm{Unif}([k]^n)} \bar{L}_{(x^{(j)}_{i_1}, \dots, x^{(j)}_{i_p})} = 0$$

for all $\boldsymbol{i} \in [k]^n$ and $j \in \{1, 2\}$. Therefore, upon fully expanding the powers of R and R' appearing on either side of (8) and taking the expectation over $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}$ first, we will have that only the diagonal terms contribute to such an expansion, and each contribution is non-negative. There are more such terms in the expression with R', and so (8) holds (see (Kunisky, 2024, Lemma 4.5) for full details of this argument). Thus we also have

$$\mathsf{CAdv}_{\leq D}(\mathbb{Q}, \mathbb{P})^2 \leq \mathop{\mathbb{E}}_{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \sim \mathrm{Unif}([k]^p)} \exp^{\leq D}(R'(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)})).$$

Now, define

$$z_{a,b} = z_{a,b}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) := \#\{i \in [n] : x^{(1)}_i = a, x^{(2)}_i = b\}.$$

First note that we may rewrite R' by grouping like terms as

$$\begin{aligned} R'(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) &= \sum_{\boldsymbol{a} \in [k]^p} \sum_{\boldsymbol{b} \in [k]^p} \frac{1}{p!} \mathop{\mathbb{E}}_{y \sim \mu_{\mathsf{avg}}} \left[\left(\frac{d\mu_{\boldsymbol{a}}}{d\mu_{\mathsf{avg}}}(y) - 1 \right) \left(\frac{d\mu_{\boldsymbol{b}}}{d\mu_{\mathsf{avg}}}(y) - 1 \right) \right] z_{a_1, b_1} \cdots z_{a_p, b_p} \\ &= \langle \boldsymbol{T}, \boldsymbol{z}^{\otimes p} \rangle, \end{aligned}$$

where $\boldsymbol{T}$ is the characteristic tensor. Finally, the proof is complete after observing that, when $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \sim \mathrm{Unif}([k]^p)$ independently, then the law of $\boldsymbol{z}$ is precisely $\mathsf{Mult}(n, k^2)$, but where in the expression above we rearrange its entries into a $k \times k$ matrix. ∎

Now, note that $\mathbb{E}\boldsymbol{z} = \frac{n}{k^2}\mathbf{1}$, for $\mathbf{1}$ the all-ones vector (of dimension k^2, but implicitly indexed by $[k] \times [k]$). We define for $\boldsymbol{z}$ as in the Lemma its centered version,

$$\bar{\boldsymbol{z}} := \boldsymbol{z} - \frac{n}{k^2}\mathbf{1},$$

whereby we have

$$\boldsymbol{z} = (\boldsymbol{z} - \mathbb{E}\boldsymbol{z}) + \mathbb{E}\boldsymbol{z} = \bar{\boldsymbol{z}} + \frac{n}{k^2}\mathbf{1}.$$

The following straightforward calculation describes how the marginal characteristic tensors arise from plugging the above observation into the result of the Lemma.

Proposition 42 *In the above setting, if the marginal order of the GSBM is at least p_*, then*

$$\langle \boldsymbol{T}, \boldsymbol{z}^{\otimes p} \rangle = \sum_{j=p_*}^{p} \binom{p}{j} n^{p-j} \langle \boldsymbol{T}^{(j)}, \bar{\boldsymbol{z}}^{\otimes (j)} \rangle.$$

Proof Observe that $\boldsymbol{T}$ is a symmetric tensor. We then expand directly, grouping like terms:

$$\langle \boldsymbol{T}, \boldsymbol{z}^{\otimes p} \rangle = \left\langle \boldsymbol{T}, \left(\bar{\boldsymbol{z}} + \frac{n}{k^2}\mathbf{1} \right)^{\otimes p} \right\rangle$$

where since $\boldsymbol{T}$ is symmetric, in each of the 2^p terms arising we may group together all of the occurrences of $\mathbf{1}$ and all of the occurrences of $\bar{\boldsymbol{z}}$, giving an expansion as in the binomial theorem,

$$= \sum_{j=0}^{p} \binom{p}{j} n^{p-j} \frac{1}{k^{2(p-j)}} \langle \boldsymbol{T}, \underbrace{\mathbf{1} \otimes \cdots \otimes \mathbf{1}}_{p-j \text{ times}} \otimes \underbrace{\bar{\boldsymbol{z}} \otimes \cdots \otimes \bar{\boldsymbol{z}}}_{j \text{ times}} \rangle$$

but now by definition of the marginalized tensors,

$$= \sum_{j=0}^{p} \binom{p}{j} n^{p-j} \langle \boldsymbol{T}^{(j)}, \bar{\boldsymbol{z}}^{\otimes (p-j)} \rangle,$$

giving the result since all terms with $j < p^*$ vanish by the definition of marginal order. ∎

Finally, what we will actually use is the following bound on this overlap and therefore on the advantage. For the sake of clarity, we give both the form involving the centered multinomial vector $\bar{\boldsymbol{z}}$ following from the above, and the normalized version in terms of the Pearson χ^2 statistic that will connect to the bounds we have developed in Appendix A.3.

Corollary 43 *In the above setting, if the marginal order of the GSBM is at least p_*, then*

$$\begin{aligned} \mathsf{CAdv}_{\le D}(\mathbb{Q}, \mathbb{P}) &\le \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathsf{Mult}(n, k^2)} \exp^{\le D} \left(\sum_{j=p_*}^{p} \binom{p}{j} \|\boldsymbol{T}^{(j)}\|_{\mathsf{inj}}\, n^{p-j} \|\bar{\boldsymbol{z}}\|^j \right) \\ &= \mathop{\mathbb{E}}_{X \sim \chi^2_{\mathsf{Pear}}(n, k^2)} \exp^{\le D} \left(\sum_{j=p_*}^{p} \binom{p}{j} \frac{1}{k^j} \|\boldsymbol{T}^{(j)}\|_{\mathsf{inj}}\, n^{p-j/2} X^{j/2} \right). \end{aligned}$$

Proof The first part follows by combining Proposition 36, Proposition 42, and Lemma 41. The second part follows from observing that the law of $\frac{k^2}{n}\|\bar{\boldsymbol{z}}\|^2$ is precisely $\chi^2_{\mathsf{Pear}}(n, k^2)$. ∎

By way of intuition, in the final expression, note that we expect $X = \Theta(1)$ typically as $n \to \infty$. Thus, by far the largest term in the truncated exponential will correspond to the smallest value of j, which is $j = p_*$. So, this bound indeed expresses that the marginal order of the GSBM governs the low coordinate degree advantage.

Before proceeding, we also establish the following bound that will be useful throughout.

