Reducibility and Computational Lower Bounds for Problems with Planted Sparse Structure

Matthew Brennan[*]   Guy Bresler[†]   Wasim Huleihel[‡]

June 21, 2018

Abstract

Recently, research in unsupervised learning has gravitated towards exploring statistical-computational gaps induced by sparsity. A line of work initiated in [BR13a] has aimed to explain these gaps through reductions from conjecturally hard problems in complexity theory. However, the delicate nature of average-case reductions has limited the development of techniques and often led to weaker hardness results that only apply to algorithms robust to different noise distributions or that do not need to know the parameters of the problem. We introduce several new techniques to give a web of average-case reductions showing strong computational lower bounds based on the planted clique conjecture. Our new lower bounds include:

- **Planted Independent Set:** We show tight lower bounds for detecting a planted independent set of size k in a sparse Erdős-Rényi graph of size n with edge density $\tilde{\Theta}(n^{-\alpha})$.

- **Planted Dense Subgraph:** If $p > q$ are the edge densities inside and outside of the community, we show the first lower bounds for the general regime $q = \tilde{\Theta}(n^{-\alpha})$ and $p - q = \tilde{\Theta}(n^{-\gamma})$ where $\gamma \geq \alpha$, matching the lower bounds predicted in [CX16]. Our lower bounds apply to a deterministic community size k, resolving a question raised in [HWX15].

- **Biclustering:** We show lower bounds for the canonical simple hypothesis testing formulation of Gaussian biclustering, slightly strengthening the result in [MW15b].

- **Sparse Rank-1 Submatrix:** We show that detection in the sparse spiked Wigner model is often harder than biclustering, and are able to obtain tight lower bounds for these two problems with different reductions from planted clique.

- **Sparse PCA:** We give a reduction between sparse rank-1 submatrix and sparse PCA to obtain tight lower bounds in the less sparse regime $k \gg \sqrt{n}$, when the spectral algorithm is optimal over the natural SDP. We give an alternate reduction recovering the lower bounds of [BR13a, GMZ17] in the simple hypothesis testing variant of sparse PCA. We also observe a subtlety in the complexity of sparse PCA that arises when the planted vector is biased.

- **Subgraph Stochastic Block Model:** We introduce a model where two small communities are planted in an Erdős-Rényi graph of the same average edge density and give tight lower bounds yielding different hard regimes than planted dense subgraph.

Our results demonstrate that, despite the delicate nature of average-case reductions, using natural problems as intermediates can often be beneficial, as is the case in worst-case complexity. Our main technical contribution is to introduce a set of techniques for average-case reductions that: (1) maintain the level of signal in an instance of a problem; (2) alter its planted structure; and (3) map two initial high-dimensional distributions simultaneously to two target distributions approximately under total variation. We also give algorithms matching our lower bounds and identify the information-theoretic limits of the models we consider.

[*]Massachusetts Institute of Technology. Department of EECS. Email: `brennanm@mit.edu`.

[†]Massachusetts Institute of Technology. Department of EECS. Email: `guy@mit.edu`.

[‡]Massachusetts Institute of Technology. Research Laboratory of Electronics. Email: `wasimh@mit.edu`.

Contents

1 Introduction

The field of statistics is undergoing a dramatic conceptual shift, with computation moving from the periphery to center stage. Prompted by the demands of modern data analysis, researchers realized two decades ago that a new approach to estimation was needed for high-dimensional problems in which the dimensionality of the data is at least as large as the sample size. High-dimensional problems are inherently underdetermined, often precluding nontrivial rates of estimation. However, this issue typically disappears if the underlying signal is known to have an appropriate structure, such as low rank or sparsity. As a result, high-dimensional structured estimation problems have received significant attention in both the statistics and computer science communities. Prominent examples include estimating a sparse vector from linear observations, sparse phase retrieval, low-rank matrix estimation, community detection, subgraph and matrix recovery problems, random constraint satisfiability, sparse principal component analysis and covariance matrix estimation. Although structural assumptions can yield nontrivial estimation rates, the statistically optimal estimators for these problems typically entail an exhaustive search over the set of possible structures and are thus not efficiently computable. Conversely, all known efficient algorithms for these problems are statistically suboptimal, requiring more data than strictly necessary. This phenomenon has led to a number of conjectured statistical-computational gaps for high-dimensional problems with structure. This raises an intriguing question: how are these gaps related to one another and are they emerging for a common reason?

In the last few years, several lines of work have emerged to make rigorous the notion of what is and what is not achievable statistically by efficient algorithms. In the seminal work of [BR13a], a conjectured statistical-computational gap for sparse principal component analysis (PCA) was shown to follow from the planted clique conjecture. This marked the first result basing the hardness of a natural statistics problem on an average-case hardness assumption and produced a framework for showing statistical-computational gaps by approximately mapping in total variation. This subsequently led to several more reductions from the planted clique conjecture to show statistical-computational gaps for problems including submatrix detection/biclustering [MW15b], submatrix localization [CLR15], planted dense subgraph [HWX15], RIP certification [WBP16], sparse PCA and sparse canonical correlation analysis [WBS16, GMZ17]. We draw heavily from the framework for average-case reductions laid out in these papers. More recently, focus has shifted to showing unconditional hardness results for restricted models of computation and classes of algorithms. An exciting line of work has emerged surrounding applications of the Sum of Squares (SOS) semidefinite programming hierarchy to problems with statistical-computational gaps. SOS Lower bounds have been shown for planted clique [BHK+16] and for sparse PCA [KNV15, MW15a, HKP+17]. Tight computational lower bounds have also been shown in the statistical query model for planted clique and planted random k-SAT [FGR+12, FPV15].

One reason behind this focus on showing hardness in restricted models of computation is that average-case reductions are inherently delicate, creating obstacles to obtaining satisfying hardness results. As described in [Bar17], these technical obstacles have left us with an unsatisfying theory of average-case hardness. Reductions in worst-case complexity typically take a *general* instance of a problem A to a *structured* instance of a problem B. For example, a classic reduction from 3SAT to INDEPENDENT-SET produces a very specific type of graph with a cluster of seven vertices per clause corresponding to each satisfying assignment such that two vertices are connected if together they yield an inconsistent assignment. If such a reduction were applied to a random 3SAT instance, the resulting graph instance would be far from any natural graph distribution. Unlike reductions in worst-case complexity, average-case reductions between natural decision problems need to precisely map the distributions on instances to one another without destroying the underlying signal in

polynomial-time. The delicate nature of this task has severely limited the development of techniques and left open reductions between decision problems that seem to be obviously equivalent from the standpoint of algorithm design. For example, it remains unknown whether refuting random constraint satisfaction problems with $10m$ clauses is equivalent to refuting those with $11m$ clauses or whether the planted clique conjecture at edge density $1/2$ implies the conjecture at edge density 0.49. There are also a variety of negative results further demonstrating new obstacles associated with average-case complexity, including that it likely cannot be based on worst-case complexity [BT06a]. For more on average-case complexity, we refer to the survey of [BT+06b].

In order to overcome these average-case difficulties, prior reductions have often made assumptions on the robustness of the underlying algorithm such as that it succeeds for any noise distributions from a fixed class as in [BR13a, WBS16, CLR15]. This corresponds to composite vs. composite hypothesis testing formulations of detection problems, where the composite null hypothesis H_0 consists of the class of noise distributions. Other reductions have shown hardness for precise noise distributions but for algorithms that do not need to exactly know the parameters of the given instance [MW15b, GMZ17]. This typically corresponds to simple vs. composite hypothesis testing where the composite alternative H_1 consists of models defined by varying parameters such as the sparsity k or signal strength. The strongest prior reduction from planted clique is that to the sparsest regime of planted dense subgraph in [HWX15]. A lower bound is shown for a simple vs. simple hypothesis testing variant of the problem, with each consisting of a single distribution. However, the community in their formulation of planted dense subgraph was binomially distributed and therefore still assumed to be unknown exactly to the algorithm. Prior reductions have also shown hardness at particular points in the parameter space, deducing that an algorithm cannot always perform better than a conjectured computational barrier rather than showing that no algorithm can ever perform better. For example, prior reductions for sparse PCA have only shown tight hardness around the single parameter point where the signal is $\theta = \tilde{\Theta}(1)$ and the sparsity is $k = \tilde{\Theta}(\sqrt{n})$. Simplifying parameters in their reductions, both [BR13a] and [GMZ17] approximately map a planted clique instance on n vertices with clique size k to an instance of sparse PCA with $\theta \approx \tilde{\Theta}(k^2/n)$ which is only tight to the conjectured barrier of $\theta^* = \Theta(\sqrt{k^2/n})$ when $k = \tilde{\Theta}(\sqrt{n})$.

These assumptions leave a subtle disparity between the existing average-case lower bounds for many problems and algorithmic upper bounds. Many algorithmic results assume a canonical generative model or implicitly assume knowledge of parameters. For example, even in the recent literature on robust algorithms for problems with sparsity in [BDLS17, Li17], the setup is in the context of specific canonical generating models, such as the spiked covariance model for sparse PCA. Even when corrupted by adversarial noise, the spiked covariance model is far in distribution from many sub-Gaussian formulations of sparse PCA. Despite existing average-case lower bounds, hardness for the canonical generative models for many problems has remained open. This includes biclustering with a flat planted $k \times k$ submatrix selected uniformly at random in Gaussian noise, sparse PCA with a k-sparse principal component chosen uniformly at random to have entries equal to $\pm 1/\sqrt{k}$ and planted dense subgraph with deterministic community size.

1.1 Overview

The aim of this paper is threefold: (1) to demonstrate that a web of average-case reductions among problems with statistical-computational gaps is feasible even for showing strong computational lower bounds; (2) to introduce a number of new techniques for average-case reductions between problems; and (3) to fully characterize the computationally hard regime of several models. The graph of our reductions is shown in Figure 1.1. Our new lower bounds are as follows.

- **Planted Independent Set:** We show tight lower bounds for detecting a planted independent
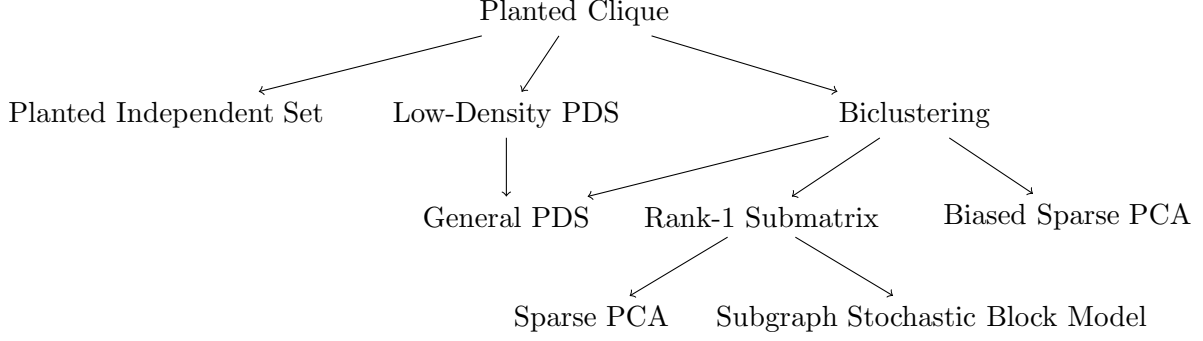
Figure 1: Graph of average-case reductions for detection problems showing tight statistical-computational gaps given the planted clique conjecture.

set of size k in a sparse Erdős-Rényi graph of size n with edge density $\tilde{\Theta}(n^{-\alpha})$.

- **Planted Dense Subgraph:** If $p > q$ are the edge densities inside and outside of the community, we show the first lower bounds for the general regime $q = \tilde{\Theta}(n^{-\alpha})$ and $p - q = \tilde{\Theta}(n^{-\gamma})$ where $\gamma \geq \alpha$, matching the lower bounds predicted in [CX16]. Our lower bounds apply to a deterministic community size k, resolving a question raised in [HWX15].

- **Biclustering:** We show lower bounds for Gaussian biclustering as a simple hypothesis testing problem to detect a uniformly at random planted flat $k \times k$ submatrix. Our alternative reduction matches the barriers in [MW15b], where a computational lower bound was shown for a composite hypothesis testing variant of biclustering. We show hardness for the natural simple hypothesis testing problem where the $k \times k$ submatrix is chosen uniformly at random and has equal entries.

- **Sparse Rank-1 Submatrix:** We show that detection in the sparse spiked Wigner model has a different computational threshold from biclustering when $k \gg \sqrt{n}$. Surprisingly, we are able to obtain tight lower bounds matching these different detection thresholds with different reductions from planted clique.

- **Sparse PCA:** We give a reduction between rank-1 submatrix and sparse PCA to obtain tight lower bounds in the less sparse regime $k \gg \sqrt{n}$, when the spectral algorithm is optimal over the SDP. This yields the first tight characterization of a computational barrier for sparse PCA over an entire parameter regime. We also give an alternate reduction recovering the lower bounds of [BR13a] and [GMZ17] in the canonical simple hypothesis testing variant of sparse PCA.

- **Biased Sparse PCA:** We show that any assumption on the sparse principal component having a constant fraction more or fewer positive entries than negative entries yields a detection-recovery gap that is not present in sparse PCA.

- **Subgraph Stochastic Block Model:** We introduce a model where two small communities are planted in an Erdős-Rényi graph of the same average edge density. Parallel to the difference between biclustering and sparse rank-1 submatrix when $k \gg \sqrt{n}$, we show that detection in this model is much harder than in planted dense subgraph when $k \gg \sqrt{n}$.

Our lower bounds for planted independent set, the general regime of planted dense subgraph, sparse rank-1 submatrix, sparse PCA when $k \gg \sqrt{n}$, biased sparse PCA and the subgraph stochastic

block model are novel. As previously mentioned, lower bounds for sparse PCA when $k \ll \sqrt{n}$, for biclustering and for planted dense subgraph in the sparsest regime were previously known. In each of these cases, we strengthen the existing lower bounds to the apply to the canonical generative model. We show computational lower bounds for simple vs. simple hypothesis testing in all cases other than for sparse PCA, rank-1 submatrix and the subgraph stochastic block model all in the regime $k \gg \sqrt{n}$. This is a consequence of our underlying reduction technique, reflection cloning, and appears unavoidable given our methods. However, we do show that the distribution we reduce to is in some sense close to the canonical generative model.

Our results demonstrate that, despite the delicate nature of average-case reductions, using natural problems as intermediates can often be beneficial as in reductions in worst-case complexity. Our main technical contribution is to introduce several techniques for mapping problems approximately in total variation without degrading the underlying planted sparse structure. These techniques are:

- **Distributional Lifting:** A variant of graph lifts that iteratively maps to intermediate matrices with entries from chosen distributions. Varying the underlying distribution produces different relationships between the resulting edge density and size of a planted subgraph. This is in introduced in Sections 4, 5 and 6.

- **Rejection Kernels:** A general framework for a change in measure of the underlying noise distribution while preserving the planted sparse structure. This unifies ideas introduced in [HWX15], [MW15b] and [GMZ17]. This is introduced in Section 5.

- **Reflection Cloning:** A method of increasing the sparsity of a planted rank-1 structure in noise while preserving the level of signal in the planted structure. This is introduced in Section 7.

- **Random Rotations for Sparse PCA:** An average-case connection between the sparse spiked Wigner model and the spiked covariance model of sparse PCA. This is introduced in Section 8.

We consider two main variants of distributional lifting using matrices with Poisson and Gaussian entries as intermediates. Poisson and Gaussian lifting lead to two very different parameter scalings, which when combined fully characterize general planted dense subgraph. Reflection cloning can be viewed as a more randomness-efficient variant of Gaussian lifting that we use to show sharper lower bounds for the sparse spiked Wigner model and sparse PCA. We also give algorithms matching our lower bounds and identify the information-theoretic limits of the problems that we consider in Section 9.

1.2   Hardness Results from an Algorithmic Perspective

In this section, we motivate our computational lower bounds and techniques using algorithms for sparse PCA as an example. Consider the detection problem for sparse PCA where either $X_1, X_2, \ldots, X_n$ are sampled i.i.d. from $N(0, I_d)$ or are sampled i.i.d. from $N(0, I_d + \theta vv^\top)$ for some latent k-sparse unit vector v with nonzero entries equal to $\pm 1/\sqrt{k}$. The task is to detect which of the two distributions the samples originated from. For now assume that $d = \Theta(n)$. Now consider the following four algorithms:

1. **Semidefinite Programming:** Form the empirical covariance matrix $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^\top$
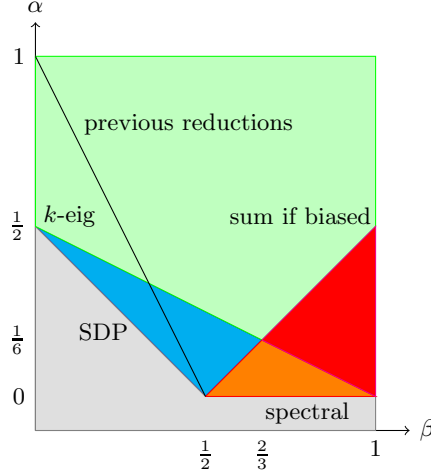
7

Figure 2: Algorithms for sparse PCA with $d = \Theta(n)$, $k = \tilde{\Theta}(n^\beta)$ and $\theta = \tilde{\Theta}(n^{-\alpha})$. Lines represent the strongest guarantees of each algorithm. The line marked as previous reductions shows the strongest previously known planted clique lower bounds for sparse PCA when $k \lesssim \sqrt{n}$. No planted clique lower bounds were known for $k \gg \sqrt{n}$.

and solve the convex program

$$\max_Z \quad \mathrm{Tr}\left(\hat{\Sigma}Z\right)$$
$$\text{s.t.} \quad \mathrm{Tr}(Z) = 1, |Z|_1 \le k, Z \succeq 0$$

As shown in [BR13a], thresholding the resulting maximum solves the detection problem as long as $\theta = \tilde{\Omega}(\sqrt{k^2/n})$.

2. **Spectral Algorithm:** Threshold the maximum eigenvalue of $\hat{\Sigma}$. If the data are sampled from $N(0, I_d)$, then the largest eigenvalue is with high probability at most

$$\lambda_{\max}(\hat{\Sigma}) \le \frac{d}{n} + \sqrt{\frac{d}{n}} + 1 + o(1)$$

by standard bounds on the singular values of random Gaussian matrices. Since $d = \Theta(n)$, this algorithm succeeds as long as $\theta = \Omega(1)$. This algorithm was considered in [KNV15].

3. **Sum Test:** Sum the entries of $\hat{\Sigma}$ and threshold the absolute value of the sum. If v has sum exactly zero, then this test will not succeed. However, if we assume that $\hat{\Sigma}$ has even 51% of its nonzero entries of one sign, then this test succeeds if $\theta = \tilde{\Omega}(\sqrt{n}/k)$.

4. **k-Sparse Eigenvalue:** Compute and threshold the k-sparse unit vector u that maximizes $u^\top \hat{\Sigma} u$. This can be found by finding the largest eigenvector of each $k \times k$ principal submatrix of $\hat{\Sigma}$. Note that this takes exponential time. It was shown in [BR13a] that this succeeds as long as $\theta = \tilde{\Omega}(\sqrt{k/n})$.

The boundaries at which these algorithms begin to succeed are shown in Figure 2 for the regime $k = \tilde{\Theta}(n^\beta)$ and $\theta = \tilde{\Theta}(n^{-\alpha})$. The computational lower bound mapping to $\theta \approx k^2/n$ in [BR13a] and [GMZ17] is also drawn. As shown, the only point in the parameter diagram for which it matches an algorithmic upper bound is $\alpha = 0$ and $\beta = 1/2$, corresponding to when $\theta = \tilde{\Theta}(1)$ and $k = \tilde{\Theta}(\sqrt{n})$.

8

For sparse PCA with $d = \Theta(n)$, the SDP is the best known algorithm up to $k = \Theta(\sqrt{n})$, at which point the spectral algorithm has stronger guarantees. This algorithmic transition at $k = \Theta(\sqrt{n})$ is characteristic of all of the problems we consider. For the biased variant of sparse PCA where the sum test succeeds, the sum test always does strictly better than the spectral algorithm. Furthermore, the biased variant ceases to have a statistical computational gap around $k = \Theta(n^{2/3})$. While the sum test yields an improved algorithm for detection, unlike the other three algorithms considered above, it does not translate into an algorithm for recovering the support of the sparse component. Given a conjecture about recovery in planted dense subgraph, we show that the best recovery algorithm for biased sparse PCA can only match the guarantees of the spectral algorithm. Thus the biased variant induces a detection-recovery gap when $k \gg \sqrt{n}$. We show that the disappearance of a statistical computation gap at $k = \Theta(n^{2/3})$ and a detection-recovery gap when $k \gg \sqrt{n}$ are features of the problems we consider that admit a sum test. These are biased sparse PCA, planted independent set, planted dense subgraph and biclustering. Distributional lifting gives tight planted clique lower bounds for these problems.

In contrast, rank-1 submatrix, the subgraph stochastic block model and sparse PCA do not admit a sum test. Given the planted clique conjecture, rank-1 submatrix and sparse PCA have no detection-recovery gap and retain their statistical-computational gap for all sparsities k. Reflection cloning shows tight lower bounds for these problems in the regime $k \gg \sqrt{n}$, where spectral algorithms become optimal. It is surprising that the planted clique conjecture can tightly capture completely different sets of computational barriers for different problems, illustrating its power as an average-case hardness assumption. Although analogues of the sum test, spectral algorithms and semidefinite programs all have equivalent guarantees up to logarithmic factors for planted clique, reductions from planted clique show tight hardness in problems for which this is not true.

1.3 Prior Work

This work is part of a growing body of literature giving rigorous evidence for computational-statistical gaps in high-dimensional inference problems. We focus on average-case reductions to directly relate computational-statistical gaps in different problems, as opposed to giving worst-case evidence for hardness in statistical problems [ZWJ14, HMRW14, CPR16]. A survey of prior results on computational-statistical gaps with a focus on predictions from statistical physics can be found in [BPW18] and a general analysis of gaps for algorithms from several convex relaxation hierarchies can be found in [CJ13].

Planted Clique and Independent Set. Our computational lower bounds are based on average-case reductions from the problem of finding a planted clique of size k in an Erdős-Rényi graph with n vertices. The planted clique problem was introduced in [AKS98], where a spectral algorithm was shown to recover the planted clique if $k = \Omega(\sqrt{n})$. A number of algorithms for planted clique have since been studied, including approximate message passing, semidefinite programming, nuclear norm minimization and several combinatorial approaches [FK00, McS01, FR10, AV11, DGGP14, DM15a, CX16]. All of these algorithms require that $k = \Omega(\sqrt{n})$, which has led to the planted clique conjecture that no polynomial time algorithm can recover the planted clique if $k = o(\sqrt{n})$. It was also shown in [AAK+07] that recovering and detecting the planted clique are equivalent up to $\log n$ factors in k. There have been a number of previous average-case reductions from the planted clique conjecture, which we discuss in more detail in the prior work section on average-case reductions.

Several works have considered finding independent sets in sparse Erdős-Rényi graphs, similar to the regime with edge density $q = \tilde{\Theta}(n^{-\alpha})$ where $\alpha \in (0, 2)$ we consider here. In [COE15, GS14,

RV$^+$17], the authors examine greedy and local algorithms to find independent sets in the regime $q = \tilde{\Theta}(n^{-1})$ in random regular, Erdős-Rényi and other random graphs. In [FO05], a spectral algorithm is given to find a planted independent set in the regime $q = \tilde{\Theta}(n^{-1})$ and in [CO03], the planted independent set recovery problem is shown to be possible in polynomial time in the regime $\alpha \in (0, 1)$ when $q \gg \frac{n}{k^2}$ even in a semirandom model. The algorithms of [CX16] also apply to recovering planted independent sets after taking the complement of the input graph.

Planted Dense Subgraph and Community Detection. The planted dense subgraph detection problem was considered in [ACV$^+$14, BI13, VAC$^+$15, HWX15] and generalizations of the recovery problem were considered in [CX16, HWX16b, Mon15, CC18]. In [HWX15], a reduction from planted clique was given for the regime $p = cq$ for some constant $c > 1$ and $q = \tilde{\Theta}(n^{-\alpha})$ and k is binomially distributed, where k, n, p and q are the size of the community, size of the graph, community edge density and graph edge density, respectively. Our results strengthen this lower bound to apply for deterministic k and for all $p > q$ with $p - q = O(q)$ where $q = \tilde{\Theta}(n^{-\alpha})$. When $p = \omega(q)$, the resulting regime is the planted dense subgraph problem considered in [BCC$^+$10]. The computational barrier for this problem is conjectured to be the log-density threshold $k = \tilde{\Theta}(n^{\log_q p})$ when $k \ll \sqrt{n}$ and is achieved by very different algorithms than those that are optimal when $p = O(q)$ [CDK12, CDM17]. Recently, it was shown in [CM18] that $\tilde{\Omega}(\log n)$ rounds of the Sherali-Adams hierarchy cannot solve the planted dense subgraph detection problem below the log-density threshold in the regime $p = \omega(q)$. Hardness below the log-density threshold has been used as an average-case assumption in several reductions, as outlined in the prior work section on average-case reductions.

Community detection in the stochastic block model has been the focus of an extensive body of literature surveyed in [Abb17]. It recently has been shown that the two-community stochastic block model does not exhibit statistical-computational gaps for partial and exact recovery, which are possible when the edge density scales like $\Theta(n^{-1})$ [MNS12, MNS13, Mas14] and $\Theta(n^{-1} \log n)$ [MNS14, HWX16a, ABH16], respectively. In contrast, the subgraph variant of the two-community stochastic block model that we introduce has computational-statistical gaps for partial recovery, exact recovery and detection, given the planted clique conjecture. The k-block stochastic block model is also conjectured to have statistical-computational gaps starting at $k \geq 4$ [AS15].

Biclustering and the Spiked Wigner Model. Gaussian biclustering was considered as a detection problem in [BI13, MW15b, MRZ15] and as a recovery problem in [SWP$^+$09, KBRS11, BKR$^+$11, CLR15, CX16, HWX16b]. In [MW15b], a reduction from planted clique was given for a simple vs. composite hypothesis testing variant of the biclustering detection problem, where the size and mean entries of the planted submatrix were allowed to vary. In [CLR15], submatrix localization with sub-Gaussian noise was shown to be hard assuming a variant of the planted clique conjecture for regular graphs.

A large body of literature has studied the spectrum of the spiked Wigner model [Péc06, FP07, CDMF$^+$09, BGN11]. Spectral algorithms and information-theoretic lower bounds for the spiked Wigner model detection and recovery problems were considered in [MRZ15, PWB16, PWBM16]. The sparse spiked Wigner model where the sparsity k of the planted vector satisfies $k = \Theta(n)$ was studied in [PWB16, PWBM16, BMV$^+$18]. The sparse spiked Wigner model with $k = \tilde{\Theta}(n^\beta)$ for some $\beta \in (0, 1)$ was considered in [HKP$^+$17], where the authors showed sum of squares lower bounds matching our planted clique reductions.

Sparse PCA. Since its introduction in [JL04], sparse principal component analysis has been studied broadly in the statistics and computer science communities. A number of algorithms solving sparse PCA under the spiked covariance model have been proposed [AW09, Ma13, CMW+13, BR13b, BR13a, SSM13, KNV15, DM14, WBS16]. The information-theoretic limits for detection and recovery in the spiked covariance model have also been examined extensively [AW09, VL12, BR13b, BJNP13, CMW+13, WBS16, CMW15]. The computational limits of sparse PCA problems have also been considered in the literature. Degree four SOS lower bounds for the spiked covariance model were shown in [MW15a]. In [BR13a], the authors give a reduction from planted clique to a sub-Gaussian composite vs. composite hypothesis testing formulation of sparse PCA as a detection problem, and [BR13b] gives a reduction from planted clique showing hardness for semidefinite programs. In [GMZ17], the authors give a reduction from planted clique to a simple vs. composite hypothesis testing formulation of detection in the spiked covariance model matching our reduction when $k \ll \sqrt{n}$. In [WBS16], the authors give a reduction from planted clique to a sub-Gaussian variant of the sparse PCA recovery problem. As mentioned in the introduction, these planted clique lower bounds do not match the conjectured algorithmic upper bounds when k differs in a polynomial factor from $\sqrt{n}$.

Average-Case Reductions. While the theory of worst-case complexity has flourished to the point that many natural problems are now known to be NP-hard or even NP-complete, the theory of average-case complexity is far less developed. In the seminal work of [Lev86], it was shown that an average-case complete problem exists. However, no natural problem with a natural distribution on inputs has yet been shown to be average-case complete. As mentioned in this section, there are obfuscations to basing average-case complexity on worst-case complexity [BT06a]. For more on the theory of average-case complexity, see Section 18 of [AB09] and [BT+06b].