Proposition 44 *Let $X \sim \chi^2_{\mathsf{Pear}}(n, k^2)$. Then, $\|X\|_\infty \le k^2 n$.*

Proof Recall that we have $X = \frac{k^2}{n} \sum_{i=1}^{k^2} (z_i - \frac{n}{k^2})^2$ for $\boldsymbol{z} \sim \mathsf{Mult}(n, k^2)$. Since $\sum_{i=1}^{k^2} z_i = n$, we may interpret $\sum_{i=1}^{k^2} (z_i - \frac{n}{k^2})^2$ as a variance, and thus bound

$$\begin{aligned}
\sum_{i=1}^{k^2} \left(z_i - \frac{n}{k^2}\right)^2 &\le \sum_{i=1}^{k^2} z_i^2 \\
&\le \left(\sum_{i=1}^{k^2} z_i\right)^2 \\
&\le n^2,
\end{aligned}$$

and the result follows from substituting this into the expression for X. ■

B.2. Proof of Theorem 15: General Marginal Order

Proof We allow C to be a constant changing from line to line, which always needs only to be taken sufficiently large depending on the parameters p, k, ℓ (which do not change with n). Let $X_n \sim \chi^2_{\mathsf{Pear}}(n, k^2)$. Using Proposition 44, we may bound crudely, for all $p_* \le j \le p$,

$$\begin{aligned}
R_{n,j} &:= \binom{p}{j} \frac{1}{k^j} \|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}}\, n^{p - \frac{j}{2}} X_n^{\frac{j}{2}} \\
&\le C \|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}}\, n^{p - \frac{p_*}{2}} X_n^{\frac{p_*}{2}}, \\
R_n &:= \sum_{j=2}^{p} R_{n,j} \\
&\le C \left(\max_{p_* \le j \le p} \|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}} \right) n^{p - \frac{p_*}{2}} X_n^{\frac{p_*}{2}}.
\end{aligned}$$

Using Corollary 43, we have

$$\begin{aligned}
\mathsf{CAdv}_{\le D(n)}(\mathbb{Q}_n, \mathbb{P}_n)^2 &\le \mathbb{E} \exp^{\le D(n)}(R_n) \\
&= \sum_{d=0}^{D(n)} \frac{1}{d!} \mathbb{E} R_n^d \\
&\le \sum_{d=0}^{D(n)} \frac{1}{d!} \left(C \left(\max_{p_* \le j \le p} \|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}} \right) n^{p - \frac{p_*}{2}} \right)^d \mathbb{E} X_n^{\frac{p_* d}{2}}
\end{aligned}$$

and now, provided that $D(n) \le n/C$, we have by Corollary 34

$$\begin{aligned}
&\le C \sum_{d=0}^{D(n)} \frac{1}{d!} \left(C \left(\max_{p_* \le j \le p} \|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}} \right) n^{p - \frac{p_*}{2}} \right)^d d^{\frac{3}{2}} (p_* d)^{\frac{p_* d}{2}} \\
&\le C \sum_{d=0}^{D(n)} \frac{1}{d!} \left(C \left(\max_{p_* \le j \le p} \|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}} \right) n^{p - \frac{p_*}{2}} d^{\frac{p_*}{2}} \right)^d
\end{aligned}$$

and by Proposition 37,

$$\leq C \sum_{d=0}^{D(n)} \left(C \left(\max_{p_* \leq j \leq p} \|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}} \right) n^{p - \frac{p_*}{2}} d^{\frac{p_*}{2} - 1} \right)^d$$
$$\leq C \sum_{d=0}^{\infty} \left(C \left(\max_{p_* \leq j \leq p} \|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}} \right) n^{p - \frac{p_*}{2}} D(n)^{\frac{p_*}{2} - 1} \right)^d,$$

and the result follows since our assumption implies that the inner expression is, say, at most $1/2$ for sufficiently large n, whereby this series converges. ■

B.3. Proof of Theorem 16: Marginal Order 2

Proof We will at first follow the same ideas as the previous proof, but eventually will need to be much more precise. Let $X_n \sim \chi^2_{\mathsf{Pear}}(n, k^2)$, and define

$$R_{n,j} := \binom{p}{j} \frac{1}{k^j} \|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}}\, n^{p - \frac{j}{2}} X_n^{\frac{j}{2}} \text{ for } 2 = p_* \leq j \leq p,$$
$$R_n := \sum_{j=2}^{p} R_{n,j}.$$

Then, Corollary 43 states that

$$\mathsf{CAdv}_{\leq D(n)}(\mathbb{Q}_n, \mathbb{P}_n) \leq \mathbb{E} \exp^{\leq D(n)}(R_n).$$

We will use Lemma 38, which concerns precisely such expressions, to show that this is bounded as $n \to \infty$.

We first establish some preliminary bounds using the assumptions of the Theorem. Recall that we assume, for constants $C > 0$ and $\varepsilon \in (0, 1)$ and sufficiently large n, that

$$\|\boldsymbol{T}_n^{(2)}\| \leq (1 - \varepsilon) \frac{k^2}{p(p-1)} \frac{1}{n^{p-1}},$$
$$\|\boldsymbol{T}_n^{(j)}\|_{\mathsf{inj}} \leq C \frac{1}{n^{p-1}} \text{ for all } 3 \leq j \leq p.$$

Let us slightly abuse notation and let the constant C from the statement of the theorem increase from line to line in the proof, but only depending on the constant parameters p, k, ℓ, ε from the statement of the Theorem (not on the growing parameter n). Plugging this in, we then find

$$R_{n,2} \leq (1 - \varepsilon) \frac{1}{2} X_n, \tag{9}$$
$$R_{n,j} \leq C n^{-\frac{j}{2} + 1} X_n^{\frac{j}{2}} \text{ for all } 3 \leq j \leq p. \tag{10}$$

Now, we prove a tail bound for R_n in order to apply Lemma 38. By the union bound and the above observations,

$$\begin{aligned}\mathbb{P}[R_n \geq t] &\leq \mathbb{P}\left[R_{n,2} \geq \left(1 - \frac{\varepsilon}{2}\right) t\right] + \sum_{j=3}^{p} \mathbb{P}\left[R_{n,j} \geq \frac{\varepsilon}{2p} t\right] \\ &\leq \mathbb{P}\left[X_n \geq \frac{1 - \frac{\varepsilon}{2}}{1 - \varepsilon} 2t\right] + \sum_{j=3}^{p} \mathbb{P}\left[X_n \geq \frac{1}{C} n^{1 - \frac{2}{j}} t^{\frac{2}{j}}\right]\end{aligned}$$

and provided that we take C even larger and restrict to $t \leq n/C$, we may ensure that all of the terms in the latter sum are at most the first probability, whereby

$$\leq p \cdot \mathbb{P}\left[X_n \geq \frac{1 - \frac{\varepsilon}{2}}{1 - \varepsilon} 2t\right] \quad \text{if } t \leq \frac{n}{C},$$

where, defining $\varepsilon' > 0$ appropriately depending on ε, this is equivalently

$$\leq p \cdot \mathbb{P}\left[X_n \geq (1 + \varepsilon') 2t\right] \quad \text{if } t \leq \frac{n}{C},$$

and finally, using the tail bound of Lemma 32, for $\delta \in (0, 1)$ to be chosen momentarily, we have

$$\leq p \left(1 + \frac{2}{\delta}\right)^d \exp\left(-\frac{(1 - \delta)^2 (1 + \varepsilon')}{1 + \frac{1 - \delta}{3} \sqrt{\frac{k^2 - 1}{n}} \sqrt{t}} t\right) \quad \text{if } t \leq \frac{n}{C}$$

where choosing δ small enough and C large enough again, we have

$$\leq C \exp\left(-\left(1 + \frac{\varepsilon'}{2}\right) t\right) \quad \text{if } t \leq \frac{n}{C}.$$

Thus, Lemma 38 applies with the choices

$$\begin{aligned}A(n) &:= \frac{n}{C}, \\ f(t) &:= C \exp\left(-\frac{\varepsilon'}{2} t\right).\end{aligned}$$

It remains to check that the condition of Lemma 38 relating $A(n)$ and $D(n)$ holds. Recall that this requires that

$$A(n) \geq D(n) \left(2 \vee \log\left(\frac{\|R_n\|_\infty}{D(n)}\right)\right),$$

while the assumption of the Theorem guarantees us that

$$D(n) \leq C \frac{n}{\log n}.$$

We must control $\|R_n\|_\infty$; to that end, recall that by Proposition 44, for $X_n \sim \chi^2_{\mathsf{Pear}}(n, k^2)$, we have $\|X_n\|_\infty \le k^2 n$. Accordingly, taking C suitably large, by the estimates in (9) and (10), we have

$$\|R_n\|_\infty \le Cn.$$

Thus, for sufficiently large n,

$$D(n)\left(2 \vee \log\left(\frac{\|R_n\|_\infty}{D(n)}\right)\right) \le Cn\frac{\log\log n}{\log n} \le A(n).$$

Thus the final condition of Lemma 38 is verified, the Lemma applies, and its result gives

$$\mathsf{CAdv}_{\le D(n)}(\mathbb{Q}_n, \mathbb{P}_n) \le \mathbb{E}\exp^{\le D(n)}(R_n) = O(1)$$

as $n \to \infty$, completing the proof. ■

Appendix C. Applications

C.1. Characteristic Tensors for Discrete Observations

We first make a few observations about our results over GSBMs where Ω is a finite set, which is the setting all of our applications will occur in.

Definition 45 *We call a GSBM* discrete *if Ω is a finite set, in which case we identify $\Omega = [\ell]$. We call a discrete GSBM* non-degenerate *if, for all $y \in [\ell]$, there exists some $\boldsymbol{a} \in [k]^p$ such that $\mu_{\boldsymbol{a}}(y) > 0$. An equivalent condition is that $\mu_{\mathsf{avg}}(y) > 0$ for all $y \in [\ell]$.*

In words, non-degeneracy just asks that there is no $y \in [\ell]$ that occurs in observations $\boldsymbol{Y} \sim \mathbb{Q}$ or $\boldsymbol{Y} \sim \mathbb{P}$ with probability zero and thus can effectively be removed from the model without changing it. This is implicitly achieved by the definition of the Radon-Nikodym derivative in our previous formulation, but for more explicit formulas below it will be convenient to make this assumption.

Proposition 46 *In a non-degenerate discrete GSBM, the characteristic tensor has entries*

$$T_{(a_1,b_1),\dots,(a_p,b_p)} = \sum_{y\in[\ell]} \frac{1}{\mu_{\mathsf{avg}}(y)}(\mu_{(a_1,\dots,a_p)}(y) - \mu_{\mathsf{avg}}(y))(\mu_{(b_1,\dots,b_p)}(y) - \mu_{\mathsf{avg}}(y)).$$

Said differently, if we view $\boldsymbol{M}(y) \in (\mathbb{R}^k)^{\otimes p}$ for each $y \in \ell$ as the tensor with $M(y)_{a_1,\dots,a_p} = \mu_{(a_1,\dots,a_p)}(y) - \mu_{\mathsf{avg}}(y)$, then

$$\boldsymbol{T} = \frac{1}{p!}\sum_{y\in[\ell]} \frac{1}{\mu_{\mathsf{avg}}(y)}\boldsymbol{M}(y)^{\otimes 2}, \tag{11}$$

with the caveat that we "flatten" pairs of dimensions in the tensor power $\boldsymbol{M}(y)^{\otimes 2}$ (as in the Kronecker product of matrices).

Because of the above formulas, it will also be convenient to introduce a notation for the centered channel measures:

$$\bar{\mu}_{\boldsymbol{a}}(y) := \mu_{\boldsymbol{a}}(y) - \mu_{\mathsf{avg}}(y).$$

C.2. Graph Stochastic Block Models

C.2.1. WARMUP: SYMMETRIC TWO COMMUNITY STOCHASTIC BLOCK MODEL

As a warmup, consider the stochastic block model with two communities and interaction matrix

$$\boldsymbol{Q} = \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix} \tag{12}$$

for some constants $\alpha, \beta \geq 0$.

Writing this in our GSBM form, we have $p = k = \ell = 2$ (pairwise observations, of membership in one of two communities, taking binary values), and, writing the probability measures $\mu_{\boldsymbol{a}}$ as vectors, we have:

$$\begin{aligned} \mu_{(1,1)} = \mu_{(2,2)} &= \left(\frac{\alpha}{n}, 1 - \frac{\alpha}{n}\right), \\ \mu_{(1,2)} = \mu_{(2,1)} &= \left(\frac{\beta}{n}, 1 - \frac{\beta}{n}\right), \\ \mu_{\mathsf{avg}} &= \left(\frac{\alpha+\beta}{2n}, 1 - \frac{\alpha+\beta}{2n}\right), \\ \bar{\mu}_{(1,1)} = \bar{\mu}_{(2,2)} &= \left(\frac{\alpha-\beta}{2n}, -\frac{\alpha-\beta}{2n}\right), \\ \bar{\mu}_{(1,2)} = \bar{\mu}_{(2,1)} &= \left(-\frac{\alpha-\beta}{2n}, \frac{\alpha-\beta}{2n}\right). \end{aligned}$$

Thus the characteristic matrix is:

$$\begin{aligned} \boldsymbol{T} &= \frac{1}{2}\left(\frac{1}{\frac{\alpha+\beta}{2n}} \begin{bmatrix} \frac{\alpha-\beta}{2n} & -\frac{\alpha-\beta}{2n} \\ -\frac{\alpha-\beta}{2n} & \frac{\alpha-\beta}{2n} \end{bmatrix}^{\otimes 2} + \frac{1}{1 - \frac{\alpha+\beta}{2n}} \begin{bmatrix} -\frac{\alpha-\beta}{2n} & \frac{\alpha-\beta}{2n} \\ \frac{\alpha-\beta}{2n} & -\frac{\alpha-\beta}{2n} \end{bmatrix}^{\otimes 2}\right) \\ &= \frac{1}{n} \cdot \left(\frac{(\alpha-\beta)^2}{4(\alpha+\beta)} + \frac{1}{n - \frac{\alpha+\beta}{2}}\right) \cdot \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}^{\otimes 2}, \end{aligned}$$

and its operator norm is

$$\|\boldsymbol{T}\| = \frac{1}{n} \cdot \frac{(\alpha-\beta)^2}{\alpha+\beta} + O\left(\frac{1}{n^2}\right).$$

Theorem 16 gives that functions of coordinate degree $O(n / \log n)$ cannot achieve strong separation in this stochastic block model once the above quantity is smaller by a constant factor than $\frac{k^2}{2n} = \frac{2}{n}$, which coincides with the Kesten-Stigum threshold. That is, we find that, once

$$\frac{(\alpha-\beta)^2}{2(\alpha+\beta)} < 1,$$

the above class of functions cannot achieve strong separation.

C.2.2. PROOF OF THEOREM 18: GENERAL STOCHASTIC BLOCK MODEL

We now advance to the general case advertised in Theorem 18. The calculations in the proof will be completely analogous to but slightly more abstract than those above.

Proof Suppose that, as in the setting of the Theorem, we have a general stochastic block model, for which the interaction matrix is $\boldsymbol{Q} \in [0,1]^{k \times k}_{\text{sym}}$. Now we still have $p = \ell = 2$ (pairwise binary observations) but potentially with larger k (number of communities).