As previously mentioned, there have been a number of average-case reductions from planted clique to average-case problems in both the computer science and statistics literature. These include reductions to testing k-wise independence [AAK+07], biclustering detection and recovery [MW15b, CLR15, CW18], planted dense subgraph [HWX15], RIP certification [WBP16, KZ14], matrix completion [Che15], minimum circuit size and minimum Kolmogorov time-bounded complexity [HS17a] and sparse PCA [BR13b, BR13a, WBS16, GMZ17]. The planted clique conjecture has also been used as a hardness assumption for average-case reductions in cryptography [JP00, ABW10], as described in Sections 2.1 and 6 of [Bar17]. There have also been a number of average-case reductions from planted clique to show worst-case lower bounds such as hardness of approximation. Planted clique has been used to show worst-case hardness of approximating densest k-subgraph [AAM+11], finding approximate Nash equilibria [MV09, HK11, ABC13], signalling [Dug14, BCKS16], approximating the minmax value of 3-player games [EHV12], aggregating pairwise comparison data [SBW16] and finding endogenously formed communities [BBB+13].

A number of average-case reductions in the literature have started with different average-case assumptions than the planted clique conjecture. Variants of planted dense subgraph have been used to show hardness in a model of financial derivatives under asymmetric information [ABBG11], link prediction [BB18], finding dense common subgraphs [CNW18] and online local learning of the size of a label set [ACLR15]. Hardness conjectures for random constraint satisfaction problems have been used to show hardness in improper learning complexity [DLSS14], learning DNFs [DSS16] and hardness of approximation [Fei02]. There has also been a recent reduction from a hypergraph variant of the planted clique conjecture to tensor PCA [ZX17].

Lower Bounds for Classes of Algorithms. As described in the introduction, recently there has been a focus on showing unconditional hardness results for restricted models of computation and classes of algorithms. In [Jer92], it was shown that the Metropolis process cannot find large cliques in samples from planted clique. The fundamental limits of spectral algorithms for biclustering and low-rank planted matrix problems were examined in [MRZ15]. Integrality gaps for SDPs solving sparse PCA, planted dense subgraph and submatrix localization were shown in [KNV15] and [CX16]. SOS lower bounds have been shown for a variety of average-case problems, including planted clique [DM15b, RS15, HKP+16, BHK+16], sparse PCA [MW15a], sparse spiked Wigner and tensor PCA [HKP+17], maximizing random tensors on the sphere [BGL17] and random CSPs [KMOW17]. Lower bounds for relaxations of planted clique and maximum independent set in the Lovász-Schrijver hierarchy are shown in [FK03] and lower bounds for Sherali-Adams relaxations of planted dense subgraph in the log-density regime are shown in [CM18]. Tight lower bounds have been shown in the statistical query model for planted clique [FGR+13], random CSPs [FPV15] and robust sparse mean estimation [DKS16]. It also has been recently shown that planted clique with $k \ll \sqrt{n}$ is hard for regular resolution [ABDR+18]. In [HS17b], a meta-algorithm for Bayesian estimation based on low-degree polynomials, SDPs and tensor decompositions is introduced and shown to achieve the best known upper bound for the k-block stochastic block model, with a matching lower bound for the meta-algorithm.

1.4 Notation

In this paper, we adopt the following notational conventions. Let $\mathcal{L}(X)$ denote the distribution law of a random variable X. Given a distribution $\mathbb{P}$, let $\mathbb{P}^{\otimes n}$ denote the distribution of $(X_1, X_2, \ldots, X_n)$ where the X_i are i.i.d. according to $\mathbb{P}$. Similarly, let $\mathbb{P}^{\otimes m \times n}$ denote the distribution on $\mathbb{R}^{m \times n}$ with i.i.d. entries distributed as $\mathbb{P}$. Given a finite or measurable set $\mathcal{X}$, let $\text{Unif}[\mathcal{X}]$ denote the uniform distribution on $\mathcal{X}$. Let d_{TV}, d_{KL} and χ^2 denote total variation distance, Kullback-Leibler divergence and χ^2 divergence, respectively. Given a measurable set $\mathcal{X}$, let $\Delta(\mathcal{X})$ denote the set of all distributions π on $\mathcal{X}$. If $\mathcal{X}$ is itself a set of distributions, we refer to $\Delta(\mathcal{X})$ as the set of priors on $\mathcal{X}$. Throughout the paper, C refers to any constant independent of the parameters of the problem at hand and will be reused for different constants.

Let $N(\mu, \sigma^2)$ denote a normal random variable with mean μ and variance σ^2 when $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{\geq 0}$. Let $N(\mu, \Sigma)$ denote a multivariate normal random vector with mean $\mu \in \mathbb{R}^d$ and covariance matrix Σ, where Σ is a $d \times d$ positive semidefinite matrix. Let $\beta(x, y)$ denote a beta distribution with parameters $x, y > 0$ and let $\chi^2(k)$ denote a χ^2-distribution with k degrees of freedom. Let $\mathcal{B}_0(k)$ denote the set of all unit vectors $v \in \mathbb{R}^d$ with $\|v\|_0 \leq k$. Let $[n] = \{1, 2, \ldots, n\}$ and $\binom{[n]}{k}$ denote the set of all size k subsets of $[n]$. Let $\mathcal{G}_n$ denote the set of all simple graphs on vertex set $[n]$. Let the Orthogonal group on $\mathbb{R}^{d \times d}$ be $\mathcal{O}_d$. Let $\mathbf{1}_S$ denote the vector $v \in \mathbb{R}^n$ with $v_i = 1$ if $i \in S$ and $v_i = 0$ if $i \notin S$ where $S \subseteq [n]$. For subsets $S \subseteq \mathbb{R}$, let $\mathbf{1}_S$ denote the indicator function of the set S. Let Φ denote the cumulative distribution of a standard normal random variable with $\Phi(x) = \int_{-\infty}^{x} e^{-t^2/2} dt$. Given a simple undirected graph G, let $V(G)$ and $E(G)$ denote its vertex and edge sets, respectively. The notation $a(n) \gg b(n)$ will denote a growing polynomially faster in n than b. In other words, $a \gg b$ if $\liminf_{n \to \infty} \log_n a(n) > \limsup_{n \to \infty} \log_n b(n)$. The notation $a = \tilde{\Theta}(b)$ denotes the equality $\lim_{n \to \infty} \log_n a(n) = \lim_{n \to \infty} \log_n b(n)$. Here, $a \lesssim b$ denotes $a(n) = O(b(n) \cdot \text{poly}(\log n))$ or in other words $a \leq b$ up to polylogarithmic factors in n.

2   Summary of Results

2.1   Detection and Recovery Problems

We consider problems $\mathcal{P}$ with planted sparse structure as both detection and recovery tasks, which we denote by $\mathcal{P}_D$ and $\mathcal{P}_R$, respectively.

Detection.   In detection problems $\mathcal{P}_D$, the algorithm is given a set of observations and tasked with distinguishing between two hypotheses:

- a *uniform* hypothesis H_0, under which observations are generated from the natural noise distribution for the problem; and

- a *planted* hypothesis H_1, under which observations are generated from the same noise distribution but with a latent planted sparse structure.

In all of the detection problems we consider, H_0 is a simple hypothesis consisting of a single distribution and H_1 is either also simple or a composite hypothesis consisting of several distributions. When H_1 is a composite hypothesis, it consists of a set of distributions of the form P_θ where θ is the latent sparse structure of interest. Often H_1 is a simple hypothesis consisting of a single distribution which is a mixture of P_θ with θ in some sense chosen uniformly at random. In both cases, we will abuse notation and refer to H_1 as a set of distributions. Given an observation X, an algorithm $A(X) \in \{0, 1\}$ *solves* the detection problem *with nontrivial probability* if there is an $\epsilon > 0$ such that its Type I+II error satisfies that

$$\limsup_{n \to \infty} \left(\mathbb{P}_{H_0}[A(X) = 1] + \sup_{P \in H_1} \mathbb{P}_{X \sim P}[A(X) = 0] \right) \leq 1 - \epsilon$$

where n is the parameter indicating the size of X. We refer to this quantity as the asymptotic Type I+II error of A for the problem $\mathcal{P}_D$. If the asymptotic Type I+II error of A is zero, then we say A *solves* the detection problem $\mathcal{P}_D$. Our reductions under total variation all yield exact correspondences between asymptotic Type I+II errors. Specifically, they show that if a polynomial time algorithm has asymptotic Type I+II error of ϵ on the problem of interest then there is a polynomial time algorithm with asymptotic Type I+II error ϵ on the problem being reduced from.

Recovery.   In recovery problems $\mathcal{P}_R$, the algorithm is given an observation from P_θ for some latent θ from a space Θ and the task is to recover the support $S(\theta)$ of the sparse structure θ. There are several variants of the recovery task. Given a randomized algorithm with output $A(X) \in \{S(\theta) : \theta \in \Theta\}$ and a distribution π on the latent space Θ, the variants of the recovery task are as follows.

- *Partial Recovery*: A solves partial recovery if

$$\mathbb{E}_{X \sim \mathbb{E}_\pi P_\theta}[|A(X) \cap S(\theta)|] = \Omega(|S(\theta)|) \quad \text{as } n \to \infty$$

- *Weak Recovery*: A solves weak recovery if

$$\mathbb{E}_{X \sim \mathbb{E}_\pi P_\theta}[|A(X) \Delta S(\theta)|] = o(|S(\theta)|) \quad \text{as } n \to \infty$$

- *Exact Recovery*: A solves exact recovery with nontrivial probability $\epsilon > 0$ if for all $\theta \in \Theta$

$$\liminf_{n \to \infty} \mathbb{P}_{X \sim \mathbb{E}_\pi P_\theta}[A(X) = S(\theta)] \geq \epsilon$$

Here, $\mathbb{E}_\pi P_\theta$ denotes the mixture of P_θ induced by π and Δ denotes the symmetric difference between two sets. Whenever the corresponding detection problem $\mathcal{P}_D$ has a simple hypothesis H_1, π will be the prior on Θ as in H_1, which typically is a uniform prior. When $\mathcal{P}_D$ has a composite hypothesis H_1, then an algorithm A solves each of the three variants of the recovery task if the above conditions are met for *all* distributions π. We remark that this is equivalent to the above conditions being met only for distributions π with all of their mass on a single $\theta \in \Theta$. Given a problem $\mathcal{P}$, the notation $\mathcal{P}_R$ will denote the exact recovery problem, and $\mathcal{P}_{PR}$ and $\mathcal{P}_{WR}$ will denote partial and weak recovery, respectively. All of our recovery reductions will apply to all recovery variants simultaneously. In other words, a partial, weak or exact recovery algorithm for the problem of interest implies the same for the problem being reduced from. The computational and statistical barriers for partial, weak and exact recovery generally differ by sub-polynomial factors for the problems we consider. Therefore the polynomial order of the barriers for $\mathcal{P}_R$ will in general also be those for $\mathcal{P}_{PR}$ and $\mathcal{P}_{WR}$. This is discussed in more detail in Section 9.

An instance of a detection problem $\mathcal{P}_D$ hereby refers to an observation X. If $\mathcal{P}_D$ is a simple vs. simple hypothesis testing problem, then the instance X takes one of two distributions – its distribution under H_0 and H_1, which we respectively denote by $\mathcal{L}_{H_0}(X)$ and $\mathcal{L}_{H_1}(X)$. If H_1 is composite, then the distribution of X under P is denoted as $\mathcal{L}_P(X)$ for each $P \in H_1$. An instance of a recovery problem $\mathcal{P}_R$ refers to an observation from some P_θ for some latent $\theta \in \Theta$ if the corresponding detection problem has a composite H_1 or from $\mathcal{L}_{H_1}(X) = \mathbb{E}_\pi P_\theta$ if the corresponding detection problem has a simple H_1.

Computational Model. The algorithms we consider here are either unconstrained or run in randomized polynomial time. An unconstrained algorithm refers to any randomized function or Markov transition kernel from one space to another. These algorithms considered in order to show that information-theoretic lower bounds are asymptotically tight. An algorithm that runs in randomized polynomial time has access to $\text{poly}(n)$ independent random bits and must run in $\text{poly}(n)$ time where n is the size of the input. For clarity of exposition, we assume that explicit expressions can be exactly computed and assume that $N(0,1)$ and Poisson random variables can be sampled in $O(1)$ operations.

2.2   Problem Formulations

In this section, we define the problems that we show computational lower bounds for and the conjectures on which these lower bounds are based. Each problem we consider has a natural parameter n, which typically denotes the number of samples or dimension of the data, and sparsity parameter k. Every parameter for each problem is implicitly a function of n, that grows or decays polynomially in n. For example, $k = k(n) = \tilde{\Theta}(n^\beta)$ for some constant $\beta \in (0,1)$ throughout the paper. For simplicity of notation, we do not write this dependence on n. We mostly will be concerned with the polynomial order of growth of each of the parameters and not with subpolynomial factors. We now formally define the problems we consider.

Planted Clique and Independent Set. The hypotheses in the planted clique detection problem $\mathrm{PC}_D(n, k, p)$ are

$$H_0 : G \sim G(n, p) \quad \text{and} \quad H_1 : G \sim G(n, k, p)$$

where $G(n, p)$ is an Erdős-Rényi random graph with edge probability p and $G(n, k, p)$ is a sample of $G(n, p)$ with a clique of size k planted uniformly at random. As mentioned in the Prior Work section, all known polynomial time algorithms for planted clique fail if $k \ll \sqrt{n}$. This has led to the following hardness conjecture.

Conjecture 2.1 (PC Conjecture). *Fix some constant $p \in (0,1)$. Suppose that $\{A_n\}$ is a sequence of randomized polynomial time algorithms $A_n : \mathcal{G}_n \to \{0,1\}$ and k_n is a sequence of positive integers satisfying that $\limsup_{n\to\infty} \log_n k_n < \frac{1}{2}$. Then if G is an instance of $\mathrm{PC}_D(n,k,p)$, it holds that*

$$\liminf_{n\to\infty} \left(\mathbb{P}_{H_0}\left[A_n(G) = 1\right] + \mathbb{P}_{H_1}\left[A_n(G) = 0\right] \right) \geq 1.$$

The hardness assumption we use throughout our results is the planted clique conjecture. Other than to show hardness for planted dense subgraph in the sparsest regime, we will only need the planted clique conjecture with edge density $p = 1/2$. An interesting open problem posed in [HWX15] is to show that the PC conjecture at $p = 1/2$ implies it for any fixed constant $p < 1/2$.

The hypotheses in the planted independent set detection problem $\mathrm{PIS}_D(n,k,p)$ are

$$H_0 : G \sim G(n,p) \quad \text{and} \quad H_1 : G \sim G_I(n,k,p)$$

where $G_I(n,k,p)$ is a sample of $G(n,p)$ where all of the edges of vertex set of size k removed uniformly at random. The recovery $\mathrm{PC}_R(n,k,p)$ and $\mathrm{PIS}_R(n,k,p)$ problems are to estimate the latent clique and independent set supports given samples from $G(n,k,p)$ and $G_I(n,k,p)$, respectively.

Planted Dense Subgraph. The hypotheses in the planted dense subgraph detection problem $\mathrm{PDS}_D(n,k,p,q)$ are

$$H_0 : G \sim G(n,q) \quad \text{and} \quad H_1 : G \sim G(n,k,p,q)$$

where $G(n,k,p,q)$ is the distribution on $\mathcal{G}_n$ formed by selecting a size k subset S of $[n]$ uniformly at random and joining every two nodes in S with probability p and every other two nodes with probability q. The recovery problem $\mathrm{PDS}_R(n,k,p,q)$ is to estimate the latent planted dense subgraph support S given samples from $G(n,k,p,q)$.

A phenomenon observed in [HWX15] and in [CX16] is that planted dense subgraph appears to have a detection-recovery gap in the regime where $k \gg \sqrt{n}$. The following is a formulation of the conjectured sharper recovery lower bound.

Conjecture 2.2 (PDS Recovery Conjecture). *Suppose that $G \sim G(n,k,p,q)$ and*

$$\liminf_{n\to\infty} \log_n k > \frac{1}{2} \quad \text{and} \quad \limsup_{n\to\infty} \log_n \left(\frac{k^2(p-q)^2}{q(1-q)} \right) < 1$$

then there is no sequence of randomized polynomial-time algorithms $A_n : \mathcal{G}_n \to \binom{[n]}{k}$ such that $A_n(G)$ achieve exact recovery of the vertices in the latent planted dense subgraph as $n \to \infty$.

This conjecture asserts that the threshold of n/k^2 on the signal $\frac{(p-q)^2}{q(1-q)}$ of an instance of PDS is tight for the recovery problem. In contrast, our results show that the tight detection threshold for PDS given the PC conjecture is lower, at n^2/k^4. We will use this conjecture to establish similar detection-recovery gaps for biased sparse PCA and biclustering. We note that a related detection-recovery gap for BC was shown in [CLR15]. The same lower bound was established for strong recovery algorithms that solve biclustering for all sub-Gaussian noise distributions assuming hardness of planted clique for a different distribution on random graphs than Erdős-Rényi.

Subgraph Stochastic Block Model. We introduce a planted subgraph variant of the two community stochastic block model. Detection in this model cannot be solved with the edge-thresholding test that produced the conjectural detection-recovery gap in planted dense subgraph.

Let $G_B(n, k, q, \rho)$ denote the set of distributions on $\mathcal{G}_n$ generated a graph G as follows. Fix any two positive integers k_1 and k_2 satisfying that

$$\frac{k}{2} - k^{1-\delta} \leq k_1, k_2 \leq \frac{k}{2} + k^{1-\delta}$$

where $\delta = \delta_{\mathrm{SSBM}} > 0$ is a small constant that will remained fixed throughout the paper. Let $S = [k_1]$ and $T = [k_1 + k_2]\backslash[k_1]$. Then generate the edges of G independently as follows:

1. include edges within S or within T with probability at least $q + \rho$;

2. include edges between S and T with probability at most $q - \rho$; and

3. include all other edges with probability q.

Then permute the vertices of G according to a permutation selected uniformly at random. The communities of the graph are defined to be the images of S and T under this permutation. Note that $G_B(n, k, q, \rho)$ defines a set of distributions since k_1 and k_2 are permitted to vary and the edges between S and T are included independently with a probability at least $q + \rho$ for each edge. Thus given the random permutation, G is distributed as an inhomogeneous random graph with independent edges. The subgraph stochastic block model detection problem $\mathrm{SSBM}_D(n, k, q, \rho)$ has hypotheses given by

$$H_0 : G \sim G(n, q) \quad \text{and} \quad H_1 : G \sim \mathbb{P} \quad \text{for some} \quad \mathbb{P} \in G_B(n, k, q, \rho)$$

Biclustering. Let $\mathcal{M}_{n,k} \subseteq \mathbb{R}^{n \times n}$ be the set of sparse matrices supported on a $k \times k$ submatrix with each nonzero entry equal to 1. The biclustering detection problem $\mathrm{BC}_D(n, k, \mu)$ has hypotheses

$$H_0 : M \sim N(0, 1)^{\otimes n \times n} \quad \text{and} \quad H_1 : M \sim \mu \cdot A + N(0, 1)^{\otimes n \times n} \text{ where } A \sim \mathrm{Unif}\,[\mathcal{M}_{n,k}]$$

The recovery problem BC_R is to estimate the latent support matrix A given a sample from H_1.

Rank-1 Submatrix and Sparse Spiked Wigner. In rank-1 submatrix, sparse spiked Wigner and sparse PCA, the planted sparse vectors will have sufficiently large entries for support recovery to be possible. We consider the following set of near-uniform magnitude unit vectors

$$\mathcal{V}_{d,k} = \left\{ v \in \mathbb{S}^{d-1} : k - \frac{k}{\log k} \leq \|v\|_0 \leq k \text{ and } |v_i| \geq \frac{1}{\sqrt{k}} \text{ for } i \in \mathrm{supp}(v) \right\}$$

The function $\log k$ can be replaced by any sub-polynomially growing function but is given explicitly for simplicity. The detection problem $\mathrm{ROS}_D(n, k, \mu)$ has hypotheses

$$H_0 : M \sim N(0, 1)^{\otimes n \times n} \quad \text{and} \quad H_1 : M \sim \mu \cdot rc^\top + N(0, 1)^{\otimes n \times n} \text{ where } r, c \in \mathcal{V}_{n,k}$$

The recovery problem ROS_R is to estimate the latent supports $\mathrm{supp}(r)$ and $\mathrm{supp}(c)$. An $n \times n$ GOE matrix $\mathrm{GOE}(n)$ is a symmetric matrix with i.i.d. $N(0, 1)$ entries below its main diagonal and i.i.d. $N(0, 2)$ entries on its main diagonal. The sparse spiked Wigner detection problem SSW_D has hypotheses

$$H_0 : M \sim \mathrm{GOE}(n) \quad \text{and} \quad H_1 : M \sim \mu \cdot rr^\top + \mathrm{GOE}(n) \text{ where } r \in \mathcal{V}_{n,k}$$

and the recovery problem SSW_R is to estimate the latent supports $\mathrm{supp}(r)$. A simple intermediate variant that will be useful in our reductions is SROS, which is ROS constrained to have a symmetric spike $r = c$. Note that if M is an instance of $\mathrm{SROS}_D(n, k, \mu)$ then it follows that $\frac{1}{\sqrt{2}}(M + M^\top)$ is an instance of $\mathrm{SSW}_D(n, k, \mu/\sqrt{2})$.

Sparse PCA. Detection in the *spiked covariance model* of $\mathrm{SPCA}_D(n, k, d, \theta)$ has hypotheses

$$H_0 : X_1, X_2, \ldots, X_n \sim N(0, I_d)^{\otimes n} \quad \text{and}$$

$$H_1 : X_1, X_2, \ldots, X_n \sim N\left(0, I_d + \theta vv^\top\right)^{\otimes n} \text{ where } v \in \mathcal{V}_{d,k}$$

The recovery task SPCA_R is to estimate $\mathrm{supp}(v)$ given observations $X_1, X_2, \ldots, X_n$ sampled from $N\left(0, I_d + \theta vv^\top\right)^{\otimes n}$ where $v \in \mathcal{V}_{d,k}$. We also consider a simple hypothesis testing variant USPCA_D of sparse PCA with hypotheses

$$H_0 : X_1, X_2, \ldots, X_n \sim N(0, I_d)^{\otimes n} \quad \text{and}$$

$$H_1 : X_1, X_2, \ldots, X_n \sim N\left(0, I_d + \theta vv^\top\right)^{\otimes n} \text{ where } v \sim \mathrm{Unif}[S_k]$$

where $S_k \subseteq \mathbb{R}^d$ consists of all k-sparse unit vectors with nonzero coordinates equal to $\pm 1/\sqrt{k}$.

Biased Sparse PCA. We introduce a variant of the spiked covariance model with an additional promise. In particular, v is restricted to the set $\mathcal{BV}_{d,k}$ of vectors in $\mathcal{V}_{d,k}$ with some overall positive or negative bias. Formally, if $\|v\|_0^+$ denotes the number of positive entries of v then

$$\mathcal{BV}_{d,k} = \left\{v \in \mathcal{V}_{d,k} : \|v\|_0^+ \geq \left(\frac{1}{2} + \delta\right) k \text{ or } \|v\|_0^+ \leq \left(\frac{1}{2} - \delta\right) k\right\}$$

where $\delta = \delta_{\mathrm{BSPCA}} > 0$ is an arbitrary constant that will remain fixed throughout the paper. The detection problem $\mathrm{BSPCA}_D(n, k, d, \theta)$ has hypotheses

$$H_0 : X_1, X_2, \ldots, X_n \sim N(0, I_d)^{\otimes n} \quad \text{and}$$

$$H_1 : X_1, X_2, \ldots, X_n \sim N\left(0, I_d + \theta vv^\top\right)^{\otimes n} \text{ where } v \in \mathcal{BV}_{d,k}$$

The recovery problem BSPCA_R is do estimate $\mathrm{supp}(v)$ given observations $X_1, X_2, \ldots, X_n$ sampled from $N\left(0, I_d + \theta vv^\top\right)^{\otimes n}$ where $v \in \mathcal{BV}_{d,k}$. We also consider a simple hypothesis testing variant UBSPCA_D defined similarly to USPCA_D with $v \sim \mathrm{Unif}[BS_k]$ where BS_k is the set of all k-sparse unit vectors with nonzero coordinates equal to $1/\sqrt{k}$.

2.3 Our Results

In this paper, our aim is to establish tight characterizations of the statistical-computational gaps for the problems formulated in the previous section. Each problem has three regimes for each of its detection and recovery variants. In the easy regime, there is a polynomial-time algorithm for the task. In the hard regime, the PC or PDS recovery conjecture implies that there is no polynomial-time algorithm but there is an inefficient algorithm solving the task. In the impossible regime, the task is information-theoretically impossible. Our results complete the hardness picture for each of these problems other than sparse PCA when $k \ll \sqrt{n}$ and biased sparse PCA. Our results are informally stated below and depicted visually in Figure 3.

Theorem 2.1 (Informal Main Theorem). *Given the PC and PDS recovery conjectures, the easy, hard and impossible regimes of the problems in Section 2.2 are classified in Figure 5 as following one of the configurations of regimes in Figure 4.*

(a) PIS(n, k, q) and PDS(n, k, p, q) with $p = cq = \tilde{\Theta}(n^{-\alpha})$

(b) PDS(n, k, p, q) with $q = \tilde{\Theta}(n^{-\alpha})$ and $p - q = \tilde{\Theta}(n^{-5\alpha/4})$

(c) SSBM(n, k, q, ρ) with $q = \tilde{\Theta}(1)$ and $\rho = \tilde{\Theta}(n^{-\alpha})$

(d) BC(n, k, μ) with $\mu = \tilde{\Theta}(n^{-\alpha})$

(e) ROS(n, k, μ) and SSW(n, k, μ) with $\frac{\mu}{k} = \tilde{\Theta}(n^{-\alpha})$

(f) SPCA(n, k, d, θ) with $d = \Theta(n)$ and $\theta = \tilde{\Theta}(n^{-\alpha})$

(g) BSPCA(n, k, d, θ) with $d = \Theta(n)$ and $\theta = \tilde{\Theta}(n^{-\alpha})$
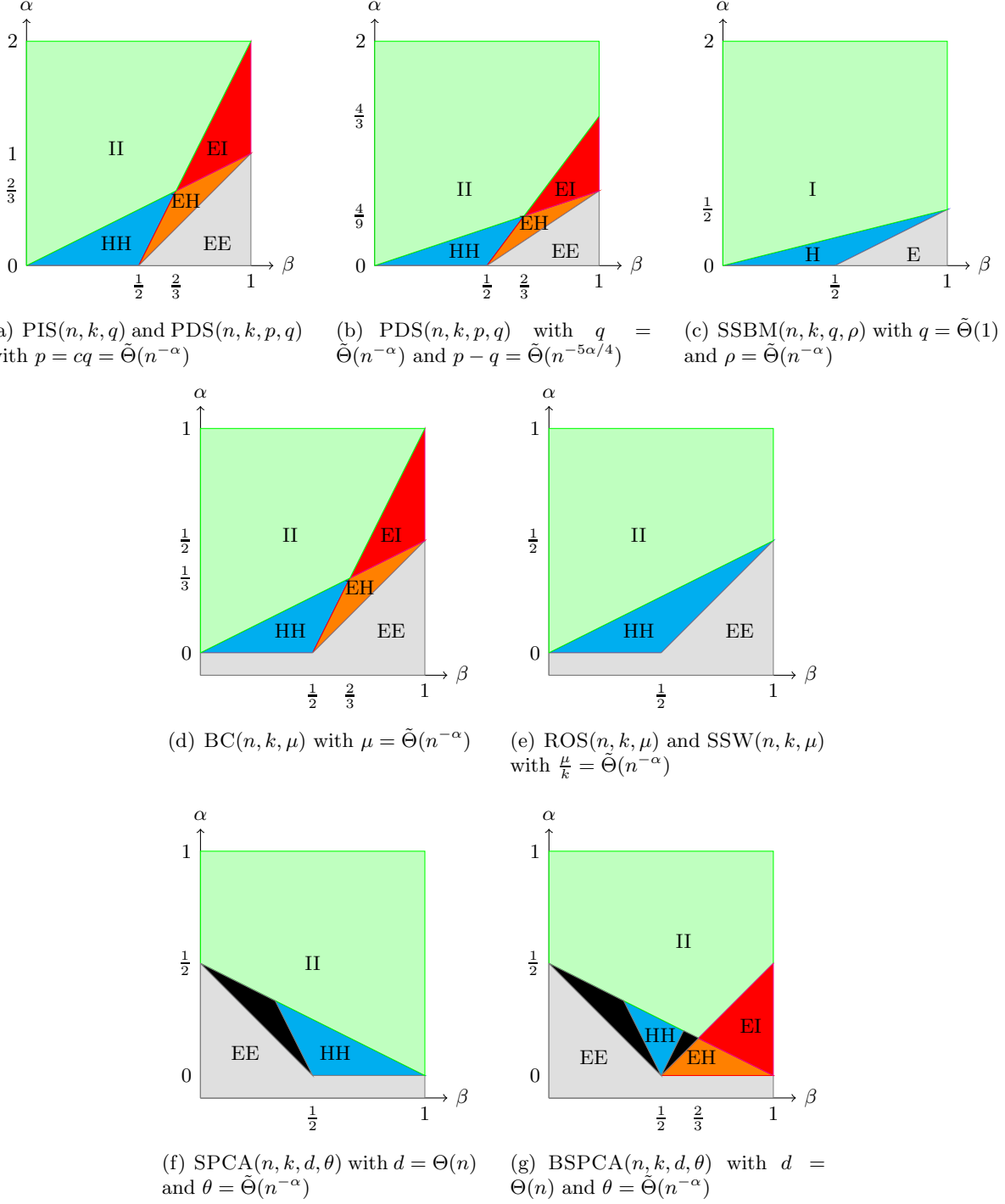
Figure 3: Parameter regimes by problem plotted as signal vs. sparsity. Sparsity is $k = \tilde{\Theta}(n^{\beta})$. First labels characterize detection and second labels characterize exact recovery. Recovery is not considered for SSBM and weak recovery is considered for SPCA and BSPCA. In Easy (E) regimes, there is a polynomial-time algorithm. In Hard (H) regimes, the PC or PDS conjecture implies there is no polynomial-time algorithm. In Impossible (I) regimes, the task is information-theoretically impossible. Hardness in black regions is open.