Again translating to our language of GSBMs, this corresponds to measures $\mu_{a,b}$ for $a, b \in [k]$ with probability masses

$$\begin{aligned}
\mu_{a,b} &= \left(\frac{1}{n}Q_{a,b}, 1 - \frac{1}{n}Q_{a,b}\right), \\
\mu_{\mathsf{avg}} &= \left(\frac{1}{n}\frac{1}{k^2}\mathbf{1}^\top \boldsymbol{Q}\mathbf{1}, 1 - \frac{1}{n}\frac{1}{k^2}\mathbf{1}^\top \boldsymbol{Q}\mathbf{1}\right), \\
\bar{\mu}_{a,b} &= \frac{1}{n}\left(Q_{a,b} - \frac{1}{k^2}\mathbf{1}^\top \boldsymbol{Q}\mathbf{1}\right) \cdot [\,1,\, -1\,].
\end{aligned}$$

Recall that we have

$$\boldsymbol{Q}\mathbf{1} = \lambda \mathbf{1}$$

for some $\lambda \geq 0$ (which amounts to asking that a vertex in any community has the same average degree), which implies that our model has marginal order at least 2. In this case, taking the inner product with $\mathbf{1}$ on either side, we see that we must have

$$\lambda = \frac{1}{k}\mathbf{1}^\top \boldsymbol{Q}\mathbf{1},$$

and in terms of this we can rewrite the above as

$$\begin{aligned}
\mu_{\mathsf{avg}} &= \left(\frac{1}{n}\frac{\lambda}{k}, 1 - \frac{1}{n}\frac{\lambda}{k}\right), \\
\bar{\mu}_{a,b} &= \frac{1}{n}\left(Q_{a,b} - \frac{\lambda}{k}\right) \cdot [\,1,\, -1\,].
\end{aligned}$$

By the same calculations as before, the characteristic matrix is then

$$\boldsymbol{T} = \frac{1}{2}\left(\frac{1}{\frac{1}{n}\frac{\lambda}{k}} + \frac{1}{1 - \frac{1}{n}\frac{\lambda}{k}}\right)\left(\frac{1}{n}\left(\boldsymbol{Q} - \frac{\lambda}{k}\mathbf{1}\mathbf{1}^\top\right)\right)^{\otimes 2},$$

where we note that $\boldsymbol{Q} - \frac{\lambda}{k}\mathbf{1}\mathbf{1}^\top$ is just $\boldsymbol{Q}$ with its unique largest eigenvalue (by the Perron-Frobenius theorem) removed. Thus the operator norm is

$$\|\boldsymbol{T}\| = \frac{1}{n}\frac{k}{2\lambda_1(\boldsymbol{Q})}\max_{j \in \{2,\dots,k\}} |\lambda_j(\boldsymbol{Q})|^2 + O\left(\frac{1}{n^2}\right).$$

As before, Theorem 16 gives that functions of coordinate degree $O(n \,/\, \log n)$ cannot achieve strong separation in this stochastic block model once the above quantity is smaller by a constant factor than $\frac{k^2}{2n}$, which coincides with the (now generalized) Kesten-Stigum threshold, i.e., once

$$\max_{j \in \{2,\dots,k\}} |\lambda_j(\boldsymbol{Q})|^2 < k\lambda_1(\boldsymbol{Q}),$$

then the above class of functions cannot achieve strong separation, as claimed. ■

C.3. Hypergraph Stochastic Block Model

C.3.1. WARMUP: SYMMETRIC TWO COMMUNITY HYPERGRAPH STOCHASTIC BLOCK MODEL

As before, let us begin with the simpler case of two symmetric communities. Let us be explicit in this manageable setting and write the values of the various tensors involved directly. To that end, let $\boldsymbol{D} \in (\{0,1\}^2)^{\otimes p}$ be the *diagonal* tensor whose entries are $D_{a,\dots,a} = 1$ for $a \in \{1,2\}$ and all other entries zero, and let $\boldsymbol{F} \in (\{0,1\}^2)^{\otimes p}$ be the *off-diagonal* tensor, $\boldsymbol{F} = [1,1]^{\otimes p} - \boldsymbol{D}$, having the opposite pattern of entries. Very concretely, we may express

$$\begin{aligned}
\boldsymbol{D} &= \boldsymbol{e}_1^{\otimes p} + \boldsymbol{e}_2^{\otimes p}, \\
\boldsymbol{F} &= \mathbf{1}^{\otimes p} - \boldsymbol{D} \\
&= \mathbf{1}^{\otimes p} - \boldsymbol{e}_1^{\otimes p} - \boldsymbol{e}_2^{\otimes p}.
\end{aligned}$$

Then, the interaction tensor of the symmetric two community HSBM is

$$\boldsymbol{Q} = \alpha \boldsymbol{D} + \beta \boldsymbol{F} \in ([0,1]^2)^{\otimes p}.$$

For the GSBM parameters, we still have $k = \ell = 2$ (two communities and binary values of the observations), but now have general p (p-ary observations) unlike the graph SBM.

Writing the rest of the setup in the GSBM language, we have:

$$\begin{aligned}
\mu_{(1,\dots,1)} = \mu_{(2,\dots,2)} &= \left(\frac{a}{\binom{n}{p-1}}, 1 - \frac{a}{\binom{n}{p-1}} \right), \\
\mu_{\boldsymbol{a}} &= \left(\frac{b}{\binom{n}{p-1}}, 1 - \frac{b}{\binom{n}{p-1}} \right) \text{ for all } \boldsymbol{a} \notin \{(1,\dots,1),(2,\dots,2)\}, \\
\mu_{\mathsf{avg}} &= \left(\frac{a + (2^{p-1}-1)b}{2^{p-1}\binom{n}{p-1}}, 1 - \frac{a + (2^{p-1}-1)b}{2^{p-1}\binom{n}{p-1}} \right), \\
\bar{\mu}_{(1,\dots,1)} = \bar{\mu}_{(2,\dots,2)} &= \left(\frac{2^{p-1}-1}{2^{p-1}\binom{n}{p-1}}(a-b), -\frac{2^{p-1}-1}{2^{p-1}\binom{n}{p-1}}(a-b) \right), \\
\bar{\mu}_{\boldsymbol{a}} &= \left(-\frac{1}{2^{p-1}\binom{n}{p-1}}(a-b), \frac{1}{2^{p-1}\binom{n}{p-1}}(a-b) \right) \text{ for all } \boldsymbol{a} \notin \{(1,\dots,1),(2,\dots,2)\}.
\end{aligned}$$

The characteristic tensor is therefore

$$\boldsymbol{T} = \boldsymbol{T}^{(p)} = \frac{1}{p!} \left(\frac{2^{p-1}\binom{n}{p-1}}{a + (2^{p-1}-1)b} + \frac{1}{1 - \frac{a+(2^{p-1}-1)b}{2^{p-1}\binom{n}{p-1}}} \right) \left(\frac{a-b}{2^{p-1}\binom{n}{p-1}} \right)^2 \boldsymbol{M}^{\otimes 2},$$

where, using our previous expressions for $\boldsymbol{D}$ and $\boldsymbol{F}$,

$$\begin{aligned}
\boldsymbol{M} &= (2^{p-1}-1)\boldsymbol{D} - \boldsymbol{F} \\
&= (2^{p-1}-1)(\boldsymbol{e}_1^{\otimes p} + \boldsymbol{e}_2^{\otimes p}) - (\mathbf{1}^{\otimes p} - \boldsymbol{e}_1^{\otimes p} - \boldsymbol{e}_2^{\otimes p}) \\
&= 2^{p-1}(\boldsymbol{e}_1^{\otimes p} + \boldsymbol{e}_2^{\otimes p}) - \mathbf{1}^{\otimes p}.
\end{aligned}$$

Note that, when $p = 2$, this is compatible with our result from the corresponding first calculation of Appendix C.2.1.