TYPE I	$k \ll n^{1/2}$	$n^{1/2} \lesssim k \ll n^{2/3}$	$n^{2/3} \ll k$
Impossible	SNR $\ll \frac{1}{k}$	SNR $\ll \frac{1}{k}$	SNR $\ll \frac{n^2}{k^4}$
Hard	SNR $\gtrsim \frac{1}{k}$	$\frac{1}{k} \lesssim$ SNR $\ll \frac{n^2}{k^4}$	None
Easy	(A) None	SNR $\gtrsim \frac{n^2}{k^4}$	SNR $\gtrsim \frac{n^2}{k^4}$
	(B) SNR $\gtrsim 1$		

TYPE II	$k \ll n^{1/2}$	$n^{1/2} \lesssim k \ll n^{2/3}$	$n^{2/3} \ll k$
Impossible	SNR $\ll \sqrt{\frac{k}{n}}$	SNR $\ll \sqrt{\frac{k}{n}}$	SNR $\ll \sqrt{\frac{n}{k^2}}$
Hard	$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll \frac{k^2}{n}$	$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll \frac{n}{k^2}$	None
Easy	SNR $\gtrsim \sqrt{\frac{k^2}{n}}$	SNR $\gtrsim \sqrt{\frac{n}{k^2}}$	SNR $\gtrsim \sqrt{\frac{n}{k^2}}$

TYPE III	$k \ll n^{1/2}$	$n^{1/2} \lesssim k$	TYPE IV	$k \ll n^{1/2}$	$n^{1/2} \lesssim k$
Impossible	SNR $\ll \frac{1}{k}$	SNR $\ll \frac{1}{k}$	Impossible	SNR $\ll \sqrt{\frac{k}{n}}$	SNR $\ll \sqrt{\frac{k}{n}}$
Hard	SNR $\gtrsim \frac{1}{k}$	$\frac{1}{k} \lesssim$ SNR $\ll \frac{n}{k^2}$	Hard	$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll \frac{k^2}{n}$	$\sqrt{\frac{k}{n}} \lesssim$ SNR $\ll 1$
Easy	(A) None	SNR $\gtrsim \frac{n}{k^2}$	Easy	SNR $\gtrsim \sqrt{\frac{k^2}{n}}$	SNR $\gtrsim 1$
	(B) SNR $\gtrsim 1$				

Figure 4: Types of hardness regimes by SNR in Theorem 2.1. TYPE I and TYPE III each have two variants A and B depending on the Easy regime when $k \ll n^{1/2}$.

PROBLEMS	PARAMETER REGIME	SNR	TYPE
$\mathrm{PIS}_D(n,k,q), \mathrm{PDS}_D(n,k,cq,q)$	$q = \tilde{\Theta}(n^{-\alpha})$ for fixed $\alpha \in [0,1)$ and $c > 1$	q	TYPE IA
$\mathrm{PIS}_R(n,k,q), \mathrm{PDS}_R(n,k,cq,q)$	$q = \tilde{\Theta}(n^{-\alpha})$ for fixed $\alpha \in [0,1)$ and $c > 1$	q	TYPE IIIA
$\mathrm{PDS}_D(n,k,p,q)$	$q = \tilde{\Theta}(n^{-\alpha})$ and $p - q = \tilde{\Theta}(n^{-\gamma})$ with $p > q$ for fixed $\alpha, \gamma \in [0,1)$	$\frac{(p-q)^2}{q(1-q)}$	TYPE IA
$\mathrm{PDS}_R(n,k,p,q)$	$q = \tilde{\Theta}(n^{-\alpha})$ and $p - q = \tilde{\Theta}(n^{-\gamma})$ with $p > q$ for fixed $\alpha, \gamma \in [0,1)$	$\frac{(p-q)^2}{q(1-q)}$	TYPE IIIA
$\mathrm{SSBM}_D(n,k,q,\rho)$	$q = \Theta(1)$ and $\rho = \tilde{\Theta}(n^{-\alpha})$ for fixed $\alpha \in [0,1)$	ρ^2	TYPE IIIA
$\mathrm{BC}_D(n,k,\mu)$	$\mu = \tilde{\Theta}(n^{-\alpha})$ for fixed $\alpha \in [0,1)$	μ^2	TYPE IB
$\mathrm{BC}_R(n,k,\mu)$	$\mu = \tilde{\Theta}(n^{-\alpha})$ for fixed $\alpha \in [0,1)$	μ^2	TYPE IIIB
$\mathrm{ROS}_D(n,k,\mu)$, $\mathrm{ROS}_R(n,k,\mu)$, $\mathrm{SSW}_D(n,k,\mu)$, $\mathrm{SSW}_R(n,k,\mu)$	$\mu = \tilde{\Theta}(n^{-\alpha})$ for fixed $\alpha \in [0,1)$	$\frac{\mu^2}{k^2}$	TYPE IIIB
$\mathrm{SPCA}_D(n,k,d,\theta)$, $\mathrm{SPCA}_{WR}(n,k,d,\theta)$, $\mathrm{BSPCA}_{WR}(n,k,d,\theta)$	$d = \Theta(n)$ and $\theta = \tilde{\Theta}(n^{-\alpha})$ for fixed $\alpha \in [0,1)$	θ	TYPE II
$\mathrm{BSPCA}_D(n,k,d,\theta)$	$d = \Theta(n)$ and $\theta = \tilde{\Theta}(n^{-\alpha})$ for fixed $\alpha \in [0,1)$	θ	TYPE IV

Figure 5: Classification of regimes for each problem as in Theorem 2.1. For each problem, k is in the regime $k = \tilde{\Theta}(n^\beta)$ where $\beta \in (0,1)$ is a constant.

We remark that all of the computational lower bounds in Theorem 2.1 follow from the PC conjecture other than those for $\mathrm{PIS}_R, \mathrm{PDS}_R, \mathrm{BC}_R$ and BSPCA_{WR} which follow from the PDS conjecture. The computational lower bounds implicit in the hard regimes in Figures 4 and 5 are the focus of the present work. Section 4 introduces PC-LIFTING to reduce from PC_D to PIS_D. Section 5 introduces rejection kernels and general DISTRIBUTIONAL-LIFTING which are then applied in Section 6 to reduce from PC_D to all regimes of PDS_D. Section 7 introduces reflection cloning to reduce from BC_D to ROS_D, SSW_D and SSBM_D. Section 8 introduces random rotations to reduce from SSW_D to SPCA_D and from BC_D to BSPCA_D. In Section 6, we also give a reduction from PDS_R to BC_R and in Section 9, we reduce from PDS_R to BSPCA_{WR}. In Section 9, we establish the algorithmic upper bounds and information-theoretic lower bounds needed to complete the proof of Theorem 2.1. In Section 10, we show that our detection lower bounds imply recovery lower bounds.

Note that this gives a complete characterization of the easy, hard and impossible regions for all of the problems we consider other than sparse PCA and biased sparse PCA. The SDP relaxation of the MLE for sparse PCA was shown in [BR13b] to succeed if $d = \Theta(n)$ and $k \ll n^{1/2-\delta}$ down to the signal level of $\theta \approx k/\sqrt{n}$, which is generally conjectured to be optimal for efficient algorithms. There is a gap between the lower bounds we prove here, which match those of [BR13b] and [GMZ17], and this conjecturally optimal threshold when $k \ll \sqrt{n}$. There is also a gap for biased sparse PCA detection when $k \gg \sqrt{n}$. These gaps are depicted as the black region in Figures 3(f) and 3(g). Showing tight hardness for sparse PCA for any parameters $k \ll \sqrt{n}$ remains an interesting open problem. Notably, we do not consider the recovery problem for the subgraph stochastic block model and only consider weak, rather than exact, recovery for sparse PCA and biased sparse PCA. While we obtain computational lower bounds that apply to these recovery variants, obtaining matching information-theoretic lower bounds and algorithms remains an open problem. These open problems are discussed in more detail in Section 11.

In Figure 3 and Theorem 2.1, we depict the implications of our hardness results for SPCA and BSPCA in the case where $d = \Theta(n)$ for simplicity. Our hardness results extend mildly beyond this regime, but not tightly. However, we remark that the assumptions $d = \Theta(n)$ or $d = \Omega(n)$ have become commonplace in the literature on lower bounds for sparse PCA. The reductions from PC to sparse PCA in [BR13a], [WBS16] and [GMZ17] construct hard instances that are only tight when $d = \Omega(n)$. The sum of squares lower bounds for sparse PCA in [MW15a] and [HKP+17] also are in the regime $d = \Theta(n)$. The SOS lower bounds in [HKP+17] are actually for the spiked Wigner model rather than the spiked covariance model of sparse PCA that we consider.

2.4 Our Techniques

Distributional Lifting. We introduce several new techniques resembling graph lifts to increase the size k of a sparse structure, while appropriately maintaining the level of signal and independence in the noise distribution. Given a graph G, the main idea behind our techniques is to replace the $\{0, 1\}$-valued edge indicators with Gaussian and Poisson random variables. We then increase the size of an instance by a factor of two iteratively, while maintaining the signal and independence, through distributional tricks such as Poisson splitting and the rotational invariance of independent Gaussians. In [HWX15], the reduction from planted clique to planted dense subgraph also expands an input graph. Rather than proceed iteratively, their method expands the graph in one step. The main technical issue arising from this is that the diagonal entries of the graph's adjacency matrix are mapped to low-density subgraphs in the hidden community. Showing that these are not detectable requires a subtle argument and that the hidden community is randomly sized according to a Binomial distribution. By proceeding incrementally as in our approach, the diagonal entries

become much easier to handle. However for many regimes of interest, incremental approaches that preserve the fact that the instance is a graph seem to unavoidably introduce dependence between edges. Our insight is to map edges to other random variables that can be preserved incrementally while maintaining independence and the desired parameter scaling to produce tight lower bounds. Using Poisson and Gaussian variants of this lifting procedure, we are able to reduce from planted clique in a wide range of parameter regimes. Gaussian lifting also recovers a simple vs. simple hypothesis testing variant of the lower bounds for biclustering shown in [MW15b] as an intermediate step towards reducing to planted dense subgraph.

Rejection Kernels. We give a simple scheme based on rejection sampling that approximately maps a sample from $\text{Bern}(p)$ to a sample from P and from $\text{Bern}(q)$ to Q where $p, q \in [0, 1]$ and P and Q are two distributions on $\mathbb{R}$. By thresholding samples from a pair of distributions P' and Q', and then mapping the resulting Bernoulli random variables to a pair of target distributions P and Q, this method yields an efficient procedure to simultaneously perform two changes of measure. This method is used in distributional lifting to map from edge indicators to a pair of chosen distributions. This framework extends and uses similar ideas to the approximate sampling methods introduced in [HWX15] and [MW15b].

Reflection Cloning. While distributional lifting appears to accurately characterize the hard regimes in many detection problems with an optimal test involving summing the entire input matrix, it is fundamentally lossy. In each iteration, these lifting techniques generate additional randomness in order to maintain independence in the noise distribution of the instance. In a problem like rank-1 submatrix detection that does not admit a sum test, these cloning techniques do not come close to showing hardness at the computational barrier. We introduce a more sophisticated cloning procedure for cases of Gaussian noise that introduces significantly less randomness in each iteration. Let $\mathcal{R}$ denote the linear operator on $n \times n$ matrices that reflects the matrix about its vertical axis of symmetry and let $\mathcal{F}$ denote the linear operator that multiplies each entry on the right half of the matrix by -1. Then one step of reflection cloning replaces a matrix W with

$$W \leftarrow \frac{1}{\sqrt{2}} \left(\mathcal{R}W^\sigma + \mathcal{F}W^\sigma \right)$$

where σ is a random permutation. Reflection cloning then repeats this for rows instead of columns. If $W = uv^\top + G$ has even dimensions and G has i.i.d. $N(0, 1)$ entries, then reflection cloning effectively doubles the sparsity of u and v while mildly decreasing the signal. Importantly, it can be checked that the Gaussian noise matrix retains the fact that it has independent entries. It is interesting to note that this cloning procedure precludes the success of a sum test by negating parts of the signal.

Random Rotations and Sparse PCA. We introduce a simple connection between sparse PCA, biclustering and rank-1 submatrix through random rotations. This yields lower bounds matching those of [GMZ17] and [BR13a]. Although often suboptimal, the random rotations map we introduce tightly gives lower bounds in the regime $k \gg \sqrt{n}$, using reflection cloning as an intermediate. This marks the first tight computational lower bound for sparse PCA over an entire parameter regime. It also illustrates the utility of natural average-case problems as reduction intermediates, suggesting that webs of reductions among problems can be useful beyond worst-case complexity.

3 Average-Case Reductions under Total Variation

The typical approach to show computational lower bounds for detection problems is to reduce an instance of one problem to a random object close in total variation distance to an instance of another problem in randomized polynomial time. More precisely, let $\mathcal{P}$ and $\mathcal{P}'$ be detection problems and X and Y be instances of $\mathcal{P}$ and $\mathcal{P}'$, respectively. Suppose we are given a polynomial-time computable map ϕ taking an object X to $\phi(X)$ with total variation distance to Y decaying to zero simultaneously under each of H_0 and H_1. Then any algorithm that can distinguish H_0 and H_1 for $\mathcal{P}'$ in polynomial time when applied to $\phi(X)$ also distinguishes H_0 and H_1 for $\mathcal{P}$. Taking $\mathcal{P}$ to be PC and $\mathcal{P}'$ to be the problem of interest then yields a computational hardness result for $\mathcal{P}'$ conditional on the PC conjecture. The general idea in this approach is formalized in the following simple lemma.

Lemma 3.1. *Let $\mathcal{P}$ and $\mathcal{P}'$ be detection problems with hypotheses H_0, H_1, H_0', H_1' and let X and Y be instances of $\mathcal{P}$ and $\mathcal{P}'$, respectively. Suppose there is a polynomial time computable map ϕ satisfying*

$$d_{TV}\left(\mathcal{L}_{H_0}(\phi(X)), \mathcal{L}_{H_0'}(Y)\right) + \sup_{\mathbb{P}\in H_1} \inf_{\pi\in\Delta(H_1')} d_{TV}\left(\mathcal{L}_{\mathbb{P}}(\phi(X)), \int_{H_1'} \mathcal{L}_{\mathbb{P}'}(Y)d\pi(\mathbb{P}')\right) \le \delta$$

If there is a polynomial time algorithm solving $\mathcal{P}'$ with Type I+II error at most ϵ, then there is a polynomial time algorithm solving $\mathcal{P}$ with Type I+II error at most $\epsilon + \delta$.

Proof. Let ψ be a polynomial time computable test function solving $\mathcal{P}'$ with Type I+II error at most ϵ. Note that for any observation X of $\mathcal{P}$, the value $\psi \circ \phi(X) \in \{0, 1\}$ can be computed in polynomial time. This is because the fact that $\phi(X)$ can be computed in polynomial time implies that $\phi(X)$ has size polynomial in the size of X. We claim that $\psi \circ \phi$ solves the detection problem $\mathcal{P}'$. Now fix some distribution $\mathbb{P} \in H_1$ and prior $\pi \in \Delta(H_1')$. By the definition of total variation,

$$\left|\mathbb{P}_{H_0}\left[\psi\circ\phi(X)=1\right] - \mathbb{P}_{H_0'}\left[\psi(Y)=1\right]\right| \le d_{\text{TV}}\left(\mathcal{L}_{H_0}(\phi(X)), \mathcal{L}_{H_0'}(Y)\right)$$

$$\left|\mathbb{P}_{X\sim\mathbb{P}}\left[\psi\circ\phi(X)=0\right] - \int_{H_1'}\mathbb{P}_{Y\sim\mathbb{P}'}\left[\psi(Y)=0\right]d\pi(\mathbb{P}')\right| \le d_{\text{TV}}\left(\mathcal{L}_{\mathbb{P}}(\phi(X)), \int_{H_1'}\mathcal{L}_{\mathbb{P}'}(Y)d\pi(\mathbb{P}')\right)$$

Also note that since π is a probability distribution,

$$\int_{H_1'}\mathbb{P}_{Y\sim\mathbb{P}'}\left[\psi(Y)=0\right]d\pi(\mathbb{P}') \le \sup_{\mathbb{P}'\in H_1'}\mathbb{P}_{Y\sim\mathbb{P}'}\left[\psi(Y)=0\right]$$

Combining these inequalities with the triangle inequality yields that

$$\mathbb{P}_{H_0}\left[\psi\circ\phi(X)=1\right] + \mathbb{P}_{X\sim\mathbb{P}}\left[\psi\circ\phi(X)=0\right] \le \epsilon + d_{\text{TV}}\left(\mathcal{L}_{H_0}(\phi(X)), \mathcal{L}_{H_0'}(Y)\right)$$
$$+ d_{\text{TV}}\left(\mathcal{L}_{\mathbb{P}}(\phi(X)), \int_{H_1'}\mathcal{L}_{\mathbb{P}'}(Y)d\pi(\mathbb{P}')\right)$$

Fixing $\mathbb{P}$ and choosing the prior π so that the second total variation above approaches its infimum yields that the right hand side above is upper bounded by $\epsilon + \delta$. The fact that this bound holds for all $\mathbb{P} \in H_1$ proves the lemma. $\square$

We remark that the second term in the total variation condition of Lemma 1 can be interpreted as ensuring that each distribution $\mathbb{P} \in H_1$ is close to a distribution formed by taking a prior π over the distributions in hypothesis H_1'. In light of this lemma, to reduce one problem to another it suffices to find such a map ϕ. In the case that $\mathcal{P}$ and $\mathcal{P}'$ are both simple hypothesis testing problems, the second term is simply $d_{\mathrm{TV}}\left(\mathcal{L}_{H_1}(\phi(X)), \mathcal{L}_{H_1'}(Y)\right)$.

Throughout the analysis of our average-case reductions, total variation will be the key object of interest. We will make use of several standard results concerning total variation, including the triangle inequality, data processing inequality and tensorization of total variation. The latter two results are stated below.

Lemma 3.2 (Data Processing)**.** *Let P and Q be distributions on a measurable space $(\mathcal{X}, \mathcal{B})$ and let $f : \mathcal{X} \to \mathcal{Y}$ be a Markov transition kernel. If $A \sim P$ and $B \sim Q$ then*

$$d_{TV}\left(\mathcal{L}(f(A)), \mathcal{L}(f(B))\right) \leq d_{TV}(P, Q)$$

Lemma 3.3 (Tensorization)**.** *Let $P_1, P_2, \ldots, P_n$ and $Q_1, Q_2, \ldots, Q_n$ be distributions on a measurable space $(\mathcal{X}, \mathcal{B})$. Then*

$$d_{TV}\left(\prod_{i=1}^{n} P_i, \prod_{i=1}^{n} Q_i\right) \leq \sum_{i=1}^{n} d_{TV}(P_i, Q_i)$$

A typical analysis of a multi-step algorithm will proceed as follows. Suppose that $A = A_2 \circ A_1$ is an algorithm with two steps A_1 and A_2. Let P_0 be the input distribution and P_2 be the target distribution that we would like to show is close in variation to $A(P_0)$. Let P_1 be an intermediate distribution which is close in total variation to $A_1(P_0)$. By the triangle inequality,

$$
\begin{aligned}
d_{\mathrm{TV}}\left(A(P_0), P_2\right) &\leq d_{\mathrm{TV}}\left(A(P_0), A_2(P_1)\right) + d_{\mathrm{TV}}\left(A_2(P_1), P_2\right) \\
&= d_{\mathrm{TV}}\left(A_2 \circ A_1(P_0), A_2(P_1)\right) + d_{\mathrm{TV}}\left(A_2(P_1), P_2\right) \\
&\leq d_{\mathrm{TV}}\left(A_1(P_0), P_1\right) + d_{\mathrm{TV}}\left(A_2(P_1), P_2\right)
\end{aligned}
$$

by the data-processing inequality. Thus total variation accumulates over the steps of a multi-step algorithm. This style of analysis will appear frequently in our reductions. Another lemma about total variation that will be useful throughout this work is as follows. The proof is given in Appendix A.

Lemma 3.4. *For any random variable Y and event A in the σ-algebra $\sigma\{Y\}$, it holds that*

$$d_{TV}\left(\mathcal{L}(Y|A), \mathcal{L}(Y)\right) = \mathbb{P}[Y \in A^c]$$

4 Densifying Planted Clique and Planted Independent Set

In this section, we give a reduction increasing the ambient edge density in planted clique while increasing the relative size of the clique, which shows tight hardness for the planted independent set problem. This reduction, which we term PC-LIFTING, serves as an introduction to the general distributional lifting procedure in the next section. Distributional lifting will subsequently be specialized to produce Poisson and Gaussian variants of the procedure, which will be used to prove hardness for biclustering and different regimes of planted dense subgraph.

4.1 Detecting Planted Generalized Diagonals

We first prove a technical lemma that will be used in all of our cloning procedures. Given a matrix M, let M^{σ_1,σ_2} denote the matrix formed by permuting rows according to σ_1 and columns according to σ_2. Let id denote the identity permutation.

Lemma 4.1. *Let P and Q be two distributions such that Q dominates P and $\chi^2(P,Q) \leq 1$. Suppose that M is an $n \times n$ matrix with all of its non-diagonal entries i.i.d. sampled from Q and all of its diagonal entries i.i.d. sampled from P. Suppose that σ is a permutation on $[n]$ chosen uniformly at random. Then*

$$d_{TV}\left(\mathcal{L}(M^{id,\sigma}), Q^{\otimes n \times n}\right) \leq \sqrt{\frac{\chi^2(P,Q)}{2}}$$

Proof. Let σ' be a permutation of $[n]$ chosen uniformly at random and independent of σ. By Fubini's theorem we have that

$$\chi^2\left(\mathcal{L}(M^{\mathrm{id},\sigma}), Q^{\otimes n \times n}\right) + 1 = \int \frac{\mathbb{E}_\sigma\left[\mathbb{P}_{M^{\mathrm{id},\sigma}}(X|\sigma)\right]^2}{\mathbb{P}_{Q^{\otimes n \times n}}(X)}dX = \mathbb{E}_{\sigma,\sigma'}\int \frac{\mathbb{P}_{M^{\mathrm{id},\sigma}}(X|\sigma)\mathbb{P}_{M^{\mathrm{id},\sigma'}}(X|\sigma')}{\mathbb{P}_{Q^{\otimes n \times n}}(X)}dX$$

Now note that conditioned on σ, the entries of $M^{\mathrm{id},\sigma}$ are independent with distribution

$$\mathbb{P}_{M^{\mathrm{id},\sigma}}(X|\sigma) = \prod_{i=1}^{n} P\left(X_{i\sigma(i)}\right) \prod_{j \neq \sigma(i)} Q\left(X_{ij}\right)$$

Therefore we have that

$$\int \frac{\mathbb{P}_{M^{\mathrm{id},\sigma}}(X|\sigma)\mathbb{P}_{M^{\mathrm{id},\sigma'}}(X|\sigma')}{\mathbb{P}_{Q^{\otimes n \times n}}(X)}dX = \int \left(\prod_{i:\sigma(i)=\sigma'(i)} \frac{P\left(X_{i\sigma(i)}\right)^2}{Q\left(X_{i\sigma(i)}\right)}\right) \left(\prod_{i:\sigma(i)\neq\sigma'(i)} P\left(X_{i\sigma(i)}\right)\right)$$

$$\times \left(\prod_{i:\sigma(i)\neq\sigma'(i)} P\left(X_{i\sigma'(i)}\right)\right) \left(\prod_{(i,j):j\neq\sigma(i),j\neq\sigma'(i)} Q\left(X_{ij}\right)\right) dX$$

$$= \prod_{i:\sigma(i)=\sigma'(i)} \left(\int \frac{P\left(X_{i\sigma(i)}\right)^2}{Q\left(X_{i\sigma(i)}\right)}dX_{i\sigma(i)}\right)$$

$$= \left(1 + \chi^2(P,Q)\right)^{|\{i:\sigma(i)=\sigma'(i)\}|}$$

If $\tau = \sigma' \circ \sigma^{-1}$, then τ is a uniformly at random chosen permutation and $Y = |\{i : \sigma(i) = \sigma'(i)\}|$ is the number of fixed points of τ. As in [Pit97], the ith moment of Y is the ith Bell number for $i \leq n$ and for $i > n$, the ith moment of Y is at most the ith Bell number. Since a Poisson distribution with rate 1 has its ith moment given by the ith Bell number for all i, it follows that for each $t \geq 0$ the MGF $\mathbb{E}[e^{tY}]$ is at most that of a Poisson with rate 1, which is $\exp(e^t - 1)$. Setting $t = \log(1 + \chi^2(P,Q)) > 0$ yields that

$$\chi^2\left(\mathcal{L}(M^{\mathrm{id},\sigma}), Q^{\otimes n \times n}\right) = \mathbb{E}\left[(1 + \chi^2(P,Q))^Y\right] - 1 \leq \exp\left(\chi^2(P,Q)\right) - 1 \leq 2 \cdot \chi^2(P,Q)$$

since $e^x \leq 1 + 2x$ for $x \in [0,1]$. Now by Cauchy-Schwarz we have that

$$d_{\mathrm{TV}}\left(\mathcal{L}(M^{\mathrm{id},\sigma}), Q^{\otimes n \times n}\right) \leq \frac{1}{2}\sqrt{\chi^2\left(\mathcal{L}(M^{\mathrm{id},\sigma}), Q^{\otimes n \times n}\right)} \leq \sqrt{\frac{\chi^2(P,Q)}{2}}$$

which completes the proof of the lemma. $\square$

4.2 Planted Clique Lifting

In this section, we analyze the reduction PC-LIFTING, which is given in Figure 6. This reduction will be shown to approximately take an instance of $\mathrm{PC}(n, n^{1/2-\epsilon}, 1/2)$ to $\mathrm{PC}(N, K, 1-q)$ where $N = \tilde{\Theta}(n^{1+\alpha/2})$, $K = \tilde{\Theta}(n^{1/2+\alpha/2-\epsilon})$ and $q = \tilde{\Theta}(n^{-\alpha})$. By taking the complement of the resulting graph, this shows planted clique lower bounds for $\mathrm{PIS}_D(N, K, q)$ up to the boundary $\frac{N^2}{K^4} \gg q$, exactly matching the computational boundary stated in Theorem 2.1. The reduction PC-LIFTING proceeds iteratively, with $\mathrm{PC}(n, k, p)$ approximately mapped at each step to $\mathrm{PC}(2n, 2k, p^{1/4})$.

Given a labelled graph G on n vertices and a permutation σ on $[n]$, let G^σ denote the labelled graph formed by permuting the vertex labels of G according to σ. Given disjoint subsets $S, T \subseteq [n]$, let $G[S]$ denote the induced subgraph on the set S and $G[S \times T]$ denote the induced bipartite subgraph between S and T. Also let $B(m, n, p)$ denote the random bipartite graph with parts of sizes m and n, respectively, where each edge is included independently with probability p. Let $G(n, p, S)$, where S is a k-subset of $[n]$, denote an instance of $G(n, k, p)$ in which the planted clique is conditioned to be on S.

Lemma 4.2 (Planted Clique Lifting)**.** *Suppose that n and ℓ are such that $\ell = O(\log n)$ and are sufficiently large. Let $w(n) > 2$ be an increasing function with $w(n) \to \infty$ as $n \to \infty$. Then $\phi = $ PC-LIFTING is a randomized polynomial time computable map $\phi : \mathcal{G}_n \to \mathcal{G}_{2^\ell n}$ such that under both H_0 and H_1, it holds that*

$$d_{TV}\left(\phi\left(\mathrm{PC}(n, k, 1/2)\right), \mathrm{PC}\left(2^\ell n, 2^\ell k, \left(1 - w(n)^{-1}\right)^{\frac{1}{4^\ell}}\right)\right) \leq \frac{2}{\sqrt{w(n)}}$$

Proof. If $\ell = O(\log n)$, this algorithm runs in randomized polynomial time. Let φ denote the function that applies a single iteration of Step 3 to an input graph H. Let ϕ_ℓ be the algorithm that outputs the value of H in ϕ after ℓ iterations of Step 3. Note that ϕ_0 outputs H after just applying Steps 1 and 2, and $\phi_{\ell+1} = \varphi \circ \phi_\ell$ for all $\ell \geq 0$.

We first consider a single iteration of Step 3 applied to $G \sim G(n, p, S)$, where $G(n, p, S)$ is the distribution of Erdős-Rényi graphs with a planted clique on a fixed vertex set $S \subseteq [n]$ of size $|S| = k$ and $p \geq 1/2$. For each pair of distinct $\{i, j\} \notin \binom{S}{2}$, it holds that $\mathbf{1}_{\{i,j\} \in E(G)} \sim \mathrm{Bern}(p)$ and hence by the probability in Step 3a, that $x^{ij} \sim \mathrm{Bern}(p^{1/4})^{\otimes 4}$. Therefore the graph H' constructed in Step 3b satisfies that:

- $S' = S \cup \{2n + 1 - i : i \in S\}$ forms a clique of size $2k$;

- $\{2n + 1 - i, i\} \in E(H')$ for each $i \in [n]$; and

- each other edge is in $E(H')$ independently with probability $p^{1/4}$.

Now consider the graph $\varphi(G) = H = (H')^\sigma$ conditioned on the set $\sigma(S')$. We will show that this graph is close in total variation to $G(2n, p^{1/4}, \sigma(S'))$. Let $T_1 = [n] \backslash S$ and $T_2 = [2n] \backslash \{2n+1-i : i \in S\}$. Note that every pair of vertices of the form $\{2n+1-i, i\}$ in H' are either both in S' or between T_1 and T_2. This implies that every pair of distinct vertices not in $\sigma(S')^2$ or $\sigma(T_1) \times \sigma(T_2)$ is in $E(H)$ independent with probability $p^{1/4}$, exactly matching the corresponding edges in $G(2n, p^{1/4}, \sigma(S'))$. Coupling these corresponding edges yields only the edges between $\sigma(T_1)$ and $\sigma(T_2)$ uncoupled. Therefore we have that

$$d_{\mathrm{TV}}\left(\mathcal{L}(H|\sigma(S')), G\left(2n, p^{1/4}, \sigma(S')\right)\right) = d_{\mathrm{TV}}\left(\mathcal{L}\left(H[\sigma(T_1) \times \sigma(T_2)]\right), B\left(n-k, n-k, p^{1/4}\right)\right)$$

Now let the $(n-k) \times (n-k)$ matrix M have 1's on its main diagonal and each other entry sampled i.i.d. from $\mathrm{Bern}(p^{1/4})$. If τ is a random permutation on $[n-k]$, then the adjacency matrix of

<div style="border:1px solid black; padding:10px;">

Algorithm PC-LIFTING

Inputs: Graph $G \in \mathcal{G}_n$, number of iterations ℓ, function w with $w(n) \to \infty$

1. For each pair of vertices $\{i,j\} \notin E(G)$, add the edge $\{i,j\}$ to $E(G)$ independently with probability $1 - 2 \cdot w(n)^{-1}$

2. Initialize $H \leftarrow G$, $m \leftarrow n$ and $p \leftarrow 1 - w(n)^{-1}$

3. Repeat for ℓ iterations:

   a. For each pair $\{i,j\}$ of distinct vertices in $[m]$, sample $x^{ij} \in \{0,1\}^4$ such that
      - If $\{i,j\} \in E(H)$, then $x^{ij} = (1,1,1,1)$
      - If $\{i,j\} \notin E(H)$, then $x^{ij} = v$ with probability

      $$\mathbb{P}\left[x^{ij} = v\right] = \frac{p^{|v|_1/4}\left(1 - p^{1/4}\right)^{4-|v|_1}}{1-p}$$

      for each $v \in \{0,1\}^4$ with $v \neq (1,1,1,1)$

   b. Construct the graph H' on the vertex set $[2m]$ such that for distinct $i, j \in [m]$
      - $\{i,j\} \in E(H')$ if $x_1^{ij} = 1$
      - $\{2m+1-i, j\} \in E(H')$ if $x_2^{ij} = 1$
      - $\{i, 2m+1-j\} \in E(H')$ if $x_3^{ij} = 1$
      - $\{2m+1-i, 2m+1-j\} \in E(H')$ if $x_4^{ij} = 1$

      and for each $i \in [m]$, add the edge $\{i, 2m+1-i\} \in E(H')$

   c. Generate a permutation σ on $[2m]$ uniformly at random

   d. Update $H \leftarrow (H')^\sigma$, $p \leftarrow p^{1/4}$ and $m \leftarrow 2m$

4. Output H

</div>

Figure 6: Planted clique lifting procedure in Lemma 4.2.

$H[\sigma(T_1) \times \sigma(T_2)]$ conditioned on $\sigma(S')$ is distributed as $\mathcal{L}\left(M^{\mathrm{id},\tau}\right)$, since T_1 and T_2 are disjoint. Therefore it follows that

$$d_{\mathrm{TV}}\left(\mathcal{L}\left(H[\sigma(T_1) \times \sigma(T_2)]\right), B\left(n-k, n-k, p^{1/4}\right)\right) = d_{\mathrm{TV}}\left(\mathcal{L}\left(M^{\mathrm{id},\tau}\right), \mathrm{Bern}(p^{1/4})^{\otimes(n-k)\times(n-k)}\right)$$
$$\leq \sqrt{\frac{\chi^2(\mathrm{Bern}(1), \mathrm{Bern}(p^{1/4}))}{2}}$$
$$\leq \sqrt{1 - p^{1/4}}$$

by Lemma 4.1 and since $p \geq 1/2$. It follows by the triangle inequality that

$$d_{\mathrm{TV}}\left(\varphi(G(n,p,S)), G(2n, 2k, p^{1/4})\right) \leq \mathbb{E}_{\sigma(S')}\left[d_{\mathrm{TV}}\left(\mathcal{L}(H|\sigma(S')), G\left(2n, p^{1/4}, \sigma(S')\right)\right)\right]$$

Letting S be chosen uniformly at random over all subsets of $[n]$ of size k, applying the triangle inequality again and combining the inequalities above yields that

$$d_{\mathrm{TV}}\left(\varphi(G(n,k,p)), G(2n, 2k, p^{1/4})\right) \leq \mathbb{E}_S\left[d_{\mathrm{TV}}\left(\varphi(G(n,p,S)), G(2n, 2k, p^{1/4})\right)\right] \leq \sqrt{1 - p^{1/4}}$$

A nearly identical but slightly simpler argument shows that

$$d_{\mathrm{TV}}\left(\varphi(G(n,p)), G(2n, p^{1/4})\right) \leq \sqrt{1 - p^{1/4}}$$

For each $\ell \geq 0$, let $p_\ell = \left(1 - w(n)^{-1}\right)^{\frac{1}{4^\ell}}$ be the value of p after ℓ iterations of Step 2. Now note that for each $\ell \geq 0$, we have by triangle inequality and data processing inequality that

$$
\begin{aligned}
d_{\mathrm{TV}}\left(\phi_{\ell+1}\left(G(n,k,1/2)\right), G\left(2^{\ell+1}n, 2^{\ell+1}k, p_{\ell+1}\right)\right) &\leq d_{\mathrm{TV}}\left(\varphi\left(\phi_\ell\left(G(n,k,1/2)\right)\right), \varphi\left(G\left(2^\ell n, 2^\ell k, p_\ell\right)\right)\right) \\
&\quad + d_{\mathrm{TV}}\left(\varphi\left(G\left(2^\ell n, 2^\ell k, p_\ell\right)\right), G\left(2^{\ell+1}n, 2^{\ell+1}k, p_{\ell+1}\right)\right) \\
&\leq d_{\mathrm{TV}}\left(\phi_\ell\left(G(n,k,1/2)\right), G\left(2^\ell n, 2^\ell k, p_\ell\right)\right) \\
&\quad + \sqrt{1 - p_{\ell+1}}
\end{aligned}
$$

and an identical inequality for $\phi_\ell(G(n,1/2))$. Noting that this total variation is zero when $\ell = 0$ and applying these inequalities inductively yields that

$$d_{\mathrm{TV}}\left(\phi_\ell\left(G(n,k,1/2)\right), G\left(2^\ell n, 2^\ell k, p_\ell\right)\right) \leq \sum_{i=1}^{\ell} \sqrt{1 - p_i}$$

and an identical inequality for $\phi_\ell(G(n,1/2))$. Now note that if $x \leq 1/2$ then $(1-x)^{1/4} \geq 1 - x/3$. Iterating this inequality yields that $1 - p_i \leq 3^{-i} w(n)^{-1}$. Therefore

$$\sum_{i=1}^{\ell} \sqrt{1 - p_i} \leq \frac{1}{\sqrt{w(n)}} \sum_{i=1}^{\ell} 3^{-i/2} < \frac{2}{\sqrt{w(n)}}$$

This completes the proof of the lemma. $\qquad\square$

The next theorem formally gives the hardness result guaranteed by the reduction analyzed above together with the PC conjecture. There will be many theorems of this form throughout the paper, which will typically resolve to applying a total variation bound guaranteed in a previous lemma with Lemma 3.1, and analyzing the asymptotic regime of several parameters.

Theorem 4.3. *Let $\alpha \in [0,2)$ and $\beta \in (0,1)$ be such that $\beta < \frac{1}{2} + \frac{\alpha}{4}$. There is a sequence $\{(N_n, K_n, q_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. *The parameters are in the regime $q = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,*

$$\lim_{n \to \infty} \frac{\log q_n^{-1}}{\log N_n} = \alpha \quad and \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. *For any sequence of randomized polynomial-time tests $\phi_n : \mathcal{G}_{N_n} \to \{0,1\}$, the asymptotic Type I+II error of ϕ_n on the problems $\mathrm{PIS}_D(N_n, K_n, q_n)$ is at least 1 assuming the PC conjecture holds with density $p = 1/2$.*

Therefore the computational boundary for $\text{PIS}_D(n, k, q)$ *in the parameter regime* $q = \tilde{\Theta}(n^{-\alpha})$ *and* $k = \tilde{\Theta}(n^\beta)$ *is* $\beta^* = \frac{1}{2} + \frac{\alpha}{4}$.

Proof. If $\beta < \alpha$ then PIS is information-theoretically impossible. Thus we may assume that $\beta \geq \alpha$. Let $\gamma = \frac{2\beta - \alpha}{2 - \alpha}$ and note that $\gamma \in (0, 1/2)$. Now set

$$\ell_n = \left\lceil \frac{\alpha \log_2 n}{2 - \alpha} \right\rceil, \qquad k_n = \lceil n^\gamma \rceil, \qquad N_n = 2^{\ell_n} n \qquad K_n = 2^{\ell_n} k_n, \qquad q_n = 1 - (1 - w(n)^{-1})^{1/4^{\ell_n}}$$

where $w(n)$ is any sub-polynomial increasing function tending to infinity. By Lemma 4.2, there is a randomized polynomial time algorithm mapping $\text{PC}_D(n, k_n, 1/2)$ to $\text{PC}_D(N_n, K_n, 1 - q_n)$ with total variation converging to zero as $n \to \infty$. Now note that flipping every edge to a non-edge and non-edge to an edge maps $\text{PC}_D(N_n, K_n, 1 - q_n)$ to $\text{PIS}_D(N_n, K_n, q_n)$. This map with Lemma 1 now implies that property 2 above holds. We now verify property 1. Note that

$$\lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \lim_{n \to \infty} \frac{\left\lceil \frac{\alpha \log_2 n}{2 - \alpha} \right\rceil \cdot \log 2 + \left(\frac{2\beta - \alpha}{2 - \alpha} \right) \log n}{\left\lceil \frac{\alpha \log_2 n}{2 - \alpha} \right\rceil \cdot \log 2 + \log n} = \frac{\frac{\alpha}{2 - \alpha} + \frac{2\beta - \alpha}{2 - \alpha}}{\frac{\alpha}{2 - \alpha} + 1} = \beta$$

Note that as $n \to \infty$, it follows that since $4^{-\ell_n} \log(1 - w(n)^{-1}) \to 0$,

$$q_n = 1 - (1 - w(n)^{-1})^{1/4^{\ell_n}} = 1 - e^{4^{-\ell_n} \log(1 - w(n)^{-1})} \sim 4^{-\ell_n} \log(1 - w(n)^{-1})$$

Now it follows that

$$\lim_{n \to \infty} \frac{\log q_n^{-1}}{\log N_n} = \lim_{n \to \infty} \frac{2 \left\lceil \frac{\alpha \log_2 n}{2 - \alpha} \right\rceil \log 2 - \log(1 - w(n)^{-1})}{\left\lceil \frac{\alpha \log_2 n}{2 - \alpha} \right\rceil \cdot \log 2 + \log n} = \frac{\frac{2\alpha}{2 - \alpha}}{\frac{\alpha}{2 - \alpha} + 1} = \alpha$$

which completes the proof. $\qquad \square$

5 Rejection Kernels and Distributional Lifting

In this section, we generalize the idea in PC-LIFTING to apply to any distribution with a natural cloning operation, analogous to Step 3a in PC-LIFTING. Before describing this general distributional lifting procedure, we first will establish several results on applying rejection kernels, a general method for changes of measure such as from Bernoulli edge indicators to Gaussians, that we will need throughout our reductions.

5.1 Rejection Kernels

All of our remaining reductions will involve approximately mapping from a pair of Bernoulli random variables, typically edge indicators in random graphs, to a given pair of random variables. Similar entry-wise transformations of measure were used in [MW15b] and [GMZ17] for mapping from Bernoulli random variables to Gaussian random variables. We generalize these maps to arbitrary distributions and give a simple algorithm using rejection sampling to implement them. The general objective is to construct a single randomized function $\text{RK} : \{0, 1\} \to \mathbb{R}$ that simultaneously maps $\text{Bern}(p)$ to the distribution f_X and $\text{Bern}(q)$ to g_X, approximately in total variation distance. For maps from instances G of planted clique, such a map with $p = 1$ and $q = 1/2$ approximately sends the edge indicators $\mathbf{1}_{\{i,j\} \in E(G)}$ to f_X if i and j are in the planted clique and to g_X otherwise.

We first describe the general structure of the maps RK and their precise total variation guarantees in the following lemma. Then we give particular rejection kernels that we will use in our reductions.

Algorithm RK(x)

Parameters: Input $B \in \{0, 1\}$, a pair of PMFs or PDFs f_X and g_X that can be efficiently computed and sampled, Bernoulli probabilities $p, q \in [0, 1]$, number of iterations N

1. Initialize $Y \leftarrow 0$

2. For N iterations do:

    a. If $B = 0$, sample $Z \sim g_X$ and if

    $$p \cdot g_X(Z) \geq q \cdot f_X(Z)$$

    then with probability $1 - \frac{q \cdot f_X(Z)}{p \cdot g_X(Z)}$, update $Y \leftarrow Z$ and break

    b. If $B = 1$, sample $Z \sim f_X$ and if

    $$(1 - q) \cdot f_X(Z) \geq (1 - p) \cdot g_X(Z)$$

    then with probability $1 - \frac{(1-p) \cdot g_X(Z)}{(1-q) \cdot f_X(Z)}$, update $Y \leftarrow Z$ and break

3. Output Y

Figure 7: Rejection kernel in Lemma 5.1

Lemma 5.1. *Let f_X and g_X be probability mass or density functions supported on subsets of $\mathbb{R}$ such that g_X dominates f_X. Let $p, q \in [0, 1]$ be such that $p > q$ and let*

$$S = \left\{ x \in \mathbb{R} : \frac{1 - p}{1 - q} \leq \frac{f_X(x)}{g_X(x)} \leq \frac{p}{q} \right\}$$

Suppose that $f_X(x)$ and $g_X(x)$ can be computed in $O(T_1)$ time and samples from f_X and g_X can be generated in randomized $O(T_2)$ time. Then there is a randomized $O(N(T_1 + T_2))$ time computable map RK $: \{0, 1\} \to \mathbb{R}$ *such that $d_{TV}(\mathrm{RK}(Bern(p)), f_X) \leq \Delta$ and $d_{TV}(\mathrm{RK}(Bern(q)), g_X) \leq \Delta$ where*

$$\Delta = \max \left\{ \frac{2 \cdot \mathbb{P}_{X \sim f_X}[X \notin S]}{p - q} + \left(\mathbb{P}_{X \sim g_X}[X \notin S] + \frac{q}{p} \right)^N, \right.$$

$$\left. \frac{2 \cdot \mathbb{P}_{X \sim g_X}[X \notin S]}{p - q} + \left(\mathbb{P}_{X \sim f_X}[X \notin S] + \frac{1 - p}{1 - q} \right)^N \right\}$$

Proof. Let RK be implemented as shown in Figure 5.1 and note that RK runs in randomized $O(N(T_1 + T_2))$ time. Define S_0 and S_1 by

$$S_0 = \left\{ x \in \mathbb{R} : \frac{f_X(x)}{g_X(x)} \leq \frac{p}{q} \right\} \quad \text{and} \quad S_1 = \left\{ x \in \mathbb{R} : \frac{1 - p}{1 - q} \leq \frac{f_X(x)}{g_X(x)} \right\}$$

Now define the distributions by the densities or mass functions

$$\varphi_0(x) = \frac{p \cdot g_X(x) - q \cdot f_X(x)}{p \cdot \mathbb{P}_{X \sim g_X}[X \in S_0] - q \cdot \mathbb{P}_{X \sim f_X}[X \in S_0]} \quad \text{for } x \in S_0$$

$$\varphi_1(x) = \frac{(1-q) \cdot f_X(x) - (1-p) \cdot g_X(x)}{(1-q) \cdot \mathbb{P}_{X \sim f_X}[X \in S_1] - (1-p) \cdot \mathbb{P}_{X \sim g_X}[X \in S_1]} \quad \text{for } x \in S_1$$

that are both zero elsewhere. Note that these are both well-defined PDFs or PMFs since they are nonnegative and normalized by the definitions of S_0 and S_1. First consider the case when $B = 0$. For the sake of this analysis, consider the iterations of Step 2 beyond the first update $Y \leftarrow Z$. Now let A_i be the event that the update $Y \leftarrow Z$ occurs in the ith iteration of Step 2a and let A_i' be the event that the first update occurs in the ith iteration. Note that $A_i' = A_1^C \cap A_2^C \cap \cdots \cap A_{i-1}^C \cap A_i$. The probability of A_i is

$$\mathbb{P}[A_i] = \int_{S_0} g_X(x) \left(1 - \frac{q \cdot f_X(x)}{p \cdot g_X(x)}\right) dx = \mathbb{P}_{X \sim g_X}[X \in S_0] - \frac{q}{p} \cdot \mathbb{P}_{X \sim f_X}[X \in S_0]$$

Now note that since the sample Z in the ith iteration and A_i are independent of $A_1, A_2, \ldots, A_{i-1}$, it holds that $f_{Y|A_i'} = f_{Y|A_i}$. The density of Y given the event A_i' is therefore given by

$$f_{Y|A_i'}(x) = f_{Y|A_i}(x) = \mathbb{P}[A_i]^{-1} \cdot g_X(x) \cdot \left(1 - \frac{q \cdot f_X(x)}{p \cdot g_X(x)}\right) = \varphi_0(x)$$

for each $x \in S_0$. If $A = A_1' \cup A_2' \cup \cdots \cup A_N' = A_1 \cup A_2 \cup \cdots \cup A_N$ is the event that the update $Y \leftarrow Z$ occurs in an iteration of Step 2a, then it follows by independence that

$$\mathbb{P}\left[A^C\right] = \prod_{i=1}^{N} (1 - \mathbb{P}[A_i]) = \left(1 - \mathbb{P}_{X \sim g_X}[X \in S_0] + \frac{q}{p} \cdot \mathbb{P}_{X \sim f_X}[X \in S_0]\right)^N$$

$$\leq \left(\mathbb{P}_{X \sim g_X}[X \notin S] + \frac{q}{p}\right)^N$$

since $S \subseteq S_0$. Note that $f_{Y|A}(x) = \varphi_0(x)$ and $\text{RK}(0)$ is Y if $B = 0$. Therefore it follows by Lemma 3.4 that

$$d_{\text{TV}}(\text{RK}(0), \varphi_0) = \mathbb{P}\left[A^C\right] \leq \left(\mathbb{P}_{X \sim g_X}[X \notin S] + \frac{q}{p}\right)^N$$

A symmetric argument shows that when $B = 1$,

$$d_{\text{TV}}(\text{RK}(1), \varphi_1) \leq \left(\mathbb{P}_{X \sim f_X}[X \notin S] + \frac{1-p}{1-q}\right)^N$$

Now note that

$$\left\| \varphi_0 - \frac{p \cdot g_X - q \cdot f_X}{p - q} \right\|_1 = \int_{S_0} \left| \frac{p \cdot g_X(x) - q \cdot f_X(x)}{p \cdot \mathbb{P}_{X \sim g_X}[X \in S_0] - q \cdot \mathbb{P}_{X \sim f_X}[X \in S_0]} - \frac{p \cdot g_X(x) - q \cdot f_X(x)}{p - q} \right| dx$$

$$+ \int_{S_0^C} \frac{q \cdot f_X(x) - p \cdot g_X(x)}{p - q} dx$$

$$= \left| 1 - \frac{p \cdot \mathbb{P}_{X \sim g_X}[X \in S_0] - q \cdot \mathbb{P}_{X \sim f_X}[X \in S_0]}{p - q} \right|$$

$$+ \frac{q \cdot \mathbb{P}_{X \sim f_X}[X \notin S_0] - p \cdot \mathbb{P}_{X \sim g_X}[X \notin S_0]}{p - q}$$

$$= \frac{2(q \cdot \mathbb{P}_{X \sim f_X}[X \notin S_0] - p \cdot \mathbb{P}_{X \sim g_X}[X \notin S_0])}{p - q}$$

$$\leq \frac{2 \cdot \mathbb{P}_{X \sim f_X}[X \notin S]}{p - q}$$

since $S_0 \subseteq S$. A similar computation shows that

$$\left\| \varphi_1 - \frac{(1 - q) \cdot f_X - (1 - p) \cdot g_X}{p - q} \right\|_1 = \frac{2((1 - p) \cdot \mathbb{P}_{X \sim g_X}[X \notin S_1] - (1 - q) \cdot \mathbb{P}_{X \sim f_X}[X \notin S_1])}{p - q}$$

$$\leq \frac{2 \cdot \mathbb{P}_{X \sim g_X}[X \notin S]}{p - q}$$

Now note that

$$f_X = p \cdot \frac{(1 - q) \cdot f_X - (1 - p) \cdot g_X}{p - q} + (1 - p) \cdot \frac{p \cdot g_X - q \cdot f_X}{p - q}$$

Therefore by the triangle inequality, we have that

$$d_{\mathrm{TV}}\left(\mathrm{RK}(\mathrm{Bern}(p)), f_X\right) \leq d_{\mathrm{TV}}\left(\mathrm{RK}(\mathrm{Bern}(p)), p \cdot \varphi_1 + (1 - p) \cdot \varphi_0\right) + d_{\mathrm{TV}}\left(p \cdot \varphi_1 + (1 - p) \cdot \varphi_0, f_X\right)$$

$$\leq p \cdot d_{\mathrm{TV}}\left(\mathrm{RK}(1), \varphi_1\right) + p \cdot \left\| \varphi_1 - \frac{(1 - q) \cdot f_X - (1 - p) \cdot g_X}{p - q} \right\|_1$$

$$+ (1 - p) \cdot d_{\mathrm{TV}}\left(\mathrm{RK}(0), \varphi_0\right) + (1 - p) \cdot \left\| \varphi_0 - \frac{p \cdot g_X - q \cdot f_X}{p - q} \right\|_1$$

$$\leq p \cdot \left(\frac{2 \cdot \mathbb{P}_{X \sim g_X}[X \notin S]}{p - q} + \left(\mathbb{P}_{X \sim f_X}[X \notin S] + \frac{1 - p}{1 - q} \right)^N \right)$$

$$+ (1 - p) \cdot \left(\frac{2 \cdot \mathbb{P}_{X \sim f_X}[X \notin S]}{p - q} + \left(\mathbb{P}_{X \sim g_X}[X \notin S] + \frac{q}{p} \right)^N \right)$$

$$\leq \Delta$$

Similarly, note that

$$g_X = q \cdot \frac{(1 - q) \cdot f_X - (1 - p) \cdot g_X}{p - q} + (1 - q) \cdot \frac{p \cdot g_X - q \cdot f_X}{p - q}$$

The same triangle inequality applications as above show that

$$d_{\mathrm{TV}}\left(\mathrm{RK}(\mathrm{Bern}(q)), g_X\right) \le q \cdot \left(\frac{2 \cdot \mathbb{P}_{X \sim f_X}[X \notin S]}{p - q} + \left(\mathbb{P}_{X \sim g_X}[X \notin S] + \frac{q}{p}\right)^N\right)$$

$$+ (1 - q) \cdot \left(\frac{2 \cdot \mathbb{P}_{X \sim g_X}[X \notin S]}{p - q} + \left(\mathbb{P}_{X \sim f_X}[X \notin S] + \frac{1 - p}{1 - q}\right)^N\right)$$

$$\le \Delta$$

completing the proof of the lemma. $\square$

We will denote RK as defined with the parameters in the lemma above as $\mathrm{RK}(p \to f_X, q \to g_X, N)$ from this point forward. We now give the particular rejection kernels we will need in our reductions and their total variation guarantees. The proofs of these guarantees are deferred to Appendix B. The first rejection kernel maps from the edge indicators in planted clique to Poisson random variables and will be essential in Poisson lifting.

Lemma 5.2. *Let n be a parameter and let $\epsilon > 0, c > 1$ and $q \in (0, 1)$ be fixed constants satisfying that $3\epsilon^{-1} \le \log_c q^{-1}$. If $\lambda = \lambda(n)$ satisfies that $0 < \lambda \le n^{-\epsilon}$, then the map*

$$\mathrm{RK}_{P1} = \mathrm{RK}(1 \to Pois(c\lambda), q \to Pois(\lambda), N)$$

where $N = \lceil 6 \log_{q^{-1}} n \rceil$ can be computed in $O(\log n)$ time and satisfies that

$$d_{TV}\left(\mathrm{RK}_{P1}(1), Pois(c\lambda)\right) = O_n(n^{-3}) \quad \text{and} \quad d_{TV}\left(\mathrm{RK}_{P1}(\mathrm{Bern}(q)), Pois(\lambda)\right) = O_n(n^{-3})$$

The next lemma gives another approximate map to Poisson random variables from Bernoulli random variables corresponding to the edge indicators in the edge-dense regime of planted dense subgraph. We use the following lemma to apply Poisson lifting after Gaussian lifting in order to deduce hardness in the general regime of PDS. The proof is very similar to that of Lemma 5.2. In typical applications of the lemma, we take $\lambda = n^{-\epsilon}$ and $c - 1 = \Theta(\rho)$ where $\rho \to 0$.

Lemma 5.3. *Let $\epsilon \in (0, 1)$ be a fixed constant and let n be a parameter. Suppose that:*

- *$\lambda = \lambda(n)$ satisfies that $0 < \lambda \le n^{-\epsilon}$;*

- *$c = c(n) > 1$ satisfies that $c = O_n(1)$; and*

- *$\rho = \rho(n) \in (0, 1/2)$ satisfies that $\rho \ge n^{-K}$ for sufficiently large n where $K = \Theta_n(1)$ is positive and*

$$(K + 3)\epsilon^{-1} \le \log_c(1 + 2\rho) = O_n(1)$$

Then the map

$$\mathrm{RK}_{P2} = \mathrm{RK}\left(\frac{1}{2} + \rho \to Pois(c\lambda), \frac{1}{2} \to Pois(\lambda), N\right)$$

where $N = \lceil 6\rho^{-1} \log n \rceil$ can be computed in poly(n) time and satisfies

$$d_{TV}\left(\mathrm{RK}_{P2}(\mathrm{Bern}(1/2 + \rho)), Pois(c\lambda)\right) = O_n(n^{-3}), \quad \text{and}$$

$$d_{TV}\left(\mathrm{RK}_{P2}(\mathrm{Bern}(1/2)), Pois(\lambda)\right) = O_n(n^{-3})$$

The next lemma of this section approximately maps from Bernoulli to Gaussian random variables, yielding an alternative to the Gaussianization step in [MW15b]. Gaussian random variables appear in two different contexts in our reductions: (1) the problems ROS, SPCA and BC have observations sampled from multivariate Gaussians; and (2) random matrices with Gaussian entries are used as intermediate in our reductions to PDS in the general regime and SSBM. In both cases, we will need the map in the following lemma. As in the proofs of the previous two lemmas, this next lemma also verifies the conditions of Lemma 5.1 for Gaussians and derives an upper bound on Δ.