We may now observe that the marginal order of this model is indeed 2: we have $\boldsymbol{M}[\mathbf{1}, \dots, \mathbf{1}, \cdot] = \mathbf{0}$ and therefore $\boldsymbol{T}^{(1)} = \mathbf{0}$, while the next-smallest marginalization is

$$\boldsymbol{T}^{(2)} = \frac{1}{p! \cdot 2^{2(p-2)}} \cdot \left(\frac{2^{p-1}\binom{n}{p-1}}{a + (2^{p-1}-1)b} + \frac{1}{1 - \frac{a+(2^{p-1}-1)b}{2^{p-1}\binom{n}{p-1}}} \right) \left(\frac{a-b}{2^{p-1}\binom{n}{p-1}} \right)^2 \\ \cdot \; 2^{2p-2} \left((\boldsymbol{e}_1^{\otimes 2} + \boldsymbol{e}_2^{\otimes 2}) - \frac{1}{2}\mathbf{1}^{\otimes 2} \right)^{\otimes 2}$$

which, canceling a factor and rewriting as a matrix, is

$$= \frac{4}{p!} \left(\frac{2^{p-1}\binom{n}{p-1}}{a + (2^{p-1}-1)b} + \frac{1}{1 - \frac{a+(2^{p-1}-1)b}{2^{p-1}\binom{n}{p-1}}} \right) \left(\frac{a-b}{2^{p-1}\binom{n}{p-1}} \right)^2 \cdot \left(\boldsymbol{I}_2 - \frac{1}{2}\mathbf{1}_2\mathbf{1}_2^\top \right)^{\otimes 2}.$$

Thus the bottom line is much the same as before: this remaining matrix is just the orthogonal projection matrix to the direction orthogonal to $\mathbf{1}_2$, and thus the operator norm of this marginalized characteristic matrix is

$$\begin{aligned} \|\boldsymbol{T}^{(2)}\| &= \frac{4}{p!} \left(\frac{2^{p-1}\binom{n}{p-1}}{a + (2^{p-1}-1)b} + \frac{1}{1 - \frac{a+(2^{p-1}-1)b}{2^{p-1}\binom{n}{p-1}}} \right) \left(\frac{a-b}{2^{p-1}\binom{n}{p-1}} \right)^2 \\ &= \frac{4}{p!} \cdot \frac{1}{\binom{n}{p-1}} \cdot \frac{(a-b)^2}{2^{p-1}(a + (2^{p-1}-1)b)} + O\left(\frac{1}{n^p}\right) \\ &= \frac{1}{n^{p-1}} \cdot \frac{4(a-b)^2}{p2^{p-1}(a + (2^{p-1}-1)b)} + O\left(\frac{1}{n^p}\right). \end{aligned}$$

Theorem 16 gives that functions of coordinate degree $O(n \,/\, \log n)$ cannot achieve strong separation in this HSBM once the above quantity is smaller by a constant factor than $\frac{k^2}{p(p-1)n^{p-1}} = \frac{4}{p(p-1)n^{p-1}}$. In particular, it suffices to have

$$\frac{(p-1)(a-b)^2}{2^{p-1}(a + (2^{p-1}-1)b)} < 1,$$

which is precisely the HSBM version of the Kesten-Stigum threshold.

C.3.2. PROOF OF THEOREM 20: GENERAL HYPERGRAPH STOCHASTIC BLOCK MODEL

Proof Suppose that, as in the setting of the Theorem, we have a general stochastic block model with a fixed symmetric interaction tensor $\boldsymbol{Q} \in ([0, 1]^k)^{\otimes p}$. We are now in a setting where $\ell = 2$ (binary observations) but p (arity of interactions) and k (number of communities) are arbitrary.

Again translating to our language of GSBMs, this corresponds to measures $\mu_{\boldsymbol{a}}$ for $\boldsymbol{a} \in [k]^p$ with probability masses

$$\begin{aligned}
\mu_{\boldsymbol{a}} &= \left[\frac{1}{\binom{n}{p-1}} Q_{\boldsymbol{a}},\ 1 - \frac{1}{\binom{n}{p-1}} Q_{\boldsymbol{a}} \right], \\
\mu_{\mathsf{avg}} &= \left[\frac{1}{\binom{n}{p-1}} \frac{1}{k^p} \langle \boldsymbol{Q}, \mathbf{1}^{\otimes p} \rangle,\ 1 - \frac{1}{\binom{n}{p-1}} \frac{1}{k^p} \langle \boldsymbol{Q}, \mathbf{1}^{\otimes p} \rangle \right], \\
\bar{\mu}_{\boldsymbol{a}} &= \frac{1}{\binom{n}{p-1}} \left(Q_{\boldsymbol{a}} - \frac{1}{k^p} \langle \boldsymbol{Q}, \mathbf{1}^{\otimes p} \rangle \right) \cdot \left[1,\ -1 \right].
\end{aligned}$$

Recall that we have assumed that

$$\boldsymbol{Q}[\mathbf{1}, \dots, \mathbf{1}, \cdot\,] = \lambda \mathbf{1}. \tag{13}$$

We again have

$$\lambda = \frac{1}{k} \langle \boldsymbol{Q}, \mathbf{1}^{\otimes p} \rangle,$$

whereby we can rewrite

$$\begin{aligned}
\mu_{\mathsf{avg}} &= \left(\frac{1}{\binom{n}{p-1}} \frac{\lambda}{k^{p-1}}, 1 - \frac{1}{\binom{n}{p-1}} \frac{\lambda}{k^{p-1}} \right), \\
\bar{\mu}_{\boldsymbol{a}} &= \frac{1}{\binom{n}{p-1}} \left(Q_{\boldsymbol{a}} - \frac{\lambda}{k^{p-1}} \right) \cdot \left[1,\ -1 \right]
\end{aligned}$$

The characteristic tensor is then

$$\boldsymbol{T}^{(p)} = \frac{1}{p!} \left(\frac{1}{\frac{1}{\binom{n}{p-1}} \frac{\lambda}{k^{p-1}}} + \frac{1}{1 - \frac{1}{\binom{n}{p-1}} \frac{\lambda}{k^{p-1}}} \right) \frac{1}{\binom{n}{p-1}^2} \left(\boldsymbol{Q} - \frac{\lambda}{k^{p-1}} \mathbf{1}^{\otimes p} \right)^{\otimes 2}.$$

We see that this model will indeed have marginal order at least 2, as the marginalized characteristic tensor $\boldsymbol{T}^{(1)}$ (a 1-tensor, or vector) will be zero by our assumption (13), while the next marginalization will be

$$\boldsymbol{T}^{(2)} = \frac{1}{p! \cdot k^{2(p-2)}} \left(\frac{1}{\frac{1}{\binom{n}{p-1}} \frac{\lambda}{k^{p-1}}} + \frac{1}{1 - \frac{1}{\binom{n}{p-1}} \frac{\lambda}{k^{p-1}}} \right) \frac{1}{\binom{n}{p-1}^2} \left(\boldsymbol{Q}[\mathbf{1}, \dots, \mathbf{1}, \cdot, \cdot] - \frac{\lambda}{k} \mathbf{1}^{\otimes 2} \right)^{\otimes 2}.$$