Lemma 5.4. *Let n be a parameter and suppose that $p = p(n)$ and $q = q(n)$ satisfy that $p > q$, $p, q \in [0, 1]$, $\min(q, 1-q) = \Omega_n(1)$ and $p - q \geq n^{-O_n(1)}$. Let $\delta = \min \left\{ \log \left(\frac{p}{q} \right), \log \left(\frac{1-q}{1-p} \right) \right\}$. Suppose that $\mu = \mu(n) \in (0, 1)$ is such that*

$$\mu \leq \frac{\delta}{2\sqrt{6 \log n + 2 \log(p - q)^{-1}}}$$

Then the map

$$\mathrm{RK}_G = \mathrm{RK}\left(p \to N(\mu, 1), q \to N(0, 1), N \right)$$

where $N = \lceil 6\delta^{-1} \log n \rceil$ can be computed in $\mathrm{poly}(n)$ time and satisfies

$$d_{TV}\left(\mathrm{RK}_G(Bern(p)), N(\mu, 1)\right) = O_n(n^{-3}) \quad and \quad d_{TV}\left(\mathrm{RK}_G(Bern(q)), N(0, 1)\right) = O_n(n^{-3})$$

5.2 Distributional Lifting

In this section, we introduce a general distributional lifting procedure to reduce from an instance of planted clique to subgraph problems with larger planted subgraphs. The key inputs to the procedure are two parameterized families of distributions P_λ and Q_λ that have a natural cloning map, as described below.

The general distributional lifting procedure begins with an instance $G \in \mathcal{G}_n$ of a planted dense subgraph problem such as planted clique and applies a rejection kernel element-wise to its adjacency matrix. This yields a symmetric matrix M with zeros on its main diagonal, i.i.d. entries sampled from P_{λ_0} on entries corresponding to clique edges and i.i.d. entries sampled from Q_{λ_0} elsewhere. As an input to the procedure, we assume a random cloning map f_{cl} that exactly satisfies

$$f_{\mathrm{cl}}(P_\lambda) \sim P_{g_{\mathrm{cl}}(\lambda)}^{\otimes 4} \quad \text{and} \quad f_{\mathrm{cl}}(Q_\lambda) \sim Q_{g_{\mathrm{cl}}(\lambda)}^{\otimes 4}$$

for some parameter update function g_{cl}. There is a natural cloning map f_{cl} for Gaussian and Poisson distributions, the two families we apply distributional lifting with. Applying this cloning map entry-wise to M and arranging the resulting entries correctly yields a matrix M of size $2n \times 2n$ with zeros on its diagonal and a planted submatrix of size $2k \times 2k$. The only distributional issue that arises are the anti-diagonal entries, which are now all from $Q_{g_{\mathrm{cl}}(\lambda)}$ although some should be from $P_{g_{\mathrm{cl}}(\lambda)}$. We handle these approximately in total variation by randomly permuting the rows and columns and applying Lemma 4.1. Iterating this procedure ℓ times yields a matrix M' of size $2^\ell n \times 2^\ell n$ with a planted submatrix of size $2^\ell k \times 2^\ell k$. If $\lambda_{i+1} = g_{\mathrm{cl}}(\lambda_i)$, then M' has all i.i.d. entries from Q_{λ_ℓ} under H_0 and a planted submatrix with i.i.d. entries from P_{λ_ℓ} under H_1. We then threshold the entries of M' to produce the adjacency matrix of a graph with i.i.d. edge indicators, conditioned on the vertices in the planted subgraph.

A natural question is: what is the purpose of the families P_λ and Q_λ? If the initial and final distributions are both graph distributions with Bernoulli edge indicators, it a priori seems

unnecessary to use matrix distributions without Bernoulli entries as intermediates. Our main reason for introducing these intermediate distributions is that they achieve the right parameter tradeoffs to match the best known algorithms for PDS, while cloning procedures that stay within the set of graph distributions do not. Consider the target planted dense subgraph instance of $\text{PDS}(n, k, p, q)$ where $p = 2q$ and $q = \tilde{\Theta}(n^{-\alpha})$. To produce lower bounds tight with the computational barrier of $q \approx n^2/k^4$ in Theorem 2.1, a lifting procedure mapping $n \to 2n$ and $k \to 2k$ at each step would need its cloning map to satisfy

$$f_{\text{cl}}(\text{Bern}(q)) \sim Q = \text{Bern}(q/4)^{\otimes 4} \quad \text{and} \quad f_{\text{cl}}(\text{Bern}(p)) \sim P = \text{Bern}(p/4)^{\otimes 4}$$

where $p = 2q$. It is not difficult to verify that for any random map $f_{\text{cl}} : \{0, 1\} \to \{0, 1\}^4$, it would need to hold that

$$\frac{1-p}{1-q} \le \frac{P(x)}{Q(x)} \le \frac{p}{q} \quad \text{for all } x \in \{0, 1\}^4$$

However, $P(1, 1, 1, 1)/Q(1, 1, 1, 1) = 16 > 2 = p/q$, so no such map can exist. Another approach is to relax f_{cl} to be an approximate map like a rejection kernel. However, this seems to induce a large loss in total variation from the target distribution on M'. Our solution is to use a rejection kernel to map to distributions with natural cloning maps f_{cl}, such as a Poisson or Gaussian distribution, to front-load the total variation loss to this approximate mapping step and induce no entry-wise total variation loss later in the cloning procedure. We choose the precise distributions P_λ and Q_λ to match the parameter tradeoff along the computational barrier. Note that this general distributional lifting procedure can also be used to map to problems other than variants of subgraph detection, such as biclustering, by not truncating in Step 4. We remark that the PC-LIFTING reduction presented in the previous section is almost an instance of distributional lifting with $P_\lambda = \text{Bern}(1)$ for all λ and $Q_\lambda = \text{Bern}(\lambda)$, with the parameter update $g_{\text{cl}}(\lambda) = \lambda^{1/4}$. However in PC-LIFTING, the planted anti-diagonal entries between the vertices i and $2m + 1 - i$ in Step 3b are from the planted distribution P_λ, rather than Q_λ as in distributional lifting. This requires a slightly different analysis of the planted anti-diagonal entries with Lemma 4.1.

We now proceed to describe distributional lifting shown in Figure 8 and prove its guarantees. Given two distributions P and Q, let $M_n(Q)$ denote the distribution on $n \times n$ symmetric matrices with zero diagonal entries and every entry below the diagonal sampled independently from Q. Similarly, let $M_n(S, P, Q)$ denote the distribution on random $n \times n$ symmetric matrices formed by:

1. sampling the entries of the principal submatrix with indices in S below its main diagonal independently from P;

2. sampling all other entries below the main diagonal independently from Q; and

3. placing zeros on the diagonal.

Let $M_n(k, P, Q)$ denote the distribution of matrices $M_n(S, P, Q)$ where S is a size k subset of $[n]$ selected uniformly at random. Given a matrix $M \in \mathbb{R}^{n \times n}$ and index sets $S, T \subseteq [n]$, let $M[S \times T]$ denote the $|S| \times |T|$ submatrix of M with row indices in S and column indices in T. Also let $G(n, p, q, S)$ where S is a k-subset of $[n]$ denote an instance of $G(n, k, p, q)$ where the planted dense subgraph is conditioned to be on S. The guarantees of distributional lifting are as follows.

Theorem 5.5 (Distributional Lifting). *Suppose that n and ℓ are such that $\ell = O(\log n)$ and are sufficiently large. Let $p', q' \in [0, 1]$ and define the parameters:*

- *target planted and noise distribution families P_λ and Q_λ parameterized by λ;*

Algorithm DISTRIBUTIONAL-LIFTING

Inputs: Graph $G \in \mathcal{G}_n$, number of iterations ℓ, parameterized families of target planted and noise distributions P_λ and Q_λ, a TV-approximation Q'_λ to Q_λ that can be efficiently sampled, rejection kernel RK approximately mapping $\mathrm{Bern}(p') \to P_{\lambda_0}$ and $\mathrm{Bern}(q') \to Q_{\lambda_0}$, threshold t, cloning map f_{cl} and corresponding parameter map g_{cl}

1. Form the symmetric matrix $M \in \mathbb{R}^{n \times n}$ with $M_{ii} = 0$ and off-diagonal terms

$$M_{ij} = \mathrm{RK}\left(\mathbf{1}_{\{i,j\} \in E(G)}\right)$$

2. Initialize $W \leftarrow M$, $m \leftarrow n$ and $\lambda \leftarrow \lambda_0$

3. Repeat for ℓ iterations:

   a. For each pair of distinct $i, j \in [m]$, let $(x^1_{ij}, x^2_{ij}, x^3_{ij}, x^4_{ij}) = f_{\mathrm{cl}}(W_{ij})$

   b. Let $W' \in \mathbb{R}^{2m \times 2m}$ be the symmetric matrix with $W'_{ii} = 0$ and

$$W'_{ij} = x^1_{ij}$$
$$W'_{(2m+1-i)j} = x^2_{ij}$$
$$W'_{i(2m+1-j)} = x^3_{ij}$$
$$W'_{(2m+1-i)(2m+1-j)} = x^4_{ij}$$

   for all distinct $i, j \in [m]$ and

$$W'_{i,2m+1-i} \sim_{\text{i.i.d.}} Q'_\lambda$$

   for all $i \in [m]$

   c. Generate a permutation σ on $[2m]$ uniformly at random

   d. Update $W \leftarrow (W')^{\sigma,\sigma}$, $m \leftarrow 2m$ and $\lambda \leftarrow g_{\mathrm{cl}}(\lambda)$

4. Output the graph H with $\{i,j\} \in E(H)$ if $W_{ij} > t$

Figure 8: Distributional lifting procedure in Theorem 5.5.

- *a rejection kernel* RK *that can be computed in randomized poly(n) time and parameter λ_0 such that* $\mathrm{RK}(Bern(p')) \sim \tilde{P}_{\lambda_0}$ *and* $\mathrm{RK}(Bern(q')) \sim \tilde{Q}_{\lambda_0}$;

- *a cloning map f_{cl} that can be computed in randomized poly(n) time and parameter map g_{cl} such that*

$$f_{cl}(P_\lambda) \sim P^{\otimes 4}_{g_{cl}(\lambda)} \quad and \quad f_{cl}(Q_\lambda) \sim Q^{\otimes 4}_{g_{cl}(\lambda)}$$

  for each parameter λ;

- *a randomized poly(n) time algorithm for sampling from Q'_{λ_i} for each $1 \leq i \leq \ell$ where the sequence of parameters λ_i are such that $\lambda_{i+1} = g_{cl}(\lambda_i)$ for each i; and*

- *a threshold $t \in \mathbb{R}$.*

Then $\phi = $ DISTRIBUTIONAL-LIFTING with these parameters is a randomized polynomial time computable map $\phi : \mathcal{G}_n \to \mathcal{G}_{2^\ell n}$ such that under both H_0 and H_1, it holds that

$$d_{TV}\left(\phi(\mathrm{PDS}(n,k,p',q')), \mathrm{PDS}\left(2^\ell n, 2^\ell k, p, q\right)\right) \leq \binom{n}{2} \cdot \max\left\{d_{TV}\left(\tilde{P}_{\lambda_0}, P_{\lambda_0}\right), d_{TV}\left(\tilde{Q}_{\lambda_0}, Q_{\lambda_0}\right)\right\}$$
$$+ \sum_{i=1}^{\ell}\left(2^i n \cdot d_{TV}\left(Q_{\lambda_i}, Q'_{\lambda_i}\right) + \sqrt{\frac{\chi^2(Q_{\lambda_i}, P_{\lambda_i})}{2}}\right)$$

where $p = \mathbb{P}_{X \sim P_{\lambda_\ell}}[X > t]$ and $q = \mathbb{P}_{X \sim Q_{\lambda_\ell}}[X > t]$.

Proof. If $\ell = O(\log n)$, the algorithm DISTRIBUTIONAL-LIFTING runs in randomized polynomial time. Let $\phi_i(W)$ be the algorithm that outputs the value of W after i iterations of Step 4 given the value of W in Step 2 as its input. Let $\phi'_i(G)$ be the algorithm that outputs the value of W after i iterations of Step 4 given the original graph G as its input. Note that ϕ'_0 outputs the value of M after Step 1.

We first consider an iteration $\phi_1(M)$ of Step 3 applied to $M \sim M_m(S, P_\lambda, Q_\lambda)$ where $|S| = k$. By the definition of f_{cl}, if i, j are distinct and both in S, then $(x_{ij}^1, x_{ij}^2, x_{ij}^3, x_{ij}^4) \sim P_{g_{\mathrm{cl}}(\lambda)}^{\otimes 4}$. Similarly, if at least one of i or j is not in S, then $(x_{ij}^1, x_{ij}^2, x_{ij}^3, x_{ij}^4) \sim Q_{g_{\mathrm{cl}}(\lambda)}^{\otimes 4}$. Therefore the symmetric matrix W' constructed in Step 3b has independent entries below its main diagonal and satisfies that:

- $W'_{ij} \sim P_{g_{\mathrm{cl}}(\lambda)}$ for all distinct $i, j \in S' = S \cup \{2m + 1 - i : i \in S\}$ with $i + j \neq 2m + 1$;

- $W'_{ij} \sim Q_{g_{\mathrm{cl}}(\lambda)}$ for all distinct $(i, j) \notin S' \times S'$;

- $W'_{ij} \sim Q'_{g_{\mathrm{cl}}(\lambda)}$ with $i + j = 2m + 1$; and

- $W'_{ii} = 0$.

Let W'_r be the matrix with each of its entries identically distributed to those of W' except $(W'_r)_{ij} \sim Q_{g_{\mathrm{cl}}(\lambda)}$ if $i + j = 2m + 1$. Coupling entries individually yields that

$$d_{\mathrm{TV}}(\mathcal{L}(W'), \mathcal{L}(W'_r)) \leq m \cdot d_{\mathrm{TV}}\left(Q_{g_{\mathrm{cl}}(\lambda)}, Q'_{g_{\mathrm{cl}}(\lambda)}\right)$$

Now consider the matrix $W_r = (W'_r)^{\sigma, \sigma}$ conditioned on the two sets $\sigma(S)$ and $\sigma(S' \backslash S)$ where σ is a uniformly at random chosen permutation on $[2m]$. We will show that this matrix is close in total variation to $M_{2m}(\sigma(S'), P_{g_{\mathrm{cl}}(\lambda)}, Q_{g_{\mathrm{cl}}(\lambda)})$. Note that fully conditioned on σ, the entries of W_r below the main diagonal are independent and identically distributed to $M_{2m}(\sigma(S'), P_{g_{\mathrm{cl}}(\lambda)}, Q_{g_{\mathrm{cl}}(\lambda)})$ other than the entries with indices $(\sigma(i), \sigma(2m + 1 - i))$ where $i \in S'$. These entries are distributed as $Q_{g_{\mathrm{cl}}(\lambda)}$ in W_r conditioned on σ and as $P_{g_{\mathrm{cl}}(\lambda)}$ in the target distribution $M_{2m}(\sigma(S'), P_{g_{\mathrm{cl}}(\lambda)}, Q_{g_{\mathrm{cl}}(\lambda)})$. Marginalizing to only condition on the sets $\sigma(S)$ and $\sigma(S', S)$, yields that all entries $(W_r)_{ij}$ with $(i, j) \notin \sigma(S) \times \sigma(S', S) \cup \sigma(S', S) \times \sigma(S)$ are identically distributed in $W_r | \{\sigma(S), \sigma(S' \backslash S)\}$ and the target distribution. Coupling these corresponding entries yields that the total variation between $W_r | \{\sigma(S), \sigma(S' \backslash S)\}$ and the target distribution satisfies that

$$d_{\mathrm{TV}}\left(\mathcal{L}(W_r | \sigma(S), \sigma(S' \backslash S)), \ M_{2m}(\sigma(S'), P_{g_{\mathrm{cl}}(\lambda)}, Q_{g_{\mathrm{cl}}(\lambda)})\right)$$
$$= d_{\mathrm{TV}}\left(\mathcal{L}(W_r[\sigma(S) \times \sigma(S' \backslash S)]), M_k(P_{g_{\mathrm{cl}}(\lambda)})\right)$$

Now observe that $W_r[\sigma(S) \times \sigma(S' \backslash S)]$ is distributed as $\mathcal{L}(A^{\mathrm{id},\tau})$ where τ is permutation of $[k]$ selected uniformly at random and A is a $k \times k$ matrix with its diagonal entries i.i.d. $Q_{g_{\mathrm{cl}}(\lambda)}$ and its other entries i.i.d. $P_{g_{\mathrm{cl}}(\lambda)}$. By Lemma 4.1, we therefore have that

$$d_{\mathrm{TV}}\left(\mathcal{L}(W_r[\sigma(S) \times \sigma(S'\backslash S)]), M_k(P_{g_{\mathrm{cl}}(\lambda)})\right) = d_{\mathrm{TV}}\left(\mathcal{L}(A^{\mathrm{id},\tau}), M_k(P_{g_{\mathrm{cl}}(\lambda)})\right)$$
$$\leq \sqrt{\frac{1}{2} \cdot \chi^2\left(Q_{g_{\mathrm{cl}}(\lambda)}, P_{g_{\mathrm{cl}}(\lambda)}\right)}$$

Now consider the matrix $\phi_1(M) = W = (W')^{\sigma,\sigma}$. By the data processing inequality, we have that

$$d_{\mathrm{TV}}\left(\mathcal{L}(W_r|\sigma(S), \sigma(S'\backslash S)), \mathcal{L}(W|\sigma(S), \sigma(S'\backslash S)\right) \leq d_{\mathrm{TV}}(\mathcal{L}(W'), \mathcal{L}(W_r'))$$
$$\leq m \cdot d_{\mathrm{TV}}\left(Q_{g_{\mathrm{cl}}(\lambda)}, Q'_{g_{\mathrm{cl}}(\lambda)}\right)$$

The triangle inequality now implies that

$$d_{\mathrm{TV}}\left(\mathcal{L}(W|\sigma(S), \sigma(S'\backslash S)),\ M_{2m}(\sigma(S'), P_{g_{\mathrm{cl}}(\lambda)}, Q_{g_{\mathrm{cl}}(\lambda)})\right)$$
$$\leq m \cdot d_{\mathrm{TV}}\left(Q_{g_{\mathrm{cl}}(\lambda)}, Q'_{g_{\mathrm{cl}}(\lambda)}\right) + \sqrt{\frac{1}{2} \cdot \chi^2\left(Q_{g_{\mathrm{cl}}(\lambda)}, P_{g_{\mathrm{cl}}(\lambda)}\right)}$$

Letting S be chosen uniformly at random over all subsets of $[n]$ of size k and the triangle inequality now imply that

$$d_{\mathrm{TV}}\left(\phi_1(M_m(k, P_\lambda, Q_\lambda)), M_{2m}(2k, P_{g_{\mathrm{cl}}(\lambda)}, Q_{g_{\mathrm{cl}}(\lambda)})\right)$$
$$\leq \mathbb{E}_S \mathbb{E}_{\sigma(S), \sigma(S'\backslash S)}\left[d_{\mathrm{TV}}\left(\mathcal{L}(W|\sigma(S), \sigma(S'\backslash S)), M_{2m}(\sigma(S'), P_{g_{\mathrm{cl}}(\lambda)}, Q_{g_{\mathrm{cl}}(\lambda)})\right)\right]$$
$$\leq m \cdot d_{\mathrm{TV}}\left(Q_{g_{\mathrm{cl}}(\lambda)}, Q'_{g_{\mathrm{cl}}(\lambda)}\right) + \sqrt{\frac{1}{2} \cdot \chi^2\left(Q_{g_{\mathrm{cl}}(\lambda)}, P_{g_{\mathrm{cl}}(\lambda)}\right)}$$

For each $i \geq 0$, combining this inequality with the triangle inequality and data processing inequality yields that

$$d_{\mathrm{TV}}\left(\phi_{i+1}(M_m(k, P_{\lambda_0}, Q_{\lambda_0})), M_{2^{i+1}m}\left(2^{i+1}k, P_{\lambda_{i+1}}, Q_{\lambda_{i+1}}\right)\right)$$
$$\leq d_{\mathrm{TV}}\left(\phi_1 \circ \phi_i(M_m(k, P_{\lambda_0}, Q_{\lambda_0})), \phi_1\left(M_{2^i m}\left(2^i k, P_{\lambda_i}, Q_{\lambda_i}\right)\right)\right)$$
$$+ d_{\mathrm{TV}}\left(\phi_1\left(M_{2^i m}\left(2^i k, P_{\lambda_i}, Q_{\lambda_i}\right)\right), M_{2^{i+1}m}\left(2^{i+1}k, P_{\lambda_{i+1}}, Q_{\lambda_{i+1}}\right)\right)$$
$$\leq d_{\mathrm{TV}}\left(\phi_i(M_m(k, P_{\lambda_0}, Q_{\lambda_0})), M_{2^i m}\left(2^i k, P_{\lambda_i}, Q_{\lambda_i}\right)\right)$$
$$+ 2^i m \cdot d_{\mathrm{TV}}\left(Q_{\lambda_{i+1}}, Q'_{\lambda_{i+1}}\right) + \sqrt{\frac{1}{2} \cdot \chi^2\left(Q_{\lambda_{i+1}}, P_{\lambda_{i+1}}\right)}$$

Now note that the adjacency matrix $A_{ij}(G) = \mathbf{1}_{\{i,j\} \in E(G)}$ of $G \sim G(n, p', q', S)$ is distributed as $M_n(S, \mathrm{Bern}(p'), \mathrm{Bern}(q'))$. Note that ϕ_0' applies RK element-wise to the entries below the main diagonal of $A_{ij}(G)$. Coupling each of the independent entries below the diagonals of $\phi_0'(G(n, p', q', S))$ and $M_n(S, P_{\lambda_0}, Q_{\lambda_0})$ separately, we have that if $|S| = k$ then

$$d_{\mathrm{TV}}\left(\phi_0'(G(n, p', q', S)), M_n(S, P_{\lambda_0}, Q_{\lambda_0})\right) \leq \binom{k}{2} \cdot d_{\mathrm{TV}}\left(\tilde{P}_{\lambda_0}, P_{\lambda_0}\right) + \left(\binom{n}{2} - \binom{k}{2}\right) \cdot d_{\mathrm{TV}}\left(\tilde{Q}_{\lambda_0}, Q_{\lambda_0}\right)$$

Taking S to be uniformly distributed over all k element subsets of $[n]$ yields by triangle inequality,

$$d_{\mathrm{TV}}\left(\phi_0'(G(n, k, p', q')),\ M_n(k, P_{\lambda_0}, Q_{\lambda_0})\right)$$
$$\leq \mathbb{E}_S\left[d_{\mathrm{TV}}\left(\phi_0'(G(n, p', q', S)), M_n(S, P_{\lambda_0}, Q_{\lambda_0})\right)\right]$$
$$\leq \binom{n}{2} \cdot \max\left\{d_{\mathrm{TV}}\left(\tilde{P}_{\lambda_0}, P_{\lambda_0}\right), d_{\mathrm{TV}}\left(\tilde{Q}_{\lambda_0}, Q_{\lambda_0}\right)\right\}$$

Applying the bounds above iteratively, the triangle inequality and the data processing inequality now yields that

$$d_{\mathrm{TV}}\left(\phi'_\ell(G(n,k,p',q')), M_{2^\ell n}(2^\ell k, P_{\lambda_\ell}, Q_{\lambda_\ell})\right)$$

$$\leq d_{\mathrm{TV}}\left(\phi_\ell \circ \phi'_0(G(n,k,p)), \phi_\ell\left(M_n(k, P_{\lambda_0}, Q_{\lambda_0})\right)\right) + d_{\mathrm{TV}}\left(\phi_\ell(M_n(k, P_{\lambda_0}, Q_{\lambda_0})), M_{2^\ell n}(2^\ell k, P_{\lambda_\ell}, Q_{\lambda_\ell})\right)$$

$$\leq d_{\mathrm{TV}}\left(\phi'_0(G(n,k,p)), M_n(k, P_{\lambda_0}, Q_{\lambda_0})\right) + \sum_{i=1}^{\ell}\left(2^{i-1}n \cdot d_{\mathrm{TV}}\left(Q_{\lambda_i}, Q'_{\lambda_i}\right) + \sqrt{\frac{\chi^2(Q_{\lambda_i}, P_{\lambda_i})}{2}}\right)$$

$$\leq \binom{n}{2} \cdot \max\left\{d_{\mathrm{TV}}\left(\tilde{P}_{\lambda_0}, P_{\lambda_0}\right), d_{\mathrm{TV}}\left(\tilde{Q}_{\lambda_0}, Q_{\lambda_0}\right)\right\} + \sum_{i=1}^{\ell}\left(2^{i-1}n \cdot d_{\mathrm{TV}}\left(Q_{\lambda_i}, Q'_{\lambda_i}\right) + \sqrt{\frac{\chi^2(Q_{\lambda_i}, P_{\lambda_i})}{2}}\right)$$

We now deal with the case of H_0. By the same reasoning, we have that

$$d_{\mathrm{TV}}\left(\phi'_0(G(n,q')), M_n(Q_{\lambda_0})\right) \leq \binom{n}{2} \cdot d_{\mathrm{TV}}\left(\tilde{Q}_{\lambda_0}, Q_{\lambda_0}\right)$$

Now note that if $M \sim M_m(Q_{\lambda_0})$, every entry of W'_{ij} in Step 3b below the main diagonal is i.i.d. sampled from Q_{λ_1} other than those with $i + j = 2m + 1$, which are sampled from Q'_{λ_1}. Coupling entries individually implies that

$$d_{\mathrm{TV}}\left(\phi_1(M), M_{2m}(Q_{\lambda_1})\right) \leq m \cdot d_{\mathrm{TV}}\left(Q_{\lambda_1}, Q'_{\lambda_1}\right)$$

By induction, the data processing and triangle inequalities imply that

$$d_{\mathrm{TV}}\left(\phi_\ell(M), M_{2m}(Q_{\lambda_\ell})\right) \leq \sum_{i=1}^{\ell} 2^{i-1}m \cdot d_{\mathrm{TV}}\left(Q_{\lambda_i}, Q'_{\lambda_i}\right)$$

Therefore it follows that

$$d_{\mathrm{TV}}\left(\phi'_\ell(G(n,q')), M_{2^\ell n}(Q_{\lambda_\ell})\right) \leq d_{\mathrm{TV}}\left(\phi_\ell \circ \phi'_0(G(n,q')), \phi_\ell\left(M_n(Q_{\lambda_0})\right)\right)$$
$$+ d_{\mathrm{TV}}\left(\phi_\ell(M_n(Q_{\lambda_0})), M_{2^\ell n}(Q_{\lambda_\ell})\right)$$
$$\leq d_{\mathrm{TV}}\left(\phi'_0(G(n,q')), M_n(Q_{\lambda_0})\right) + \sum_{i=1}^{\ell} 2^{i-1}m \cdot d_{\mathrm{TV}}\left(Q_{\lambda_i}, Q'_{\lambda_i}\right)$$
$$\leq \binom{n}{2} \cdot d_{\mathrm{TV}}\left(\tilde{Q}_{\lambda_0}, Q_{\lambda_0}\right) + \sum_{i=1}^{\ell} 2^{i-1}m \cdot d_{\mathrm{TV}}\left(Q_{\lambda_i}, Q'_{\lambda_i}\right)$$

If $W \sim M_{2^\ell n}(2^\ell k, P_{\lambda_\ell}, Q_{\lambda_\ell})$, then the graph with adjacency matrix $A_{ij} = \mathbf{1}_{\{W_{ij} > t\}}$ is distributed as $G(2^\ell n, 2^\ell k, p, q)$ where $p = \mathbb{P}_{X \sim P_{\lambda_\ell}}[X > t]$ and $q = \mathbb{P}_{X \sim Q_{\lambda_\ell}}[X > t]$. Similarly if $W \sim M_{2^\ell n}(Q_{\lambda_\ell})$ then the graph with adjacency matrix $A_{ij} = \mathbf{1}_{\{W_{ij} > t\}}$ is distributed as $G(2^\ell n, q)$. Now combining the total variation bounds above with the data processing inequality proves the theorem. $\qquad\square$

6 Planted Dense Subgraph and Biclustering

6.1 Poisson Lifting and Lower Bounds for Low-Density PDS

In this section, we introduce Poisson lifting to give a reduction from planted clique to $\mathrm{PDS}(n, k, p, q)$ in the regime where $p = cq$ for some fixed $c > 1$ and $q = \tilde{\Theta}(n^{-\alpha})$ for some fixed $\alpha > 0$. This is the

---

Algorithm POISSON-LIFTING

Inputs: Graph $G \in \mathcal{G}_n$, iterations ℓ, parameters $\gamma, \epsilon \in (0,1)$ and $c > 1$ with $3\epsilon^{-1} \leq \log_c \gamma^{-1}$

Return the output of DISTRIBUTIONAL-LIFTING applied to G with ℓ iterations and parameters:

- initial edge densities $p' = 1$ and $q' = \gamma$

- target families $P_\lambda = \mathrm{Pois}(c\lambda)$ and $Q_\lambda = Q'_\lambda = \mathrm{Pois}(\lambda)$

- rejection kernel $\mathrm{RK}_{P1} = \mathrm{RK}\left(1 \to \mathrm{Pois}(c\lambda_0), \gamma \to \mathrm{Pois}(\lambda_0), \lceil 6\log_{\gamma^{-1}} n \rceil\right)$ and $\lambda_0 = n^{-\epsilon}$

- cloning map $f_{\mathrm{cl}}(x) = (x_1, x_2, x_3, x_4)$ computed as follows:

    1. Generate x numbers in $[4]$ uniformly at random
    2. Let x_i be the number of i's generated for each $i \in [4]$

- parameter map $g_{\mathrm{cl}}(\lambda) = \lambda/4$

- threshold $t = 0$

Figure 9: Poisson lifting procedure in Lemma 6.1.

same parameter regime as considered in [HWX15] and strengthens their lower bounds to hold for a fixed planted dense subgraph size k. Poisson lifting is a specific instance of DISTRIBUTIONAL-LIFTING with Poisson target distributions. The guarantees of Poisson lifting are captured in the following lemma.