Note now that $\boldsymbol{Q}[\mathbf{1}, \dots, \mathbf{1}, \cdot, \cdot]$ is a non-negative symmetric matrix, and by our assumption (13), its Perron-Frobenius eigenvalue is λ with eigenvector $\mathbf{1}$. Thus the matrix appearing above (as a 2-tensor) is just the matrix $\boldsymbol{Q}[\mathbf{1}, \dots, \mathbf{1}, \cdot, \cdot]$ with its top eigenvector removed. The operator norm of $\boldsymbol{T}^{(2)}$ is therefore

$$\begin{aligned}
\|\boldsymbol{T}^{(2)}\| &= \frac{1}{p! \cdot k^{p-3} \cdot \binom{n}{p-1} \cdot \lambda} \max_{j=2,\dots,k} |\lambda_j(\boldsymbol{Q}[\mathbf{1}, \dots, \mathbf{1}, \cdot, \cdot])|^2 + O\left(\frac{1}{n^p}\right) \\
&= \frac{1}{n^{p-1}} \cdot \frac{1}{p \cdot k^{p-3} \cdot \lambda} \max_{j=2,\dots,k} |\lambda_j(\boldsymbol{Q}[\mathbf{1}, \dots, \mathbf{1}, \cdot, \cdot])|^2 + O\left(\frac{1}{n^p}\right)
\end{aligned}$$

As before, Theorem 16 gives that functions of coordinate degree $O(n \,/\, \log n)$ cannot achieve strong separation in this stochastic block model once the above quantity is smaller by a constant factor than $\frac{k^2}{p(p-1)n}$, which coincides with the (now generalized) Kesten-Stigum threshold, i.e., once

$$\max_{j \in \{2,\dots,k\}} |\lambda_j(\boldsymbol{Q}[\mathbf{1},\dots,\mathbf{1},\cdot,\cdot])|^2 < \frac{k^{p-1}}{p-1} \lambda_1(\boldsymbol{Q}[\mathbf{1},\dots,\mathbf{1},\cdot,\cdot]),$$

then the above class of functions cannot achieve strong separation. ■

C.3.3. Higher Marginal Order Example: Random XOR-SAT

As an example of an interesting model with marginal order higher than 2, let us describe how a lower bound matching several prior works on the random p-XOR-SAT problem can be seen in our framework. Our first task is to embed a version of random p-XOR-SAT into a GSBM. Recall that a p-XOR-SAT instance is equivalently a system of parity equations of the form $x_{i_1} \cdots x_{i_p} = b$ with $b \in \{\pm 1\}$, to be solved over $\boldsymbol{x} \in \{\pm 1\}^n$. Each such equation is sometimes called a *clause*, and we will follow standard notation in writing m for the number of these clauses. We consider hypothesis testing problems that ask to distinguish between uniformly random such instances (in a sense we will clarify in a moment) and ones with a planted structure causing unusually many clauses to be satisfiable at once.

Consider a GSBM of order p with $\Omega = \{+1, -1, \bullet\}$, where $\bullet$ stands for a clause that is not included in an instance. We set $k = 2$, but identify the labels not with $[2]$ but again with $\{\pm 1\}$. We then define, for $\eta \in (0, 1)$,

$$\mu_{(a_1,\dots,a_p)} := \eta \delta_{a_1 \cdots a_p} + (1-\eta)\delta_\bullet,$$

which will yield the average

$$\mu_{\mathsf{avg}} = \eta \, \mathrm{Unif}(\{\pm 1\}) + (1-\eta)\delta_\bullet.$$

A sample from the null model $\mathbb{Q}$ of this GSBM will then be, in effect, a p-XOR-SAT instance where every clause is present with probability η independently and the right-hand sides are drawn uniformly at random from $\{\pm 1\}$. A sample from the planted model $\mathbb{P}$ will be an instance where every clause is again present with probability η independently, but there is a uniformly random satisfying assignment $\boldsymbol{x}$ chosen that determines the right-hand sides so that it is satisfying.

Our discussion of lower bounds will actually apply to this exactly satisfiable model; as has been widely discussed, the noiseless p-XOR-SAT problem is solvable by Gaussian elimination, but many frameworks for understanding algorithmic hardness more generally only address robust algorithms and neglect this "brittle" algebraic algorithm. See, e.g., discussion in Kunisky et al. (2022)). One may also add a small amount of noise by resampling (in the sense of Definition 25) with a small probability. In fact, our construction above is just the $(1-\eta)$-censorship (in the sense of Definition 27) of the GSBM that reveals all clauses. By our discussion in Section 4.5, rates of censorship and resampling have the same effect on our low-degree calculations, so a small amount of further resampling noise will effectively just change η slightly, and we may just as well repeat all calculations after that operation. We continue calculating with the exactly satisfiable model above for the purposes of this discussion.

We compute the centered likelihood ratios

$$\frac{d\mu_{(a_1,\dots,a_p)}}{d\mu_{\mathsf{avg}}}(y) - 1 = \left\{ \begin{array}{ll} 0 & \text{if } y = \bullet, \\ (-1)^{\mathbb{1}\{a_1 \cdots a_p = y\}} & \text{if } y \in \{\pm 1\} \end{array} \right\},$$

from which we compute the entries of the characteristic tensor, which end up with a simple form:

$$\begin{aligned} T_{(a_1,b_1),\dots,(a_p,b_p)} &= \mathop{\mathbb{E}}_{y \sim \mu_{\mathsf{avg}}} \left[\left(\frac{d\mu_{(a_1,\dots,a_p)}}{d\mu_{\mathsf{avg}}}(y) - 1 \right) \left(\frac{d\mu_{(b_1,\dots,b_p)}}{d\mu_{\mathsf{avg}}}(y) - 1 \right) \right] \\ &= \eta \mathop{\mathbb{E}}_{y \sim \mathrm{Unif}(\{\pm 1\})} (-1)^{\mathbb{1}\{a_1 \cdots a_p = y\} + \mathbb{1}\{b_1 \cdots b_p = y\}} \\ &= \eta \cdot (-1)^{\mathbb{1}\{a_1 \cdots a_p = b_1 \cdots b_p\}} \\ &= \eta \cdot a_1 b_1 \cdots a_p b_p, \end{aligned}$$

where we use that we have chosen to work with $a_i, b_j \in \{\pm 1\}$. Note that, up to an appropriate permutation of all axes, this means that

$$\boldsymbol{T} = \eta \begin{bmatrix} +1 \\ -1 \\ -1 \\ +1 \end{bmatrix}^{\otimes p}.$$

In particular then, every marginalization $\boldsymbol{T}^{(j)}$ for $j < p$ will be zero (since the vector above is orthogonal to $\mathbf{1}$), and the marginal order of this model is

$$p_* = p.$$

Thus, we may apply Theorem 15 and find that, given a sequence of degrees $D(n)$, no LCDF of degree $D(n)$ will achieve strong separation provided that

$$\eta = \eta(n) = O\left(n^{-\frac{p}{2}} D(n)^{-\frac{p-2}{2}} \right).$$

Since the number of clauses (in expectation and typically within small error) in such random formulas is about $m = \eta \cdot \binom{n}{p} \sim \eta \cdot n^p$, this condition is equivalent to having the number of clauses scale as

$$m = m(n) = O\left(n^{\frac{p}{2}} D(n)^{-\frac{p-2}{2}} \right).$$

In the regime $D(n) = n^\delta$, this lower bound precisely complements (under the heuristic that LCDF of degree $D(n)$ correspond to $\exp(\widetilde{\Theta}(D(n)))$-time algorithms) the scaling of the runtime of at least two families of subexponential time algorithms for this problem (or its variant with a small amount of noise): the spectral and sum-of-squares algorithms of Raghavendra et al. (2017), and the simpler spectral algorithms based on the Kikuchi hierarchy of Wein et al. (2019).

C.4. Group Problems

Recall that for the problems over finite groups G that we study, we write $k := |G|$, which coincides with the parameter k of the GSBMs we consider.

C.4.1. Proof of Theorem 22: Truth-or-Haar Synchronization

In the truth-or-Haar model of group synchronization, we choose n elements of G independently and uniformly at random, say $g_1, \dots, g_n$, and for each $i < j$ we observe y_{ij} which is $g_i g_j^{-1}$ with probability $\eta = \eta(n)$ and a uniformly random element of G with probability $1 - \eta$ (hence the name, since the Haar measure on G is just the uniform measure). We will ultimately be interested in the scaling $\eta(n) = \gamma/\sqrt{n}$ for a constant γ, but for convenience for now work with the η parameter instead. Let us write μ_{avg} for the uniform measure on G (which will indeed be μ_{avg} in the GSBM setting as well, as we will see momentarily).

Proof Writing this as a GSBM, the measures involved are indexed by pairs $g, h \in G$, and we have

$$\mu_{(g,h)} = \eta\delta_{gh^{-1}} + (1-\eta)\mu_{\mathsf{avg}}$$

by the above definition. We see that μ_{avg} is indeed the average of the $\mu_{(g,h)}$. We note also that this model is *not* strongly symmetric for most G, since $gh^{-1} \neq hg^{-1} = (gh^{-1})^{-1}$ unless every element of G has order 1 or 2. However, from the calculations below we will see that the characteristic matrix is symmetric and thus that the model is weakly symmetric, so our tools still apply.

Remark 47 *We see also that in fact this model can be obtained as the $(1-\eta)$-resampling (in the sense of Definition 25) of the "noiseless" GSBM where we start out with simply observing the exact group element differences, $\mu_{(g,h)} = \delta_{gh^{-1}}$.*

We then also have

$$\bar{\mu}_{(g,h)} = \eta(\delta_{gh^{-1}} - \mu_{\mathsf{avg}}).$$

Let $\boldsymbol{P}_g \in \{0,1\}^{k\times k}$ be the permutation matrix associated to the permutation of g acting by multiplication on the left on G (i.e., the image of G under the left regular representation of G). The characteristic matrix is then

$$\begin{aligned}
\boldsymbol{T} &= \frac{\eta^2}{2}\sum_{g\in G} k\left(\boldsymbol{P}_g - \frac{1}{k}\boldsymbol{1}_k\boldsymbol{1}_k^\top\right)^{\otimes 2} \\
&= \frac{\eta^2}{2}\Bigg(k\underbrace{\sum_{g\in G}\boldsymbol{P}_g^{\otimes 2}}_{=:\boldsymbol{M}} - \boldsymbol{1}_{k^2}\boldsymbol{1}_{k^2}^\top,\Bigg)
\end{aligned}$$

where we expand the tensor product and use that $\sum_{g\in G}\boldsymbol{P}_g = \boldsymbol{1}_k\boldsymbol{1}_k^\top$. From this expression it follows that $\boldsymbol{T}$ is a symmetric matrix and thus that the GSBM is weakly symmetric, since $(\boldsymbol{P}_g)^\top = \boldsymbol{P}_{g^{-1}}$ so transposition merely permutes the terms in the summation above.

We must understand the eigenvalues of this matrix. Consider the entries of the matrix $\boldsymbol{M}$: it is indexed by pairs $(g,h) \in G^2$, and we have

$$\begin{aligned}
(\boldsymbol{P}_f^{\otimes 2})_{(g_1,h_1),(g_2,h_2)} &= \mathbb{1}\{g_1g_2^{-1} = h_1h_2^{-1} = f\}, \\
\boldsymbol{M}_{(g_1,h_1),(g_2,h_2)} &= \mathbb{1}\{g_1g_2^{-1} = h_1h_2^{-1}\} \\
&= \mathbb{1}\{g_1^{-1}h_1 = g_2^{-1}h_2\}.
\end{aligned}$$

But, this latter expression means that $\boldsymbol{M}$, after suitably permuting the rows and columns, is just the block matrix $\boldsymbol{I}_k \otimes \boldsymbol{1}_k\boldsymbol{1}_k^\top$, whose diagonal $k\times k$ blocks are the all-ones matrix and whose other blocks are zero. $\mathbf{1}_{k^2}$ is an eigenvector of this matrix with eigenvalue k, but this eigenspace has dimension $k > 1$. Therefore, we have

$$\|\boldsymbol{T}\| = \frac{k^2\eta^2}{2}.$$

Theorem 16 gives that functions of coordinate degree $O(n\,/\log n)$ cannot achieve strong separation in this GSBM once the above quantity is smaller by a constant factor than $\frac{k^2}{2n}$. Thus, if we have

$$\eta = \eta(n) < \frac{1-\varepsilon}{\sqrt{n}},$$

i.e. if $\gamma < 1$, then we obtain the stated lower bound. ■

C.4.2. PROOF OF THEOREM 24: TRUTH-OR-HAAR SUMSET

Proof This model is identical to the synchronization model, but instead of observing gh^{-1} or a uniformly random element, we observe gh or a uniformly random element. In other words, our GSBM is now specified by channel measures

$$\mu_{(g,h)} = \eta\delta_{gh} + (1-\eta)\mu_{\mathsf{avg}}$$

Recall that we further assume in this case that G is abelian.

At a glance, this model might seem quite different from the previous one. Indeed, the intermediate calculations are different: we may carry out the same plan, but instead of working with the permutation matrices $\boldsymbol{P}_f$ with entries

$$(\boldsymbol{P}_f)_{gh} = \mathbb{1}\{gh^{-1} = f\},$$

we must work with $\widetilde{\boldsymbol{P}}_f$ with entries

$$(\widetilde{\boldsymbol{P}}_f)_{gh} = \mathbb{1}\{gh = f\}.$$

These are still permutations and still have $\sum_{g\in G}\widetilde{\boldsymbol{P}}_g = \mathbf{1}_k\mathbf{1}_k^\top$, so the first part of the argument remains unaffected, and we reach the characteristic matrix

$$\boldsymbol{T} = \frac{\eta^2}{2}\left(\underbrace{k\sum_{g\in G}\widetilde{\boldsymbol{P}}_g^{\otimes 2}}_{=:\widetilde{\boldsymbol{M}}} - \mathbf{1}_{k^2}\mathbf{1}_{k^2}^\top,\right).$$

In understanding $\widetilde{\boldsymbol{M}}$, we must use that now we assume G is abelian: without that assumption, we could only reach

$$\begin{aligned}
(\widetilde{\boldsymbol{P}}_f^{\otimes 2})_{(g_1,h_1),(g_2,h_2)} &= \mathbb{1}\{g_1g_2 = h_1h_2 = f\}, \\
\widetilde{\boldsymbol{M}}_{(g_1,h_1),(g_2,h_2)} &= \mathbb{1}\{g_1g_2 = h_1h_2\} \\
&= \mathbb{1}\{g_1^{-1}h_1 = g_2h_2^{-1}\}
\end{aligned}$$

but assuming that G is abelian we may continue

$$= \mathbb{1}\{g_1^{-1}h_1 = h_2^{-1}g_2\}$$
$$= \mathbb{1}\{g_1^{-1}h_1 = (g_2^{-1}h_2)^{-1}\}.$$