Lemma 6.1 (Poisson Lifting). *Suppose that n and ℓ are such that $\ell = O(\log n)$ and are sufficiently large. Fix arbitrary constants $\epsilon \in (0,1)$ and $c > 1$ and let $\lambda_0 = n^{-\epsilon}$. Suppose that γ is a small enough constant satisfying that $\log_c \gamma^{-1} \geq 3\epsilon^{-1}$. Then $\phi = $ POISSON-LIFTING is a randomized polynomial time computable map $\phi : \mathcal{G}_n \to \mathcal{G}_{2^\ell n}$ such that under both H_0 and H_1, it holds that*

$$d_{TV}\left(\phi(\mathrm{PC}(n, k, \gamma)), \mathrm{PDS}\left(2^\ell n, 2^\ell k, p, q\right)\right) = O\left(n^{-\epsilon/2}\right)$$

where $p = 1 - e^{-4^{-\ell} c\lambda_0}$ and $q = 1 - e^{-4^{-\ell} \lambda_0}$.

Proof. Let POISSON-LIFTING be the algorithm DISTRIBUTIONAL-LIFTING applied with the parameters in Figure 9. Let $\lambda_{i+1} = g_{\mathrm{cl}}(\lambda_i) = \lambda_i/4$ for each $0 \leq i \leq \ell - 1$. Note that the cloning map $f_{\mathrm{cl}}(x)$ can be computed in $O(1)$ operations. Furthermore, if $x \sim \mathrm{Pois}(\lambda)$ then Poisson splitting implies that if $f_{\mathrm{cl}}(x) = (x_1, x_2, x_3, x_4)$ then the x_i are independent and $x_i \sim \mathrm{Pois}(\lambda/4)$. Therefore,

$$f_{\mathrm{cl}}(P_\lambda) = f_{\mathrm{cl}}(\mathrm{Pois}(c\lambda)) \sim \mathrm{Pois}(c\lambda/4)^{\otimes 4} = P_{g_{\mathrm{cl}}(\lambda)}^{\otimes 4} \quad \text{and}$$
$$f_{\mathrm{cl}}(Q_\lambda) = f_{\mathrm{cl}}(\mathrm{Pois}(\lambda)) \sim \mathrm{Pois}(\lambda/4)^{\otimes 4} = Q_{g_{\mathrm{cl}}(\lambda)}^{\otimes 4}$$

Both $P_\lambda = \mathrm{Pois}(c\lambda)$ and $Q_\lambda = Q'_\lambda = \mathrm{Pois}(\lambda)$ can be sampled in $O(1)$ time, and the χ^2 divergence

39

between these distributions is

$$\chi^2\left(Q_\lambda, P_\lambda\right) = -1 + \sum_{t=0}^{\infty} \frac{\left(\frac{1}{t!}e^{-\lambda}\lambda^t\right)^2}{\frac{1}{t!}e^{-c\lambda}(c\lambda)^t} = -1 + \exp\left(c^{-1}(c-1)^2\lambda\right) \cdot \sum_{t=0}^{\infty} \frac{e^{-\lambda/c}(\lambda/c)^t}{t!}$$

$$= \exp\left(c^{-1}(c-1)^2\lambda\right) - 1 \leq 2c^{-1}(c-1)^2\lambda$$

as long as $c^{-1}(c-1)^2\lambda \leq 1$ since $e^x \leq 1 + 2x$ for $x \in [0,1]$. By Lemma 5.2, the rejection kernel RK_{P1} can be computed in $O(\log n)$ time and satisfies that

$$d_{\text{TV}}\left(\text{RK}_{\text{P1}}(1), P_{\lambda_0}\right) = O(n^{-3}) \quad \text{and} \quad d_{\text{TV}}\left(\text{RK}_{\text{P1}}(\text{Bern}(\gamma)), Q_{\lambda_0}\right) = O(n^{-3})$$

Now note that $P_{\lambda_\ell} = \text{Pois}(4^{-\ell}c\lambda_0)$ and $Q_{\lambda_\ell} = \text{Pois}(4^{-\ell}\lambda_0)$ which implies that $p = \mathbb{P}_{X \sim P_{\lambda_\ell}}[X > 0] = 1 - e^{-4^{-\ell}c\lambda_0}$ and $q = \mathbb{P}_{X \sim Q_{\lambda_\ell}}[X > 0] = 1 - e^{-4^{-\ell}\lambda_0}$. Since $\text{PDS}(n, k, 1, \gamma)$ is the same problem as $\text{PC}(n, k, \gamma)$, applying Theorem 5.5 yields that under both H_0 and H_1, we have

$$d_{\text{TV}}\left(\phi(\text{PC}(n, k, \gamma)), \text{PDS}\left(2^\ell n, 2^\ell k, p, q\right)\right)$$

$$\leq \binom{n}{2} \cdot \max\left\{d_{\text{TV}}\left(\text{RK}_{\text{P1}}(1), P_{\lambda_0}\right), d_{\text{TV}}\left(\text{RK}_{\text{P1}}(\text{Bern}(\gamma)), Q_{\lambda_0}\right)\right\} + \sum_{i=1}^{\ell} \sqrt{\frac{\chi^2(Q_{\lambda_i}, P_{\lambda_i})}{2}}$$

$$\leq \binom{n}{2} \cdot O(n^{-3}) + c^{-1/2}(c-1)\sum_{i=1}^{\ell} \sqrt{\lambda_i}$$

$$= O(n^{-1}) + c^{-1/2}(c-1)n^{-\epsilon/2}\sum_{i=1}^{\ell} 2^{-i} = O\left(n^{-1} + n^{-\epsilon/2}\right)$$

which completes the proof of the lemma. $\qquad\square$

We now use the reduction based on Poisson lifting analyzed above to prove hardness for the sparsest regime of PDS. The proof of this theorem is similar to that of Theorem 4.3 and is deferred to Appendix C.

Theorem 6.2. *Fix some $c > 1$. Let $\alpha \in [0, 2)$ and $\beta \in (0, 1)$ be such that $\beta < \frac{1}{2} + \frac{\alpha}{4}$. There is a sequence $\{(N_n, K_n, p_n, q_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. The parameters are in the regime $q = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,

$$\lim_{n \to \infty} \frac{\log q_n^{-1}}{\log N_n} = \alpha, \qquad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta \quad and \quad \lim_{n \to \infty} \frac{p_n}{q_n} = c$$

2. For any sequence of randomized polynomial-time tests $\phi_n : \mathcal{G}_{N_n} \to \{0, 1\}$, the asymptotic Type I+II error of ϕ_n on the problems $\text{PDS}_D(N_n, K_n, p_n, q_n)$ is at least 1 assuming the PC conjecture holds for each fixed density $p \leq 1/2$.

Therefore the computational boundary for $\text{PDS}_D(n, k, p, q)$ in the parameter regime $q = \tilde{\Theta}(n^{-\alpha})$, $\frac{p}{q} \to c$ and $k = \tilde{\Theta}(n^\beta)$ is $\beta^* = \frac{1}{2} + \frac{\alpha}{4}$.

In this section, we gave a planted clique lower bound for $\text{PDS}_D(n, k, p, q)$ with $\frac{p}{q} \to c$ as opposed to $p = cq$ exactly. We now will describe a simple reduction from $\text{PDS}_D(n, k, p, q)$ with $\frac{p}{q} \to c_1$ where $c_1 > c$ to $\text{PDS}_D(n, k, p_1, q_1)$ where $p_1 = cq_1$ and $q_1 = \Theta(q)$. Given an instance of

---
Algorithm GAUSSIAN-LIFTING

Inputs: Graph $G \in \mathcal{G}_n$, iterations ℓ

Return the output of DISTRIBUTIONAL-LIFTING applied to G with ℓ iterations and parameters:

- initial densities $p' = 1$ and $q' = 1/2$

- target families $P_\lambda = N(\lambda, 1)$ and $Q_\lambda = Q'_\lambda = N(0, 1)$

- rejection kernel $\mathrm{RK_G} = \mathrm{RK}\left(1 \to N(\lambda_0, 1), 1/2 \to N(0, 1), N\right)$ where $N = \lceil 6 \log_2 n \rceil$ and $\lambda_0 = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$

- cloning map $f_{\mathrm{cl}}(x) = (x_1, x_2, x_3, x_4)$ computed as follows:

    1. Generate $G_1, G_2, G_3 \sim_{\mathrm{i.i.d.}} N(0, 1)$
    2. Compute (x_1, x_2, x_3, x_4) as

    $$x_1 = \frac{1}{2}\left(x + G_1 + G_2 + G_3\right)$$
    $$x_2 = \frac{1}{2}\left(x - G_1 + G_2 - G_4\right)$$
    $$x_3 = \frac{1}{2}\left(x + G_1 - G_2 - G_3\right)$$
    $$x_4 = \frac{1}{2}\left(x - G_1 - G_2 + G_3\right)$$

- parameter map $g_{\mathrm{cl}}(\lambda) = \lambda/2$

- threshold $t = 0$

Figure 10: Gaussian lifting procedure in Lemma 6.3.

$\mathrm{PDS}_D(n, k, p, q)$, add in every non-edge independently with probability $\rho = \frac{p-cq}{c-1+p-cq}$ which is in $(0, 1)$ since $c_1 > c$ implies that $p > cq$ for large enough n. This yields an instance of PDS_D with $p_1 = 1 - (1-\rho)(1-p) = p + \rho - \rho p$ and $q_1 = 1 - (1-\rho)(1-q) = q + \rho - \rho q$. The choice of ρ implies that $p_1 = cq_1$ exactly and $\rho = \Theta(q)$ since $\frac{p}{q} \to c_1 > c$. Applying this reduction after POISSON-LIFTING yields that $\mathrm{PDS}_D(n, k, cq, q)$ has the same planted clique lower bound as in the previous theorem.

6.2 Gaussian Lifting and Lower Bounds for High-Density PDS and BC

In parallel to the previous section, here we introduce Gaussian lifting to give a reduction from planted clique to the dense regime of $\mathrm{PDS}(n, k, p, q)$ where $q = \Theta(1)$ and $p - q = \tilde{\Theta}(n^{-\alpha})$ for some fixed $\alpha > 0$. As an intermediate step, we also give a reduction strengthening the lower bounds for BC shown in [MW15b] to hold for the canonical simple vs. simple hypothesis testing formulation of BC.

The next lemma we prove is an analogue of Lemma 6.1 for GAUSSIAN-LIFTING and follows the

same structure of verifying the preconditions for and applying Lemma 5.5.

Lemma 6.3 (Gaussian Lifting). *Suppose that n and ℓ are such that $\ell = O(\log n)$ and are sufficiently large and let*

$$\mu = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$$

Then $\phi = \text{GAUSSIAN-LIFTING}$ is a randomized polynomial time computable map $\phi : \mathcal{G}_n \to \mathcal{G}_{2^\ell n}$ such that under both H_0 and H_1, it holds that

$$d_{TV}\left(\phi(\text{PC}(n,k,1/2)), \text{PDS}\left(2^\ell n, 2^\ell k, \Phi\left(2^{-\ell}\mu\right), 1/2\right)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

Proof. Let GAUSSIAN-LIFTING be the algorithm DISTRIBUTIONAL-LIFTING applied with the parameters in Figure 10. Let $\lambda_{i+1} = g_{\text{cl}}(\lambda_i) = \lambda_i/2$ for each $0 \le i \le \ell - 1$. Note that the cloning map $f_{\text{cl}}(x)$ can be computed in $O(1)$ operations. Now suppose that $x \sim N(\lambda, 1)$. If $f_{\text{cl}}(x) = (x_1, x_2, x_3, x_4)$, then it follows that

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \frac{\lambda}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x - \lambda \\ G^1 \\ G^2 \\ G^3 \end{bmatrix}$$

Since $x - \mu, G^1, G^2$ and G^3 are zero-mean and jointly Gaussian with covariance matrix I_4, it follows that the entries of (x_1, x_2, x_3, x_4) are also jointly Gaussian. Furthermore, the coefficient matrix above is orthonormal, implying that the covariance matrix of (x_1, x_2, x_3, x_4) remains I_4. Therefore it follows that $f_{\text{cl}}(N(\lambda, 1)) \sim N(\lambda/2, 1)^{\otimes 4}$. Applying this identity with $\lambda = 0$ yields that $f_{\text{cl}}(N(0,1)) \sim N(0,1)^{\otimes 4}$. Thus f_{cl} is a valid cloning map for P_λ and Q_λ with parameter map $g_{\text{cl}}(\lambda) = \lambda/2$.

Observe that $P_\lambda = N(\lambda, 1)$ and $Q_\lambda = Q'_\lambda = N(0,1)$ can be sampled in $O(1)$ time in the given computational model. Note that the χ^2 divergence between these distributions is

$$\chi^2(Q_\lambda, P_\lambda) = -1 + \int_{-\infty}^{\infty} \frac{\left(\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right)^2}{\frac{1}{\sqrt{2\pi}}e^{-(x-\lambda)^2/2}} dx = -1 + \frac{e^{2\lambda^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x+\lambda)^2/2} dx = e^{2\lambda^2} - 1 \le 4\lambda^2$$

as long as $4\lambda^2 \le 1$ since $e^x \le 1 + 2x$ for $x \in [0,1]$, which is the case for all $\lambda = \lambda_i$. By Lemma 5.4, the rejection kernel RK_G can be computed in $\text{poly}(n)$ time and satisfies that

$$d_{TV}(\text{RK}_G(1), P_{\lambda_0}) = O(n^{-3}) \quad \text{and} \quad d_{TV}(\text{RK}_G(\text{Bern}(1/2)), Q_{\lambda_0}) = O(n^{-3})$$

Now note that $P_{\lambda_\ell} = N(2^{-\ell}\mu, 1)$ and $Q_{\lambda_\ell} = N(0,1)$ which implies that $p = \mathbb{P}_{X \sim P_{\lambda_\ell}}[X > 0] = \Phi\left(2^{-\ell}\mu\right)$ and $q = \mathbb{P}_{X \sim Q_{\lambda_\ell}}[X > 0] = 1/2$. Since $\text{PDS}(n, k, 1, 1/2)$ is the same problem as $\text{PC}(n, k, 1/2)$, applying Theorem 5.5 yields that under both H_0 and H_1, we have

$$d_{TV}\left(\phi(\text{PC}(n,k,1/2)), \text{PDS}\left(2^\ell n, 2^\ell k, p, q\right)\right)$$

$$\le \binom{n}{2} \cdot \max\left\{d_{TV}(\text{RK}_G(1), P_{\lambda_0}), d_{TV}(\text{RK}_G(\text{Bern}(1/2)), Q_{\lambda_0})\right\} + \sum_{i=1}^{\ell} \sqrt{\frac{\chi^2(Q_{\lambda_i}, P_{\lambda_i})}{2}}$$

$$\le \binom{n}{2} \cdot O(n^{-3}) + \sqrt{2} \cdot \sum_{i=1}^{\ell} \lambda_i$$

$$= O(n^{-1}) + \mu\sqrt{2} \cdot \sum_{i=1}^{\ell} 2^{-i} = O\left(n^{-1} + \frac{1}{\sqrt{\log n}}\right)$$

42

Algorithm BC-REDUCTION

Inputs: Graph $G \in \mathcal{G}_n$, iterations ℓ

1. Set W to be the output of GAUSSIAN-LIFTING applied to G with ℓ iterations without the thresholding in Step 4 of DISTRIBUTIONAL-LIFTING

2. Replace the diagonal entries with $W_{ii} \sim_{\text{i.i.d.}} N(0, 2)$

3. Generate an antisymmetric $2^\ell n \times 2^\ell n$ matrix A of with i.i.d. $N(0, 1)$ random variables below its main diagonal and set

$$W \leftarrow \frac{1}{\sqrt{2}}(W + A)$$

4. Generate a permutation σ of $[2^\ell n]$ uniformly at random and output $W^{\text{id},\sigma}$

Figure 11: Reduction to biclustering in Lemma 6.5.

which completes the proof of the lemma. □

We now use this GAUSSIAN-LIFTING reduction to deduce hardness for the dense variant of planted dense subgraph, which has a slightly different computational boundary of than the sparsest variant, given by $p - q \approx n/k^2$. The proof of the next theorem is deferred to Appendix C

Theorem 6.4. *Let $\alpha \in [0, 2)$ and $\beta \in (0, 1)$ be such that $\beta < \frac{1}{2} + \frac{\alpha}{2}$. There is a sequence $\{(N_n, K_n, p_n, q_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. *The parameters are in the regime $q = \Theta(1)$, $p - q = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,*

$$\lim_{n \to \infty} \frac{\log(p_n - q_n)^{-1}}{\log N_n} = \alpha \quad and \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. *For any sequence of randomized polynomial-time tests $\phi_n : \mathcal{G}_{N_n} \to \{0, 1\}$, the asymptotic Type I+II error of ϕ_n on the problems $\text{PDS}_D(N_n, K_n, p_n, q_n)$ is at least 1 assuming the PC conjecture holds with density $p = 1/2$.*

Therefore the computational boundary for $\text{PDS}_D(n, k, p, q)$ in the parameter regime $q = \Theta(1)$, $p - q = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\beta^* = \frac{1}{2} + \frac{\alpha}{2}$.

Note that to prove this theorem, it was only necessary to map to instances with ambient density $q = 1/2$. We remark that it is possible to map from $q = 1/2$ to any constant q by removing edges with a constant probability $\rho < 1$ or removing non-edges with probability ρ. Note that this still preserves the asymptotic regime $p - q = \tilde{\Theta}(n^{-\alpha})$. We now use GAUSSIAN-LIFTING to give a reduction from planted clique to biclustering. This next lemma uses similar ingredients to the proof of Theorem 5.5 and analysis of the cloning technique in GAUSSIAN-LIFTING. The proof is deferred to Appendix C.

Lemma 6.5. *Suppose that n and ℓ are such that $\ell = O(\log n)$ and are sufficiently large and*

$$\mu = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$$

Then there is a randomized polynomial time computable map $\phi = \text{BC-REDUCTION}$ with $\phi : \mathcal{G}_n \to \mathbb{R}^{2^\ell n \times 2^\ell n}$ such that under H_0 and H_1, it holds that

$$d_{TV}\left(\phi(\text{PC}(n, k, 1/2)), \text{BC}\left(2^\ell n, 2^\ell k, 2^{-\ell-1/2}\mu\right)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

Note that Lemma 6.5 provides a randomized polynomial time map that exactly reduces from $\text{PC}_D(n, k, 1/2)$ to $\text{BC}_D(2^\ell n, 2^\ell k, 2^{-\ell-1/2}\mu)$. This reduction yields tight computational lower bounds for a simple vs. simple hypothesis testing variant of biclustering as stated in Theorem 6.6. This follows from setting ℓ_n, k_n, N_n and K_n as in Theorem 6.4 and $\mu_n = 2^{-\ell_n-1/2}\mu$, then applying an identical analysis as in Theorem 6.4. Note that when $\beta < \frac{1}{2}$, this choice sets $\ell_n = 0$ and deduces that BC_D is hard when $\alpha > 0$.

Theorem 6.6. *Let $\alpha > 0$ and $\beta \in (0, 1)$ be such that $\beta < \frac{1}{2} + \frac{\alpha}{2}$. There is a sequence $\{(N_n, K_n, \mu_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. *The parameters are in the regime $\mu = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,*

$$\lim_{n \to \infty} \frac{\log \mu_n^{-1}}{\log N_n} = \alpha \quad and \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. *For any sequence of randomized polynomial-time tests $\phi_n : \mathbb{R}^{N_n \times N_n} \to \{0, 1\}$, the asymptotic Type I+II error of ϕ_n on the problems $\text{BC}_D(N_n, K_n, \mu_n)$ is at least 1 assuming the PC conjecture holds with density $p = 1/2$.*

Therefore the computational boundary for $\text{BC}_D(n, k, \mu)$ in the parameter regime $\mu = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\beta^* = \frac{1}{2} + \frac{\alpha}{2}$ and $\alpha^* = 0$ when $\beta < \frac{1}{2}$.

We now deduce the computational barrier for the biclustering recovery problem from the PDS recovery conjecture. The obtained boundary of $\beta^* = \frac{1}{2} + \alpha$ is stronger than the detection boundary $\beta^* = \frac{1}{2} + \frac{\alpha}{2}$ in the previous theorem. First we will need the following lemma, which gives the necessary total variation guarantees for our reduction. We omit details that are identical to the proof of Lemma 6.5.

Lemma 6.7. *Suppose that n, μ and $\rho \geq n^{-1}$ are such that*

$$\mu = \frac{\log(1 + 2\rho)}{2\sqrt{6 \log n + 2 \log 2}}$$

Then there is a randomized polynomial time computable map $\phi = \text{BC-RECOVERY}$ with $\phi : \mathcal{G}_n \to \mathbb{R}^{n \times n}$ such that for any subset $S \subseteq [n]$ with $|S| = k$, it holds that

$$d_{TV}\left(\phi\left(G(n, 1/2 + \rho, 1/2, S)\right), \int \mathcal{L}\left(\mu \cdot \mathbf{1}_S \mathbf{1}_T^\top + N(0, 1)^{\otimes n \times n}\right) d\pi(T)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

where π is the uniform distribution on subsets of $[n]$ of size k.

Algorithm BC-RECOVERY

Inputs: Graph $G \in \mathcal{G}_n$, density bias ρ

1. Let $\text{RK}_G = \text{RK}\left(\frac{1}{2} + \rho \to N(\mu, 1), \frac{1}{2} \to N(0, 1), N\right)$ where $\mu = \frac{\log(1+2\rho)}{2\sqrt{6\log n + 2\log 2}}$ and $N = \lceil 6\log_{1+2\rho} n \rceil$ and compute the symmetric matrix $W \in \mathbb{R}^{n \times n}$ with

$$W_{ij} = \text{RK}_G\left(\mathbf{1}_{\{i,j\} \in E(G)}\right)$$

   for all $i \neq j$ and $W_{ii} \sim_{\text{i.i.d.}} N(0, 2)$

2. Generate an antisymmetric matrix $A \in \mathbb{R}^{n \times n}$ of with i.i.d. $N(0, 1)$ random variables below its main diagonal and set

$$W \leftarrow \frac{1}{\sqrt{2}}(W + A)$$

3. Generate a permutation σ of $[n]$ uniformly at random and output $W^{\text{id}, \sigma}$

Figure 12: Reduction to biclustering recovery in Lemma 6.7.

Proof. Let $\phi = $ BC-RECOVERY be as in Figure 12. Applying Lemma 5.4, it holds that RK_G can be computed in $\text{poly}(n)$ time and that

$$d_{\text{TV}}\left(\text{RK}_G(\text{Bern}(1/2 + \rho)), N(\mu, 1)\right) = O(n^{-3}) \quad \text{and} \quad d_{\text{TV}}\left(\text{RK}_G(\text{Bern}(1/2)), N(0, 1)\right) = O(n^{-3})$$

Let W_1 and W_2 be the values of W after Steps 1 and 2, respectively, applied to an input graph $G \sim G(n, 1/2 + \rho, 1/2, S)$. Let M be a sample from $M_n(S, N(\mu, 1), N(0, 1))$ with i.i.d. $N(0, 2)$ random variables on its diagonal. Coupling entries individually yields that

$$d_{\text{TV}}\left(\mathcal{L}(W_1), \mathcal{L}(M)\right) \leq \binom{k}{2} \cdot d_{\text{TV}}\left(\text{RK}_G(\text{Bern}(1/2 + \rho)), N(\mu, 1)\right)$$

$$+ \left(\binom{n}{2} - \binom{k}{2}\right) \cdot d_{\text{TV}}\left(\text{RK}_G(\text{Bern}(1/2)), N(0, 1)\right)$$

$$= \binom{n}{2} \cdot O(n^{-3}) = O(n^{-1})$$

An identical argument as in Lemma 6.5 now shows that W_2 is at total variation distance $O(n^{-1})$ from $\mu \cdot \mathbf{1}_S \mathbf{1}_S^\top + N(0, 1)^{\otimes n \times n}$ with all of its diagonal entries replaced with i.i.d. samples from $N(0, 1)$. The same permutation argument applying Lemma 4.1 now yields that

$$d_{\text{TV}}\left(\mathcal{L}((W_2')^{\text{id}, \sigma}), \int \mathcal{L}\left(\mu \cdot \mathbf{1}_S \mathbf{1}_T^\top + N(0, 1)^{\otimes n \times n}\right) d\pi(T)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

Applying the triangle and data processing inequalities as in the conclusion of Lemma 6.5 completes the proof of the lemma. $\square$

With this lemma, we now deduce the recovery barrier for biclustering from the PDS and PC conjectures. Note that the recovery barrier of $\beta^* = \frac{1}{2} + \alpha$ and detection barrier of $\beta^* = \frac{1}{2} + \frac{\alpha}{2}$ indicates that recovery is conjectured to be strictly harder than detection for the formulations we consider in the regime $\beta > \frac{1}{2}$.

Theorem 6.8. *Let $\alpha > 0$ and $\beta \in (0,1)$. There is a sequence $\{(N_n, K_n, \mu_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. *The parameters are in the regime $\mu = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,*

$$\lim_{n \to \infty} \frac{\log \mu_n^{-1}}{\log N_n} = \alpha \quad and \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. *If $\beta \geq \frac{1}{2}$ and $\beta < \frac{1}{2} + \alpha$, then the following holds. Let $\epsilon > 0$ be fixed and let M_n be an instance of $\mathrm{BC}_R(N_n, K_n, \mu_n)$. There is no sequence of randomized polynomial-time computable functions $\phi_n : \mathbb{R}^{N_n \times N_n} \to \binom{[N_n]}{k}^2$ such that for all sufficiently large n the probability that $\phi_n(M_n)$ is exactly the pair of latent row and column supports of M_n is at least ϵ, assuming the PDS recovery conjecture.*

3. *If $\beta < \frac{1}{2}$ and $\alpha > 0$, then the following holds. There is no sequence of randomized polynomial-time computable functions $\phi_n : \mathbb{R}^{N_n \times N_n} \to \binom{[N_n]}{k}^2$ such that for all sufficiently large n the probability that $\phi_n(M_n)$ is exactly the pair of latent row and column supports of M_n is at least ϵ, assuming the PC conjecture.*

Therefore, given the PDS recovery conjecture, the computational boundary for $\mathrm{BC}_R(n, k, \mu)$ in the parameter regime $\mu = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\beta^* = \frac{1}{2} + \alpha$ when $\beta \geq \frac{1}{2}$ and $\alpha^* = 0$ when $\beta < \frac{1}{2}$.

Proof. First we consider the case when $\beta \geq 1/2$ and $\beta < \frac{1}{2} + \alpha$. Now set

$$k_n = \lceil n^\beta \rceil, \qquad \rho_n = n^{-\alpha}, \qquad N_n = n \qquad K_n = k_n, \qquad \mu_n = \frac{\log(1 + 2\rho_n)}{2\sqrt{6 \log n + 2 \log 2}}$$

Assume for contradiction that there is a sequence of randomized polynomial-time computable functions ϕ_n as described above and let ϕ_n^r denote the restriction of ϕ_n to output latent row support only. Let $\varphi_n = \mathrm{BC\text{-}RECOVERY}$ be the reduction in Lemma 6.7, let $G_n \sim G(n, S, 1/2 + \rho_n, 1/2)$ and let $M_n = \varphi_n(G_n)$ where S is a k_n-subset of $[n]$. Let $\mathcal{L}_{n,S,T} = \mathcal{L}\left(\mu_n \mathbf{1}_S \mathbf{1}_T^\top + N(0,1)^{\otimes n \times n}\right)$ and let $\mathcal{L}_{n,S} = \int \mathcal{L}_{n,S,T} d\pi(T)$ where π is the uniform distribution over k_n-subsets of $[n]$ be the distribution of an instance of $\mathrm{BC}_R(N_n, K_n, \mu_n)$ conditioned on the event that the row support of its planted submatrix is S. Now observe that

$$\left| \mathbb{P}_{M \sim \mathcal{L}(M_n)} \left[\phi_n^r(M) = S \right] - \mathbb{P}_{M \sim \mathcal{L}_{n,S}} \left[\phi_n^r(M) = S \right] \right| \leq d_{\mathrm{TV}} \left(\mathcal{L}(M_n), \mathcal{L}_{n,S} \right) = O\left(\frac{1}{\sqrt{\log n}} \right)$$

because of Lemma 6.7. Now note that $\mathbb{P}_{M \sim \mathcal{L}_{n,S}} [\phi_n^r(M) = S] = \mathbb{E}_{T \sim \pi} \mathbb{P}_{M \sim \mathcal{L}_{n,S,T}} [\phi_n^r(M) = S] \geq \epsilon$ for sufficiently large n by assumption. Therefore it follows that

$$\mathbb{P}[\phi_n^r \circ \varphi_n(G_n) = S] \geq \epsilon - O\left(\frac{1}{\sqrt{\log n}} \right)$$

which is at least $\epsilon/2$ for sufficiently large n. Now observe that

$$\lim_{n \to \infty} \frac{\log k_n}{\log n} = \beta \quad and \quad \lim_{n \to \infty} \log_n \left(\frac{k_n^2 \rho_n^2}{\frac{1}{4} - \rho_n^2} \right) = 2\beta - 2\alpha < 1$$

Since the sequence of functions $\phi_n^r \circ \varphi_n$ can be computed in randomized polynomial time, this contradicts the PDS recovery conjecture. Therefore no such sequence of functions ϕ_n exists for the parameter sequence $\{(N_n, K_n, \mu_n)\}_{n \in \mathbb{N}}$ defined above. Now note that as $n \to \infty$,

$$\mu_n = \frac{\log(1 + 2\rho_n)}{2\sqrt{6 \log n + 2 \log 2}} \sim \frac{\rho_n}{\sqrt{6 \log n}} = \frac{n^{-\alpha}}{\sqrt{6 \log n}}$$

Therefore it follows that

$$\lim_{n\to\infty} \frac{\log \mu_n^{-1}}{\log N_n} = \lim_{n\to\infty} \frac{\alpha \log n + \frac{1}{2}\log(6\log n)}{\log n} = \alpha \quad \text{and} \quad \lim_{n\to\infty} \frac{\log K_n}{\log N_n} = \beta$$

This completes the proof in the case that $\beta \geq 1/2$. Now consider the case where $\beta < 1/2$. Set $\rho_n = 1/2$ and all other parameters as above. Let $G_n \sim G(n, k_n, S)$ and repeat the same argument as above to obtain that $\mathbb{P}[\phi_n^r \circ \varphi_n(G_n) = S] \geq \epsilon - o(1)$. Now consider the algorithm $\phi_n' : \mathcal{G}_n \to \{0,1\}$ that computes $S' = \phi_n^r \circ \varphi_n(G_n)$ and checks if S' is a clique, outputting a 1 if it is and 0 or 1 uniformly at random otherwise. If $G_n \sim G(n, 1/2)$, then with probability $1 - o(1)$ the largest clique of G_n is less than $(2 + \epsilon)\log_2 n$ for any fixed $\epsilon > 0$. It follows by the definition of ϕ_n that $|S'| = k_n = \Theta(n^\beta) = \omega(\log n)$ and thus with probability $1 - o(1)$, ϕ_n' outputs a random bit. Therefore $\mathbb{P}_{G_n \sim G(n,1/2)}[\phi_n'(G_n) = 1] = 1/2 + o(1)$. If $G_n \sim G(n, k, 1/2)$, then with probability at least $\epsilon - o(1)$, S' is the support of the planted clique and ϕ_n' outputs a 1. Otherwise, ϕ_n' outputs a random bit. Therefore $\mathbb{P}_{G_n \sim G(n,k,1/2)}[\phi_n'(G_n) = 0] = (1 - \epsilon)/2 + o(1)$. Therefore it follows that the Type I+II error of ϕ_n' is

$$\mathbb{P}_{G_n \sim G(n,1/2)}[\phi_n'(G_n) = 1] + \mathbb{P}_{G_n \sim G(n,k,1/2)}[\phi_n'(G_n) = 0] = 1 - \frac{\epsilon}{2} + o(1)$$

which contradicts the PC conjecture. This completes the proof of the theorem. $\qquad\square$

Since the reduction BC-RECOVERY exactly preserves the latent support S of the instance of PDS_R when mapping to BC_R, the same reduction shows hardness of partial recovery if the PDS conjecture is strengthened to hold for partial recovery. The same is true for weak recovery.

6.3 Lower Bounds for General PDS

In this section we give a reduction to the general regime of PDS where $q = \tilde{\Theta}(n^{-\alpha})$ and $p - q = \tilde{\Theta}(n^{-\beta})$ where $\beta > \alpha$. Note that in order to completely characterize PDS when $p - q = O(q)$, we also need the computational lower bound shown in Section 6.1 when $\alpha = \beta$. We now give this reduction, which applies GAUSSIAN-LIFTING and POISSON-LIFTING in sequence. Its correctness follows from combining the guarantees of rejection kernels, GAUSSIAN-LIFTING and POISSON-LIFTING with the data processing and triangle inequalities. The proof is deferred to Appendix C.

Lemma 6.9. *Fix some arbitrary $\epsilon \in (0,1)$. Suppose that n, ℓ_1 and ℓ_2 are such that $\ell_1, \ell_2 = O(\log n)$ and are sufficiently large. Let $\ell = \ell_1 + \ell_2$ and*

$$\mu = \frac{\log 2}{2\sqrt{6\log n + 2\log 2}}$$

Then there is a randomized polynomial time computable map $\phi = \text{GENERAL-PDS-REDUCTION}$ with $\phi : \mathcal{G}_n \to \mathcal{G}_{2^\ell n}$ such that under both H_0 and H_1, it holds that

$$d_{TV}\left(\phi(\text{PC}(n, k, 1/2)), \text{PDS}\left(2^\ell n, 2^\ell k, p_{\ell_1, \ell_2}, q_{\ell_1, \ell_2}\right)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

where p_{ℓ_1, ℓ_2} and q_{ℓ_1, ℓ_2} are defined to be

$$p_{\ell_1, \ell_2} = 1 - \exp\left(4^{-\ell_2}\left(2^{\ell_1}n\right)^{-\epsilon} \cdot \left(2\Phi\left(2^{-\ell_1}\mu\right)\right)^{\epsilon/4}\right) \quad \text{and} \quad q_{\ell_1, \ell_2} = 1 - \exp\left(4^{-\ell_2}\left(2^{\ell_1}n\right)^{-\epsilon}\right)$$

Algorithm GENERAL-PDS-REDUCTION

Inputs: Graph $G \in \mathcal{G}_n$, iterations ℓ_1, ℓ_2

1. Let H be the output of GAUSSIAN-LIFTING applied to G with ℓ_1 iterations

2. Update H to be the output of POISSON-LIFTING applied to H with ℓ_2 iterations where the rejection kernel is replaced with

$$\text{RK}_{\text{P2}} = \text{RK}\left(\frac{1}{2} + \rho \to \text{Pois}(c\lambda), \frac{1}{2} \to \text{Pois}(\lambda), N\right)$$

where the rejection kernel has natural parameter $2^{\ell_1} n$ and satisfies $\lambda = (2^{\ell_1} n)^{-\epsilon}$, $N = \lceil 6\rho^{-1} \log(2^{\ell_1} n) \rceil$, $\rho = \Phi(2^{-\ell_1}\mu) - 1/2$ and $c = \left(2\Phi\left(2^{-\ell_1}\mu\right)\right)^{\epsilon/4}$

3. Output H

Figure 13: Reduction to the general regime of planted dense subgraph in Lemma 6.9.

We now use this reduction combining GAUSSIAN-LIFTING and POISSON-LIFTING to identify the computational barrier in the general PDS problem. As in previous theorems deducing the computational barrier implied by a reduction, the proof resolves to a calculation we defer to Appendix C.

Theorem 6.10. *Let $\alpha, \gamma \in [0, 2)$ and $\beta \in (0, 1)$ be such that $\gamma \geq \alpha$ and $\beta < \frac{1}{2} + \frac{\gamma}{2} - \frac{\alpha}{4}$. There is a sequence $\{(N_n, K_n, p_n, q_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. *The parameters are in the regime $p - q = \tilde{\Theta}(N^{-\gamma})$, $q = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,*

$$\lim_{n \to \infty} \frac{\log q_n^{-1}}{\log N_n} = \alpha, \quad \lim_{n \to \infty} \frac{\log(p_n - q_n)^{-1}}{\log N_n} = \gamma \quad and \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. *For any sequence of randomized polynomial-time tests $\phi_n : \mathcal{G}_{N_n} \to \{0, 1\}$, the asymptotic Type I+II error of ϕ_n on the problems $\text{PDS}_D(N_n, K_n, p_n, q_n)$ is at least 1 assuming the PC conjecture holds for $p = 1/2$.*

Therefore the computational boundary for $\text{PDS}_D(n, k, p, q)$ in the parameter regime $p - q = \tilde{\Theta}(n^{-\gamma})$, $q = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ where $\gamma \geq \alpha$ is $\beta^* = \frac{1}{2} + \frac{\gamma}{2} - \frac{\alpha}{4}$.

7 Reflection Cloning and Subgraph Stochastic Block Model

7.1 Reflecting Cloning and Rank-1 Submatrix

Suppose that n is even and fixed. Let $\mathcal{R}$ denote the linear operator on $\mathbb{R}^{n \times n}$ matrices that reflects a matrix horizontally about its vertical axis of symmetry. Let $\mathcal{F}$ denote the linear operator that multiplies each entry on the right half of a matrix by -1. The reflection cloning reduction is to

iteratively update the matrix W to

$$W \leftarrow \frac{1}{\sqrt{2}} \left(\mathcal{R} W^\sigma + \mathcal{F} W^\sigma \right)$$

where σ is chosen uniformly at random and then to update W similarly with vertical analogues of $\mathcal{R}$ and $\mathcal{F}$. This achieves the same scaling of μ as GAUSSIAN-LIFTING but does not increase n. This ends up tightly achieving the right parameter scaling to deduce the sharper hardness of ROS_D and, indirectly, SPCA_D over problems that admit sum-tests such as PIS_D, PDS_D, BC_D and BSPCA_D. The parameter scaling in these problems is captured exactly by the cloning methods in the previous two sections. Note that REFLECTION-CLONING causes r' and c' to have negative entries and hence cannot show hardness for BC, unlike GAUSSIAN-LIFTING. We remark that all previous cloning methods are in some sense lossy, introducing independent sources of randomness at each entry of the input matrix. In contrast, the only randomness introduced in REFLECTION-CLONING are random permutations of rows and columns, and ends up achieving a much sharper scaling in μ.

We remark that if $r, c \in \{-1, 0, 1\}$ then r' and c' have most of their entries in $\{-1, 0, 1\}$. Entries outside of $\{-1, 0, 1\}$ result from permutations σ in Step 2a that yield collisions between elements in the support of r and c i.e. when i and $n + 1 - i$ are both in the support of either r or c in an iteration of Step 2. Although few in number, these entries turn out to be information-theoretically detectable. We note that if it were possible to reduce from $r, c \in \{-1, 0, 1\}$ to $r', c' \in \{-1, 0, 1\}$ with this property, then this would strengthen our hardness results for $\mathrm{SSBM}_D, \mathrm{SSW}_D, \mathrm{ROS}_D$ and SPCA_D to hold for the canonical simple vs. simple hypothesis testing formulations of these problems. Given a vector $v \in \mathbb{R}^n$ and permutation σ of $[n]$, let r^σ denote the vector formed by permuting the indices of r according to σ.

Lemma 7.1 (Reflection Cloning). *Suppose n is even and $\ell = O(\log n)$. There is a randomized polynomial-time computable map $\phi = $ REFLECTION-CLONING with $\phi : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ and*

1. *It holds that*
$$\phi \left(N(0,1)^{\otimes n \times n} \right) \sim N(0,1)^{\otimes n \times n}$$

2. *Consider any $\lambda > 0$ and any pair of vectors $r, c \in \mathbb{Z}^n$. Then there is a distribution π over vectors $r', c' \in \mathbb{Z}^n$ with $\|r'\|_2^2 = 2^\ell \|r\|_2^2$ and $\|c'\|_2^2 = 2^\ell \|c\|_2^2$ such that*

$$\phi \left(\lambda \cdot rc^\top + N(0,1)^{\otimes n \times n} \right) \sim \int \mathcal{L} \left(\frac{\lambda}{2^\ell} \cdot r'c'^\top + N(0,1)^{\otimes n \times n} \right) d\pi(r', c')$$

where it holds with probability at least $1 - 4\|r\|_0^{-1} - 4\|c\|_0^{-1}$ over π that

$$2^\ell \|r\|_0 \geq \|r'\|_0 \geq 2^\ell \|r\|_0 \left(1 - \max \left(\frac{2C\ell \cdot \log(2^\ell \|r\|_0)}{\|r\|_0}, \frac{2^\ell \|r\|_0}{n} \right) \right)$$

$$2^\ell \|c\|_0 \geq \|c'\|_0 \geq 2^\ell \|c\|_0 \left(1 - \max \left(\frac{2C\ell \cdot \log(2^\ell \|c\|_0)}{\|c\|_0}, \frac{2^\ell \|c\|_0}{n} \right) \right)$$

for some constant C if $\|r\|_0$ and $\|c\|_0$ are sufficiently large and at most $2^{-\ell-1}n$. Furthermore, if $r = c$ then $r' = c'$ holds almost surely.

Proof. Let $\phi(M)$ be implemented by the procedure REFLECTION-CLONING(M, ℓ) as in Figure 6. If $\ell = O(\log n)$, this algorithm runs in randomized polynomial time. Let $\phi_1(W)$ denote the map that takes the value W prior to an iteration of Step 2 as its input and outputs the value of W after this iteration.

49

Algorithm REFLECTION-CLONING

Inputs: Matrix $M \in \mathbb{R}^{n \times n}$ where n is even, number of iterations ℓ

1. Initialize $W \leftarrow M$

2. For $i = 0, 1, \ldots, \ell - 1$ do:

   a. Generate a permutation σ of $[n]$ uniformly at random

   b. Let $W' \in \mathbb{R}^{n \times n}$ have entries

   $$W'_{ij} = \frac{1}{2}\left(W^{\sigma,\sigma}_{ij} + W^{\sigma,\sigma}_{(n+1-i)j} + W^{\sigma,\sigma}_{(n+1-i)(n+1-j)} + W^{\sigma,\sigma}_{(n+1-i)(n+1-j)}\right)$$

   $$W'_{(n+1-i)j} = \frac{1}{2}\left(W^{\sigma,\sigma}_{ij} - W^{\sigma,\sigma}_{(n+1-i)j} + W^{\sigma,\sigma}_{(n+1-i)(n+1-j)} - W^{\sigma,\sigma}_{(n+1-i)(n+1-j)}\right)$$

   $$W'_{i(n+1-j)} = \frac{1}{2}\left(W^{\sigma,\sigma}_{ij} + W^{\sigma,\sigma}_{(n+1-i)j} - W^{\sigma,\sigma}_{(n+1-i)(n+1-j)} - W^{\sigma,\sigma}_{(n+1-i)(n+1-j)}\right)$$

   $$W'_{(n+1-i)(n+1-j)} = \frac{1}{2}\left(W^{\sigma,\sigma}_{ij} - W^{\sigma,\sigma}_{(n+1-i)j} - W^{\sigma,\sigma}_{(n+1-i)(n+1-j)} + W^{\sigma,\sigma}_{(n+1-i)(n+1-j)}\right)$$

   for each $1 \le i, j \le n/2$

   c. Set $W \leftarrow W'$

3. Output W

Figure 14: Reflection cloning procedure in Lemma 7.1.

If $W \sim N(0,1)^{\otimes n \times n}$, then it follows by a similar argument as in Lemma 7 that $\phi_1(W) \sim N(0,1)^{\otimes n \times n}$. Specifically, for each $1 \le i, j \le n/2$ we have that

$$\begin{bmatrix} W'_{ij} \\ W'_{(n+1-i)j} \\ W'_{i(n+1-j)} \\ W'_{(n+1-i)(n+1-j)} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} W^{\sigma,\sigma}_{ij} \\ W^{\sigma,\sigma}_{(n+1-i)j} \\ W^{\sigma,\sigma}_{i(n+1-j)} \\ W^{\sigma,\sigma}_{(n+1-i)(n+1-j)} \end{bmatrix}$$

where σ is the random permutation generated in Step 2a. It holds that $W^{\sigma,\sigma} \sim N(0,1)^{\otimes n \times n}$ and therefore that the vector of entries on the right hand side above is distributed as $N(0,1)^{\otimes 4}$. Since the coefficient matrix is orthogonal, it follows that the vector on the left hand side is also distributed as $N(0,1)^{\otimes 4}$. Since the σ-algebras $\sigma\{W^{\sigma,\sigma}_{ij}, W^{\sigma,\sigma}_{(n+1-i)j}, W^{\sigma,\sigma}_{i(n+1-j)}, W^{\sigma,\sigma}_{(n+1-i)(n+1-j)}\}$ are independent as (i,j) ranges over $[n/2]^2$, it follows that $W' = \phi_1(W) \sim N(0,1)^{\otimes n \times n}$. Iterating, it now follows that $\phi(N(0,1)^{\otimes n \times n}) \sim N(0,1)^{\otimes n \times n}$, establishing Property 1.

Now consider the case when $W = \lambda \cdot rc^\top + U$ where $U \sim N(0,1)^{\otimes n \times n}$. Note that W' can be expressed in terms of $W^{\sigma,\sigma}$ as

$$W' = \frac{1}{2}(A+B)^\top W^{\sigma,\sigma}(A+B)^\top = \frac{\lambda}{2}(Ar^\sigma + Br^\sigma)(Ac^\sigma + Bc^\sigma)^\top + \frac{1}{2}(A+B)^\top U^{\sigma,\sigma}(A+B)$$

where B is the $n \times n$ matrix with ones on its anti-diagonal and zeros elsewhere, and A is given by

$$A = \begin{bmatrix} I_{n/2} & 0 \\ 0 & -I_{n/2} \end{bmatrix}$$

Note that $\frac{1}{2}(A+B)^\top U^{\sigma,\sigma}(A+B)$ is distributed as $\phi_1(U) \sim N(0,1)^{\otimes n \times n}$. Since $A+B$ is symmetric and satisfies that $(A+B)^2 = 2 \cdot I_n$, we have that

$$\|Ar^\sigma + Br^\sigma\|_2^2 = 2\|r^\sigma\|_2^2 = 2\|r\|_2^2 \quad \text{and} \quad \|Ac^\sigma + Bc^\sigma\|_2^2 = 2\|c^\sigma\|_2^2 = 2\|c\|_2^2$$

Let $r_0 = r$ and $r_{i+1} = Ar^{\sigma_i} + Br^{\sigma_i}$ where σ_i is the permutation generated in the ith iteration of Step 2. It follows by induction that $r_\ell \in \mathbb{Z}^n$ and $\|r_\ell\|_2^2 = 2^\ell\|r\|_2^2$ hold almost surely. Analogously define c_i for each $0 \le i \le \ell$ and note that

$$\phi\left(\lambda \cdot rc^\top + N(0,1)^{\otimes n \times n}\right) \sim \mathcal{L}\left(\frac{\lambda}{2^\ell} \cdot r_\ell c_\ell^\top + N(0,1)^{\otimes n \times n}\right)$$

Furthermore note that if $r_0 = r = c = c_0$, then $r_i = c_i$ for all i holds almost surely. Thus it suffices to show that the desired bounds on $\|r_\ell\|_0$ and $\|c_\ell\|_0$ hold with high probability in order to establish Property 2.

Note that since $A + B$ has two nonzero entries per row and column, it follows that $2\|r_i\|_0 \ge \|r_{i+1}\|_0$ for all i. Now consider the collision set

$$S_i = \{\{j, n+1-j\} : j, n+1-j \in \text{supp}(r_i^\sigma)\}$$

Note that if $j \in \text{supp}(r_i^\sigma)$, then j is only not in the support of r_{i+1} if it is in some unordered pair in S_i. Also note that $(r_{i+1})_j + (r_{i+1})_{n+1-j} = 2r_j^\sigma \ne 0$ if j is in some unordered pair in S_i. Therefore at most one element per unordered pair of S_i can be absent from the support of r_{i+1}. This implies that $2\|r_i\|_0 - \|r_{i+1}\|_0 \le |S_i|$ for each i. For each $1 \le j \le n/2$, let X_j be the indicator for the event that $\{j, n+1-j\} \in S_i$. Let $t = \|r_i\|_0$ and note that $|S_i| = X_1 + X_2 + \cdots + X_{n/2}$. For any subset $T \subseteq [n/2]$, it follows that if $|T| \le n/2$ then

$$\mathbb{E}\left[\prod_{j \in T} X_j\right] = \frac{t(t-1)\cdots(t-2|T|+1)}{n(n-1)\cdots(n-2|T|+1)} \le \left(\frac{t}{n}\right)^{2|T|}$$

and if $|T| > n/2$, then this expectation is zero. Let $Y \sim \text{Bin}(n/2, t^2/n^2)$ and note that the above inequality implies that $\mathbb{E}[|S_i|^k] \le \mathbb{E}[Y^k]$ for all $j \ge 0$. This implies that if $\theta \ge 0$, then

$$\mathbb{E}[\exp(\theta|S_i|)] \le \mathbb{E}[\exp(\theta Y)] = \left(1 + (e^\theta - 1) \cdot \frac{t^2}{n^2}\right)^{n/2} \le \exp\left((e^\theta - 1) \cdot \frac{t^2}{2n}\right)$$

A Chernoff bound now yields that

$$\mathbb{P}[|S_i| \ge k] \le \exp\left((e^\theta - 1) \cdot \frac{t^2}{2n} - \theta k\right)$$

Setting $k = t^2/n$ and $\theta = \ln 2$ yields that

$$\mathbb{P}\left[|S_i| \ge \frac{t^2}{n}\right] \le \left(\frac{e}{4}\right)^{\frac{t^2}{2n}} \le \frac{1}{t}$$

51

if $t^2/2n \geq \log_{4/e} t$. If $t^2/2n < \log_{4/e} t$, setting $\theta = \ln 2$ and $k = \frac{1}{\ln 2}\left(\log_{4/e} t + \ln t\right) = C \log t$ yields

$$\mathbb{P}\left[|S_i| \geq C \log t\right] \leq \exp\left(\frac{t^2}{2n} - (\ln 2)k\right) = \frac{1}{t}$$

Therefore with probability at least $1 - 1/\|r_i\|_0$, it follows that $|S_i| < \max(C \log \|r_i\|_0, \|r_i\|_0^2/n)$. Note that this inequality implies that

$$\|r_{i+1}\|_0 \geq 2\|r_i\|_0 - |S_i| \geq 2\|r_i\|_0\left(1 - \max\left(\frac{C \log \|r_i\|_0}{\|r_i\|_0}, \frac{\|r_i\|_0}{n}\right)\right)$$

We now will show by induction on $0 \leq j \leq \ell$ that as long as $\|r\|_0$ is sufficiently large, we have

$$\|r_j\|_0 \geq 2^j \|r\|_0\left(1 - \max\left(\frac{2Cj \cdot \log(2^j \|r\|_0)}{\|r\|_0}, \frac{2^j \|r\|_0}{n}\right)\right) \tag{1}$$

holds with probability at least

$$1 - \frac{4(1 - 2^{-j})}{\|r\|_0}$$

The claim is vacuously true when $j = 0$. Now assume that this holds for a particular j. Now note that since $j \leq \ell$ and $\|r\|_0$ is sufficiently large but at most $2^{-\ell-1}n$, we have that

$$\max\left(\frac{2Cj \cdot \log(2^j \|r\|_0)}{\|r\|_0}, \frac{2^j \|r\|_0}{n}\right) \leq \frac{1}{2}$$

Therefore $\|r_j\|_0 \geq 2^{j-1}\|r\|_0$. We now condition on the value of $\|r_j\|_0$ and the event (7.1) holds. Under this conditioning, it follows by the induction hypothesis that with probability at least $1 - 1/\|r_j\|_0 \geq 1 - 2^{1-j}\|r\|_0^{-1}$, we have that

$$\|r_{j+1}\|_0 \geq 2^{j+1}\|r\|_0\left(1 - \max\left(\frac{2Cj \cdot \log(2^j \|r\|_0)}{\|r\|_0}, \frac{2^j \|r\|_0}{n}\right)\right)\left(1 - \max\left(\frac{C \log \|r_j\|_0}{\|r_j\|_0}, \frac{\|r_j\|_0}{n}\right)\right)$$

$$2^{j+1}\|r\|_0\left(1 - \max\left(\frac{2C(j+1) \cdot \log(2^{j+1}\|r\|_0)}{\|r\|_0}, \frac{2^{j+1}\|r\|_0}{n}\right)\right)$$

since $2^{j-1}\|r\|_0 \leq \|r_j\| \leq 2^j \|r\|_0$. Marginalizing over $\|r_j\|_0$ and whether or not (7.1) holds yields that the above inequality holds with probability at least

$$\left(1 - \frac{4(1 - 2^{-j})}{\|r\|_0}\right)\left(1 - \frac{2^{1-j}}{\|r\|_0^{-1}}\right) \geq 1 - \frac{4(1 - 2^{-j})}{\|r\|_0} - \frac{2^{1-j}}{\|r\|_0} = 1 - \frac{4(1 - 2^{-j-1})}{\|r\|_0}$$

which completes the induction. Obtaining symmetric results for c and taking $j = \ell$ completes the proof of the lemma. $\qquad\square$

Combining this reflection cloning procedure with the reduction from planted clique to biclustering yields a reduction from planted clique to rank-1 submatrix. The proof of the next lemma and theorem are deferred to Appendix D.

Lemma 7.2. *Suppose that n is even and $2^\ell k < \frac{n}{\log k}$ where n is sufficiently large. Let*

$$\mu = \frac{\log 2}{2\sqrt{6 \log n} + 2 \log 2}$$

52

Algorithm ROS-REDUCTION

Inputs: Graph $G \in \mathcal{G}_n$, number of iterations ℓ

1. Compute the output W of BC-REDUCTION applied to G with zero iterations

2. Return the output of REFLECTION-CLONING applied to W with ℓ iterations

---

Figure 15: Reduction to rank-1 submatrix in Lemma 7.2.

There is a randomized polynomial time computable map $\phi = $ ROS-REDUCTION with $\phi : \mathcal{G}_n \to \mathbb{R}^{n \times n}$ such that if G is an instance of $PC(n, k, 1/2)$ then under H_0, it holds that

$$d_{TV}\left(\mathcal{L}_{H_0}(\phi(G)), N(0,1)^{\otimes n \times n}\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

and, under H_1, there is a prior π on pairs of unit vectors in $\mathcal{V}_{n,2^\ell k}$ such that

$$d_{TV}\left(\mathcal{L}_{H_1}(\phi(G)), \int \mathcal{L}\left(\frac{\mu k}{\sqrt{2}} \cdot uv^\top + N(0,1)^{\otimes n \times n}\right) d\pi(u,v)\right) = O\left(\frac{1}{\sqrt{\log n}} + k^{-1}\right)$$

This lemma provides a polynomial time map from an instance of $PC_D(n, k, 1/2)$ to $N(0,1)^{\otimes n \times n}$ under H_0 and to a distribution in the composite hypothesis H_1 of $ROS_D(n, 2^\ell k, 2^{-1/2}\mu)$ under H_1. Now we deduce the hard regime of ROS_D given the planted clique conjecture as in the next theorem. Here, we consider the asymptotic regime $\frac{\mu}{k} = \tilde{\Theta}(n^{-\alpha})$ to be consistent with Figure 3. The purpose of this parameterization is to focus on the factor $\frac{\mu}{k}$ required to normalize entries in the planted submatrix to have magnitude approximately 1. This enables a valid comparison between the hardness of ROS_D and BC_D.

Theorem 7.3. *Let $\alpha > 0$ and $\beta \in (0,1)$ be such that $\beta < \frac{1}{2} + \alpha$. There is a sequence $\{(N_n, K_n, \mu_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. *The parameters are in the regime $\frac{\mu}{K} = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,*

$$\lim_{n \to \infty} \frac{\log(K_n \mu_n^{-1})}{\log N_n} = \alpha \quad and \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. *For any sequence of randomized polynomial-time tests $\phi_n : \mathcal{G}_{N_n} \to \{0,1\}$, the asymptotic Type I+II error of ϕ_n on the problems $ROS_D(N_n, K_n, \mu_n)$ is at least 1 assuming the PC conjecture holds with density $p = 1/2$.*

Therefore the computational boundary for $ROS_D(n, k, \mu)$ in the parameter regime $\frac{\mu}{k} = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\beta^* = \frac{1}{2} + \alpha$ and $\alpha^* = 0$ when $\beta < \frac{1}{2}$.