Thus $\widetilde{\boldsymbol{M}}$ is the same as $\boldsymbol{M}$ from the group synchronization case, but with an involution applied to the columns. After permuting the rows and columns such that this involution reverses the order of the columns, $\widetilde{\boldsymbol{M}}$ is then $\widetilde{\boldsymbol{I}}_k \otimes \mathbf{1}_k\mathbf{1}_k^\top$, where $\widetilde{\boldsymbol{I}}_k$ is the $k \times k$ "anti-identity" matrix:

$$\widetilde{\boldsymbol{I}}_k = \begin{bmatrix} & & & & 1 \\ & & & 1 & \\ & & \iddots & & \\ & 1 & & & \\ 1 & & & & \end{bmatrix}.$$

As this matrix is symmetric, so is $\widetilde{\boldsymbol{M}}$, and we see that $\boldsymbol{T}$ is symmetric also and so this model is weakly symmetric. But, this would not necessarily be the case if G were not abelian (like for small non-abelian groups including the symmetric group $\mathrm{Sym}([3])$).

It is easily verified that $\widetilde{\boldsymbol{I}}_k$ has the eigenvalue 1 with multiplicity $\lceil k/2 \rceil$, and the remaining eigenvalues are all -1. Thus, we still always have

$$\|\boldsymbol{T}\| = \frac{k^2\eta^2}{2},$$

and the remainder of the argument goes through as for synchronization to derive the (identical) lower bound against LCDF. ■

C.4.3. Gaussian Multi-Frequency Synchronization

We briefly summarize how our technical results in Appendix A.3 imply an improvement to the recent results of Kireeva et al. (2024) on a related synchronization model. For the sake of explanation let us focus on cyclic groups, though the improvement also holds for their results on general finite groups and also the circle group $U(1)$ (which is analyzed by reducing to cyclic groups). That work studies the following model for the cyclic groups. We write $\mathsf{GUE}(n)$ for the law of a random Hermitian matrix $\boldsymbol{W} \in \mathbb{C}^{n\times n}$ whose entries on and above the diagonal are independent, with diagonal entries distributed as $\mathcal{N}(0, 2)$ and off-diagonal entries having real and imaginary parts distributed independently as $\mathcal{N}(0, 1/2)$.

Definition 48 (Multi-frequency $\mathbb{Z}_k$ synchronization) *Let $k \geq 2$ and $\lambda \geq 0$. We consider two models of $\lceil k/2 \rceil - 1$ observations of $n \times n$ Hermitian matrices:*

1. *Under $\mathbb{Q}_n$, observe $\boldsymbol{Y}_\ell \sim \mathsf{GUE}(n)$ for $\ell = 1, \dots, \lceil k/2 \rceil - 1$.*
2. *Under $\mathbb{P}_n$, draw $\boldsymbol{x} \in \mathbb{C}^n$ a random vector such that each entry is sampled independently as $x_j \sim \mathrm{Unif}(\{\omega^0, \dots, \omega^{k-1}\})$, where $\omega = \exp(2\pi i/k)$ is a primitive root of unity. Also, draw $\boldsymbol{W}_\ell \sim \mathsf{GUE}(n)$ for $\ell = 1, \dots, \lceil k/2 \rceil - 1$. Then, observe*

$$\boldsymbol{Y}_\ell = \frac{\lambda}{n}\boldsymbol{x}^{(\ell)}\boldsymbol{x}^{(\ell)*} + \frac{1}{\sqrt{n}}\boldsymbol{W}_\ell, \text{ for } \ell = 1, \dots \lceil k/2 \rceil - 1,$$

where $\boldsymbol{x}^{(\ell)}$ denotes the ℓth entrywise power.

The general conjecture concerning this model, first proposed by Perry et al. (2016), is that testing should be computationally hard whenever $\lambda < 1$ (the same as the computational threshold for a single such spiked matrix observation, meaning that the presence of extra observations is not helpful). In Kireeva et al. (2024), it is shown that no sequence of polynomials of degree $o(n^{1/3})$ can strongly separate $\mathbb{Q}_n$ from $\mathbb{P}_n$. We improve on this scaling as follows:

Theorem 49 *For any $\lambda < 1$, there exists $\gamma = \gamma(\lambda) > 0$ such that no polynomial of degree at most γn can strongly separate $\mathbb{Q}_n$ from $\mathbb{P}_n$ in the model of Definition 48.*

Proof We will work with the polynomial advantage, denoted $\mathsf{Adv}_{\leq D}$ and given by the same expression as in our Definition 39 of the coordinate advantage, but optimizing over $\deg(f) \leq D$ for f a polynomial rather than $\mathsf{cdeg}(f) \leq D$:

$$\mathsf{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) := \left\{ \begin{array}{ll} \text{maximize} & \mathbb{E}_{\boldsymbol{y} \sim \mathbb{P}} f(\boldsymbol{y}) \\ \text{subject to} & \mathbb{E}_{\boldsymbol{y} \sim \mathbb{Q}} f(\boldsymbol{y})^2 \leq 1, \\ & f \in \mathbb{R}[\boldsymbol{y}], \\ & \deg(f) \leq D \end{array} \right\}.$$

The analog of Proposition 40 holds for the polynomial advantage and strong separation by polynomials, so that to prove the Theorem it suffices to show that $\mathsf{Adv}_{\leq D(n)}(\mathbb{P}_n, \mathbb{Q}_n) = O(1)$ whenever $D(n) \leq \gamma n$.

Let us choose $\varepsilon = \varepsilon(\lambda) > 0$ such that $\lambda^2 \cdot \frac{2+\varepsilon}{2} \leq 1 - \varepsilon$. We compute starting from Equation (6.4) of Kireeva et al. (2024), which expresses the advantage in terms of expectations over multinomial vectors. In fact, we may relate the expectations appearing there to Pearson's χ^2 statistics:

$$\begin{aligned} \mathsf{Adv}_{\leq D(n)}(\mathbb{P}_n, \mathbb{Q}_n)^2 &= \sum_{d=0}^{D(n)} \frac{1}{d!} \frac{\lambda^{2d}}{n^d} \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathsf{Mult}(n,k)} \left(\frac{k}{2} \sum_{\ell=1}^{k} \left(z_\ell - \frac{n}{k} \right)^2 \right)^d \\ &= \sum_{d=0}^{D(n)} \frac{1}{d!} \frac{\lambda^{2d}}{2^d} \mathop{\mathbb{E}}_{X \sim \chi^2_{\mathsf{Pear}}(n,k)} X^d \end{aligned}$$

and now by Corollary 34, there exist $C, \gamma > 0$ depending on our ε chosen above (and thus in turn depending only on λ) such that, if $D(n) \leq \gamma n$, then

$$\begin{aligned} &\leq C^k \sum_{d=0}^{D(n)} \frac{1}{d!} \frac{\lambda^{2d}}{2^d} d^{3/2} \left(\frac{(2+\varepsilon)d}{e} \right)^d \\ &\leq C^k \sum_{d=0}^{D(n)} \frac{1}{d!} d^{3/2} \left(\frac{(1-\varepsilon)d}{e} \right)^d \end{aligned}$$

and bounding the factorial by Proposition 37 we have

$$\leq C^k \sum_{d=0}^{\infty} d^{3/2} (1-\varepsilon)^d$$

which converges, completing the proof. ■