7.2 Sparse Spiked Wigner Matrix

We now show a computational lower bound for sparse spiked Wigner matrix detection given the planted clique conjecture. As observed in Section 2, it suffices to reduce from planted clique to

---

Algorithm SROS-REDUCTION

Inputs: Planted clique instance $G \in \mathcal{G}_n$ with clique size k where n is even, iterations ℓ

1. Let $\text{RK}_G = \text{RK}\left(1 \to N(\mu, 1), 1/2 \to N(0, 1), N\right)$ where $N = \lceil 6 \log_2 n \rceil$ and $\mu = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$ and form the symmetric matrix $W \in \mathbb{R}^{n \times n}$ with $W_{ii} = 0$ and for all $i < j$,

$$W_{ij} = \text{RK}_G\left(\mathbf{1}_{\{i,j\} \in E(G)}\right)$$

2. Sample an antisymmetric matrix $A \in \mathbb{R}^{n \times n}$ with i.i.d. $N(0, 1)$ entries below its main diagonal and sample two matrices $B, C \in \mathbb{R}^{n \times n}$ with i.i.d. $N(0, 1)$ off-diagonal and zero diagonal entries

3. Form the matrix $M \in \mathbb{R}^{n \times n}$ with

$$M_{ij} = \frac{k-1}{2\sqrt{n-1}}\left(W_{ij} + A_{ij} + B_{ij} \cdot \sqrt{2}\right) + C_{ij} \cdot \sqrt{1 - \frac{(k-1)^2}{n-1}}$$

for all $i \neq j$ and

$$M_{ii} = \frac{1}{2\sqrt{n-1}} \sum_{j=1}^{n}\left(W_{ij} + A_{ij} - B_{ij} \cdot \sqrt{2}\right)$$

4. Update M to be the output of REFLECTION-CLONING applied with ℓ iterations to M

5. Output $M^{\sigma,\sigma}$ where σ is a permutation of $[n]$ chosen uniformly at random

---

Figure 16: Reduction to sparse spiked Wigner matrix in Lemma 7.4.

SROS_D. This is because if M is an instance of $\text{SROS}_D(n, k, \mu)$, then $\frac{1}{\sqrt{2}}(M + M^\top)$ is an instance of $\text{SSW}_D(n, k, \mu/\sqrt{2})$. This transformation implies that any computational lower bound that applies to $\text{SROS}_D(n, k, \mu)$ also applies to $\text{SSW}_D(n, k, \mu/\sqrt{2})$. The rest of this section is devoted to giving a reduction to $\text{SROS}_D(n, k, \mu)$.

Our reduction uses the symmetry preserving property of reflection cloning and yields the same computational barrier as for ROS_D. However, there are several subtle differences between this reduction and that in Lemma 7.2. In order to show hardness for SROS_D, it is important to preserve the symmetry of the planted sparse structure. This requires planting the hidden entries along the diagonal of the adjacency matrix of the input graph G, which we do by an averaging trick and generating additional randomness to introduce independence. However, this induces an arbitrarily small polynomial loss in the size of the spike, unlike in the reduction to ROS_D. Although this does not affect our main theorem statement for SROS_D, which only considers poly(n) size factors, it yields a weaker lower bound than that in Theorem 7.3 when examined up to sub-polynomial factors. This reduction to SROS_D will also serve as the main sub-routine in our reduction to SSBM_D.

Lemma 7.4. *Suppose that n is even and $2^\ell k < \frac{n}{\log k}$ where n is sufficiently large. Let*

$$\mu = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$$

There is a randomized polynomial time computable map $\phi = \text{SROS-REDUCTION}$ with $\phi : \mathcal{G}_n \to \mathbb{R}^{n \times n}$ such that if G is an instance of $PC(n, k, 1/2)$ then under H_0, it holds that

$$d_{TV}\left(\mathcal{L}_{H_0}(\phi(G)), N(0,1)^{\otimes n \times n}\right) = O(n^{-1})$$

and, under H_1, there is a prior π on unit vectors in $\mathcal{V}_{n, 2^\ell k}$ such that

$$d_{TV}\left(\mathcal{L}_{H_1}(\phi(G)), \int \mathcal{L}\left(\frac{\mu k(k-1)}{2\sqrt{(n-1)}} \cdot vv^\top + N(0,1)^{\otimes n \times n}\right) d\pi(v)\right) = O\left(n^{-1}\right)$$

Proof. Let $\phi = \text{SROS-REDUCTION}$ be as in Figure 16. Applying the total variation bounds in Lemma 5.4 and entry-wise coupling as in Lemma 6.7 yields that

$$d_{\text{TV}}\left(\mathcal{L}_{H_0}(W), M_n(N(0,1))\right) = O(n^{-1}) \quad \text{and} \quad d_{\text{TV}}\left(\mathcal{L}_{H_1}(W), M_n(k, N(\mu, 1), N(0,1))\right) = O(n^{-1})$$

Let $W' \in \mathbb{R}^{n \times n}$ be such that $W' \sim M_n(N(0,1))$ under H_0 and $W' \sim M_n(k, N(\mu, 1), N(0,1))$ under H_1. Let M' denote the matrix computed in Step 3 using the matrix W' in place of W. We will first argue that under H_0, the variables $W'_{ij} + A_{ij} + B_{ij} \cdot \sqrt{2}$ and $W'_{ij} + A_{ij} - B_{ij} \cdot \sqrt{2}$ for all $i \neq j$ are independent. First note that the σ-algebras $\sigma\{W'_{ij}, A_{ij}, B_{ij}\}$ for all $i < j$ are independent. Therefore it suffices to verify that the four variables $W'_{ij} + A_{ij} \pm B_{ij} \cdot \sqrt{2}$ and $W'_{ji} + A_{ji} \pm B_{ji} \cdot \sqrt{2}$ are independent. Observe that these four variables are jointly Gaussian and satisfy that

$$\frac{1}{2} \cdot \begin{bmatrix} W'_{ij} + A_{ij} + B_{ij} \cdot \sqrt{2} \\ W'_{ij} + A_{ij} - B_{ij} \cdot \sqrt{2} \\ W'_{ji} + A_{ji} + B_{ji} \cdot \sqrt{2} \\ W'_{ji} + A_{ji} - B_{ji} \cdot \sqrt{2} \end{bmatrix} = \frac{1}{2} \cdot \begin{bmatrix} 1 & 1 & \sqrt{2} & 0 \\ 1 & 1 & -\sqrt{2} & 0 \\ 1 & -1 & 0 & \sqrt{2} \\ 1 & -1 & 0 & -\sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} W'_{ij} \\ A_{ij} \\ B_{ij} \\ B_{ji} \end{bmatrix}$$

since W' is symmetric and A is antisymmetric. Observe that $W'_{ij}, A_{ij}, B_{ij}, B_{ji}$ are independent Gaussians and since the coefficient matrix above is orthogonal, it follows that the vector on the left hand side above is distributed as $N(0,1)^{\otimes 4}$ and thus has independent entries. Since C has i.i.d. $N(0,1)$ entries off of its diagonal and W', A, B and C all have zero diagonals, it follows that $\text{Var}(M'_{ij}) = 1$ for all $i, j \in [n]$. Since the entries of M' are independent and each entry is Gaussian with variance 1, it follows that $M' \sim N(0,1)^{\otimes n \times n}$.

Now suppose that H_1 holds and let $S \subseteq [n]$ be indices of the rows and columns containing the planted $N(\mu, 1)$ entries of W'. It now follows that $W'_{ij} = \mu + W''_{ij}$ if $i \neq j$ and $i, j \in S$ and $W'_{ij} = W''_{ij}$ where $W'' \sim M_n(N(0,1))$. Now note that if $i \neq j$, we have that conditioned on S,

$$M'_{ij} = \frac{(k-1)\mu}{2\sqrt{n-1}} \cdot \mathbf{1}_{\{i,j \in S\}} + \frac{k-1}{2\sqrt{n-1}}\left(W''_{ij} + A_{ij} + B_{ij} \cdot \sqrt{2}\right) + C_{ij} \cdot \sqrt{1 - \frac{(k-1)^2}{n-1}}$$

$$\sim \frac{(k-1)\mu}{2\sqrt{n-1}} \cdot \mathbf{1}_{\{i,j \in S\}} + N(0,1)$$

by applying the previous argument to W''. Furthermore M'_{ii} has diagonal entries

$$M'_{ij} = \frac{(k-1)\mu}{2\sqrt{n-1}} \cdot \mathbf{1}_{\{i,j \in S\}} + \frac{1}{2\sqrt{n-1}} \sum_{j=1}^{n}\left(W''_{ij} + A_{ij} - B_{ij} \cdot \sqrt{2}\right)$$

$$\sim \frac{(k-1)\mu}{2\sqrt{n-1}} \cdot \mathbf{1}_{\{i \in S\}} + N(0,1)$$

55

conditioned on S. Therefore $M'|S \sim \frac{(k-1)\mu}{2\sqrt{n-1}} \cdot \mathbf{1}_S\mathbf{1}_S^\top + N(0,1)^{\otimes n \times n}$. Now by the data processing inequality, we now have that under H_0,

$$d_{\mathrm{TV}}\left(\mathcal{L}_{H_0}(M), N(0,1)^{\otimes n \times n}\right) \leq d_{\mathrm{TV}}\left(\mathcal{L}_{H_0}(W), M_n(N(0,1))\right) = O(n^{-1})$$

By the data processing and triangle inequalities, we have that

$$d_{\mathrm{TV}}\left(\mathcal{L}_{H_1}(M), \int \mathcal{L}\left(\frac{(k-1)\mu}{2\sqrt{(n-1)}} \cdot \mathbf{1}_S\mathbf{1}_S^\top + N(0,1)^{\otimes n \times n}\right) d\pi'(S)\right)$$

$$\leq \mathbb{E}_S d_{\mathrm{TV}}\left(\mathcal{L}_{H_1}(M), \frac{(k-1)\mu}{2\sqrt{(n-1)}} \cdot \mathbf{1}_S\mathbf{1}_S^\top + N(0,1)^{\otimes n \times n}\right) = O(n^{-1})$$

where π' is the uniform distribution on k-subsets of $[n]$. Now applying the same argument as in the proof of Lemma 7.2 and the fact that REFLECTION-CLONING preserves the fact that the rank-1 mean submatrix is symmetric as shown in Lemma 7.1, proves the lemma. □

We now use this lemma to deduce the computational barriers for SROS_D and SSW_D. Although the barrier matches Theorem 7.3, the parameter settings needed to achieve it are slightly different due to the polynomial factor loss in the reduction in Lemma 7.4. The proof is deferred to Appendix D.

Theorem 7.5. *Let $\alpha > 0$ and $\beta \in (0,1)$ be such that $\beta < \frac{1}{2} + \alpha$. There is a sequence $\{(N_n, K_n, \mu_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. The parameters are in the regime $\frac{\mu}{K} = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,

$$\lim_{n \to \infty} \frac{\log(K_n \mu_n^{-1})}{\log N_n} = \alpha \quad and \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. For any sequence of randomized polynomial-time tests $\phi_n : \mathcal{G}_{N_n} \to \{0,1\}$, the asymptotic Type I+II error of ϕ_n on the problems $\mathrm{SROS}_D(N_n, K_n, \mu_n)$ and $\mathrm{SSW}_D(N_n, K_n, \mu_n/\sqrt{2})$ is at least 1 assuming the PC conjecture holds with density $p = 1/2$.

Therefore the computational boundaries for $\mathrm{SROS}_D(n, k, \mu)$ and $\mathrm{SSW}_D(n, k, \mu)$ in the parameter regime $\frac{\mu}{k} = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\beta^* = \frac{1}{2} + \alpha$ and $\alpha^* = 0$ when $\beta < \frac{1}{2}$.

7.3 Subgraph Stochastic Block Model

In this section, we deduce tight hardness for the subgraph stochastic block model from the reflection cloning reduction from planted clique to SROS_D. This captures the sharper hardness of detection in the subgraph stochastic block model over planted dense subgraph and the lack of a sum test. Note that here total variation distance under H_1 refers to the total variation distance to some prior over the distributions in H_1.

Lemma 7.6. *Suppose that n is even and $\delta \in (0, 1/2)$ is such that $(2^\ell k)^{1+2\delta} = O(n)$ and $k = \Omega\left(n^\delta\right)$. Let $\mu > 0$ be such that*

$$\mu = \frac{\log 2}{2\sqrt{6\log n + 2\log 2}}$$

There is a polynomial time map $\phi = $ SSBM-REDUCTION with $\phi : \mathcal{G}_n \to \mathcal{G}_n$ such that for some mixture $\mathcal{L}_{\mathrm{SSBM}}$ of distributions in $G_B\left(n, 2^\ell k, 1/2, \rho\right)$, it holds that

$$d_{TV}(\phi(G(n, 1/2)), G(n, 1/2)) = O\left(n^{-1}\right) \quad and \quad d_{TV}(\phi(G(n, k, 1/2)), \mathcal{L}_{\mathrm{SSBM}}) = O\left(k^{-1}\right)$$

Algorithm SSBM-REDUCTION

Inputs: Planted clique instance $G \in \mathcal{G}_n$ with clique size k where n is even, iterations ℓ

1. Compute the output M of SROS-REDUCTION applied to G with ℓ iterations

2. Sample n i.i.d. Rademacher random variables $x_1, x_2, \ldots, x_n$ and update each entry of M to be $M_{ij} \leftarrow x_i x_j M_{ij}$

3. Output the graph H where $\{i, j\} \in E(H)$ if and only if $M_{ij} > 0$ for each $i < j$

Figure 17: Reduction to the subgraph stochastic block model in Lemma 7.6.

where ρ is given by

$$\rho = \Phi\left(\frac{\mu(k-1)}{2^{\ell+1}\sqrt{n-1}}\right) - \frac{1}{2}$$

Proof. Given a vector $v \in \mathbb{Z}^n$ and $\theta > 0$, let $M(\theta, v) \in \mathbb{R}^{n \times n}$ be distributed as $\theta \cdot vv^\top + N(0,1)^{\otimes n \times n}$. Let $S \subseteq [n]$ and $P \in [0,1]^{|S| \times |S|}$ be a symmetric matrix with zero diagonal entries and rows and columns indexed by vertices in S. Let $G(n, S, P, q)$ be the distribution on graphs G generated as follows:

1. if $i \neq j$ and $i, j \in S$ then the edge $\{i, j\} \in E(G)$ independently with probability P_{ij}; and

2. all other edges of G are included independently with probability q.

Now let $\tau : \mathbb{R}^{n \times n} \to \mathcal{G}_n$ be the map sending a matrix M to the graph G such that

$$E(G) = \{\{i, j\} : M_{ij} > 0 \text{ and } i < j\}$$

In other words, G is the graph formed by thresholding the entries of M below its main diagonal at zero. Suppose that $v \in \mathbb{Z}^n$ is k-sparse and $S = \mathrm{supp}(v)$. Since the entries of $M(\theta, v)$ are independent, it follows that

$$\tau\left(M(\theta, v)\right) \sim G(n, S, P, q) \quad \text{where} \quad P_{ij} = \Phi\left(\theta \cdot v_i v_j\right) \text{ for each } i, j \in S$$

Now let $S = A \cup B$ where $A = A(v) = \{i \in S : v_i > 0\}$ and $B = B(v) = \{i \in S : v_i < 0\}$. Note that since $v \in \mathbb{Z}^n$, it follows that if $i, j \in A$ or $i, j \in B$ then $v_i v_j \geq 1$ and thus $P_{ij} = \Phi\left(\theta \cdot v_i v_j\right) \geq \Phi(\theta)$. Furthermore if $(i, j) \in A \times B$ or $(i, j) \in B \times A$ then $v_i v_j \leq -1$ and $P_{ij} = \Phi\left(\theta \cdot v_i v_j\right) \leq 1 - \Phi(\theta)$. Now note that if σ is a permutation of $[n]$ chosen uniformly at random then

$$\mathcal{L}\left(\tau\left(M(\theta, v)\right)^\sigma\right) = \mathcal{L}\left(\tau\left(M(\theta, v)^{\sigma, \sigma}\right)\right) \in G_B\left(n, k, 1/2, \Phi(\theta) - 1/2\right)$$

if it also holds that $\frac{k}{2} - k^{1-\delta} \leq |A|, |B| \leq \frac{k}{2} + k^{1-\delta}$, by the definition of G_B in Section 2.2.

Let $\phi_1 = $ SROS-REDUCTION and let $W \sim N(0,1)^{\otimes n \times n}$. As shown in Lemma 7.4,

$$d_{\mathrm{TV}}\left(\phi_1(G(n, 1/2)), \mathcal{L}(W)\right) = O(n^{-1})$$

Now observe that since $N(0, 1)$ is symmetric and since the entries of W are independent, the distribution of W is invariant to flipping the signs of any subset of the entries of W. Therefore

$xx^\top \circ W \sim N(0,1)$ where $\circ$ denotes the entry-wise or Schur product on $\mathbb{R}^{n\times n}$ and $x \in \{-1,1\}^n$ is chosen uniformly at random. Therefore the data processing inequality implies that

$$d_{\mathrm{TV}}\left(\mathcal{L}\left(xx^\top \circ \phi_1(G(n,1/2))\right), \mathcal{L}(W)\right) = d_{\mathrm{TV}}\left(\mathcal{L}\left(xx^\top \circ \phi_1(G(n,1/2))\right), \mathcal{L}\left(xx^\top \circ W\right)\right)$$
$$\leq d_{\mathrm{TV}}\left(\phi_1(G(n,1/2)), \mathcal{L}(W)\right) = O(n^{-1})$$

Recalling that ϕ is the output of SSBM-REDUCTION, note that $\phi(G(n,1/2))$ is distributed as $\tau\left(xx^\top \circ \phi_1(G(n,1/2))\right)$ and that $\tau\left(N(0,1)^{\otimes n\times n}\right)$ is distributed as $G(n,1/2)$. It follows by the data processing inequality that

$$d_{\mathrm{TV}}\left(\phi(G(n,1/2)), G(n,1/2)\right) = d_{\mathrm{TV}}\left(\mathcal{L}\left(\tau\left(xx^\top \circ \phi_1(G(n,1/2))\right)\right), \mathcal{L}(\tau(W))\right) = O(n^{-1})$$

Repeating the analysis in Lemmas 7.2 and 7.4 without converting the prior in Lemma 7.1 to be over unit vectors yields that there is a prior π on vectors in $u \in \mathbb{Z}^n$ such that

$$d_{\mathrm{TV}}\left(\phi_1(G(n,k,1/2)), \int \mathcal{L}\left(\frac{\mu(k-1)}{2^{\ell+1}\sqrt{n-1}} \cdot uu^\top + N(0,1)^{\otimes n\times n}\right) d\pi(u)\right) = O\left(n^{-1}\right)$$

and such that with probability at least $1 - 8k^{-1}$ it holds that

$$2^\ell k \geq \|u\|_0 \geq 2^\ell k \left(1 - \max\left(\frac{2C\ell \cdot \log(2^\ell k)}{k}, \frac{2^\ell k}{n}\right)\right) = 2^\ell k - O\left(2^\ell(\log n)^2 + (2^\ell k)^{1-2\delta}\right)$$

where $C > 0$ is the constant in Lemma 7.1. Now let π' be the prior π conditioned on the inequality in the last displayed equation. It follows by Lemma 3.4 that $d_{\mathrm{TV}}(\pi, \pi') \leq 8k^{-1}$. Now let $\theta = \frac{\mu(k-1)}{2^{\ell+1}\sqrt{n-1}}$ and let the matrix M' be distributed as

$$M' \sim \int \mathcal{L}\left(\theta \cdot uu^\top + N(0,1)^{\otimes n\times n}\right) d\pi(u)$$

As above, let $x \in \{-1,1\}^n$ be chosen uniformly at random. The same argument above shows that

$$\mathcal{L}\left(xx^\top \circ M' \,\middle|\, x \circ u = v\right) = \mathcal{L}\left(\theta \cdot vv^\top + N(0,1)^{\otimes n\times n}\right) = \mathcal{L}(M(\theta,v))$$

As shown above, this implies that

$$\mathcal{L}\left(\tau\left(xx^\top \circ M'\right) \,\middle|\, x \circ u = v\right) \in G_B\left(n, 2^\ell k, 1/2, \Phi(\theta) - 1/2\right)$$

as long as $2^{\ell-1}k - (2^\ell k)^{1-\delta} \leq |A(v)|, |B(v)| \leq 2^{\ell-1}k + (2^\ell k)^{1-\delta}$. Let $\pi''(x,u)$ be the product distribution of $\mathcal{L}(x)$ and π' conditioned on the event that these inequalities hold for $v = x \circ u$. Now note that conditioning on u yields that $|A(x \circ u)| + |B(x \circ u)| = \|u\|_0$ and $|A(x \circ u)|$ is distributed as $\mathrm{Bin}(\|u\|_0, 1/2)$. By Hoeffding's inequality, we have that conditioned on u,

$$\mathbb{P}\left[\left||A(x \circ u)| - \frac{1}{2}\|u\|_0\right| > \sqrt{\|u\|_0 \log k}\right] \leq \frac{2}{k^2}$$

Note that if this inequality holds for $|A(x \circ u)|$, then it also holds for $|B(x \circ u)|$ since these sum to $\|u\|_0$. Therefore with probability at least $1 - 2k^{-2}$, it holds that

$$\left||A(x \circ u)| - 2^{\ell-1}k\right| \leq \left||A(x \circ u)| - \frac{1}{2}\|u\|_0\right| + O\left(2^\ell(\log n)^2 + (2^\ell k)^{1-2\delta}\right)$$
$$= O\left(\sqrt{2^\ell k \log k} + 2^\ell(\log n)^2 + (2^\ell k)^{1-2\delta}\right) \leq (2^\ell k)^{1-\delta}$$

58

for sufficiently large k and the same inequalities hold for $|B(x \circ u)|$. This verifies that the desired inequalities on $|A(x \circ u)|$ and $|B(x \circ u)|$ hold with probability at least $1 - 2k^{-2}$ over $\mathcal{L}(x) \otimes \pi(u)$. Applying Lemma 3.4 therefore yields that $d_{\text{TV}}(\mathcal{L}(x) \times \pi'(u), \pi''(x, u)) \leq 2k^{-2}$. The data processing and triangle inequalities then imply that

$$d_{\text{TV}}\left(\mathcal{L}(x) \otimes \pi(u), \pi''(x, u)\right) \leq d_{\text{TV}}\left(\mathcal{L}(x) \otimes \pi(u), \mathcal{L}(x) \otimes \pi'(u)\right) + d_{\text{TV}}\left(\mathcal{L}(x) \otimes \pi'(u), \pi''(x, u)\right)$$
$$\leq d_{\text{TV}}(\pi, \pi') + 2k^{-2} = O(k^{-1})$$

Now observe that

$$d_{\text{TV}}\left(\phi(G(n, k, 1/2), \int \mathcal{L}\left(\tau\left(xx^\top \circ M'\right)\Big|x, u\right) d\pi''(x, u)\right)$$

$$\leq d_{\text{TV}}\left(\mathcal{L}\left(\tau\left(xx^\top \circ \phi_1(G(n, k, 1/2))\right)\right), \int \mathcal{L}\left(\tau\left(xx^\top \circ M'\right)\Big|x, u\right) d\mathcal{L}(x) d\pi(u)\right)$$

$$+ d_{\text{TV}}\left(\int \mathcal{L}\left(\tau\left(xx^\top \circ M'\right)\Big|x, u\right) d\mathcal{L}(x) d\pi(u), \int \mathcal{L}\left(\tau\left(xx^\top \circ M'\right)\Big|x, u\right) d\pi''(x, u)\right)$$

$$\leq d_{\text{TV}}\left(\phi_1(G(n, k, 1/2), \mathcal{L}(M'))\right) + d_{\text{TV}}\left(\mathcal{L}(x) \otimes \pi(u), \pi''(x, u)\right)$$

$$= O(n^{-1}) + O(k^{-1}) = O(k^{-1})$$

By the definition of π'', the distribution

$$\mathcal{L}_{\text{SSBM}} = \int \mathcal{L}\left(\tau\left(xx^\top \circ M'\right)\Big|x, u\right) d\pi''(x, u)$$

is a mixture of distributions in $G_B\left(n, 2^\ell k, 1/2, \Phi(\theta) - 1/2\right)$, completing the proof of the lemma. $\qquad \square$

Applying this reduction and setting parameters similarly to Theorem 7.5 yields the following computational lower bound for the subgraph stochastic block model. The proof is a calculation deferred to Appendix D.

Theorem 7.7. *Let $\alpha \in [0, 2)$ and $\beta \in (\delta, 1 - 3\delta)$ be such that $\beta < \frac{1}{2} + \alpha$. There is a sequence $\{(N_n, K_n, q_n, \rho_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. *The parameters are in the regime $q = \Theta(1)$, $\rho = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,*

$$\lim_{n \to \infty} \frac{\log \rho_n^{-1}}{\log N_n} = \alpha, \qquad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta \quad \text{and} \quad \lim_{n \to \infty} q_n = q$$

2. *For any sequence of randomized polynomial-time tests $\phi_n : \mathcal{G}_{N_n} \to \{0, 1\}$, the asymptotic Type I+II error of ϕ_n on the problems $\text{SSBM}_D(N_n, K_n, q_n, \rho_n)$ is at least 1 assuming the PC conjecture holds with density $p = 1/2$.*

Therefore the computational boundary for $\text{SSBM}_D(n, k, q, \rho)$ in the parameter regime $q = \Theta(1)$, $\rho = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\beta^* = \frac{1}{2} + \alpha$.

8 Random Rotations and Sparse PCA

In this section, we deduce tight lower bounds for detection in sparse PCA when $k \gg \sqrt{n}$. We also show a suboptimal lower bound when $k \ll \sqrt{n}$ matching the results of [BR13a] and [GMZ17]. In both cases, we reduce from biclustering and rank-1 submatrix to biased and unbiased variants of

---

Algorithm RANDOM-ROTATION

Inputs: Matrix $M \in \mathbb{R}^{m \times n}$, parameter $\tau \in \mathbb{N}$

1. Construct the $m \times \tau n$ matrix M' such that the leftmost $m \times n$ submatrix of M' is M and the remaining entries of M' are sampled i.i.d. from $N(0,1)$

2. Let R be the leftmost $\tau n \times n$ submatrix of a random orthogonal matrix sampled from the normalized Haar measure on the orthogonal group $\mathcal{O}_{\tau n}$

3. Output $M'R$

Figure 18: Random rotation procedure in Lemma 8.2.

sparse PCA. The reductions in this section are relatively simple, with most of the work behind the lower bounds we show residing in the reductions from PC to biclustering and rank-1 submatrix. This illustrates the usefulness of using natural problems as intermediates in constructing average-case reductions. The lower bounds we prove make use of the following theorem of Diaconis and Freedman showing that the first m coordinates of a unit vector in n dimensions where $m \ll n$ are close to independent Gaussians in total variation [DF87].

Theorem 8.1 (Diaconis and Freedman [DF87]). *Suppose that $(v_1, v_2, \ldots, v_n)$ is uniformly distributed according to the Haar measure on $\mathbb{S}^{n-1}$. Then for each $1 \le m \le n - 4$,*

$$d_{TV}\left(\mathcal{L}\left(v_1, v_2, \ldots, v_m\right), N(0, n^{-1})^{\otimes m}\right) \le \frac{2(m+3)}{n - m - 3}$$

The next lemma is the crucial ingredient in the reductions of this section. Note that the procedure RANDOM-ROTATION in Figure 19 requires sampling the Haar measure on $\mathcal{O}_{\tau n}$. This can be achieved efficiently by iteratively sampling a Gaussian, projecting it onto the orthogonal complement of the vectors chosen so far, normalizing it and adding it to the current set. Repeating this for n iterations yields an implementation for the sampling part of Step 2 in Figure 19. In the next lemma, we show that RANDOM-ROTATION takes an instance $\lambda \cdot uv^\top + N(0,1)^{\otimes m \times n}$ of rank-1 submatrix to an $m \times n$ matrix with its n columns sampled i.i.d. from $N(0, I_m + \theta vv^\top)$.

Lemma 8.2 (Random Rotation). *Let $\tau : \mathbb{N} \to \mathbb{N}$ be an arbitrary function with $\tau(n) \to \infty$ as $n \to \infty$. Consider the map $\phi : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ that sends M to RANDOM-ROTATION with inputs M and τ. It follows that $\phi(N(0,1)^{\otimes m \times n}) \sim N(0,1)^{\otimes m \times n}$ and for any unit vectors $u \in \mathbb{R}^m, v \in \mathbb{R}^n$ we have that*

$$d_{TV}\left(\phi\left(\lambda \cdot uv^\top + N(0,1)^{\otimes m \times n}\right), N\left(0, I_m + \frac{\lambda^2}{\tau n} \cdot uu^\top\right)^{\otimes n}\right) \le \frac{2(n+3)}{\tau n - n - 3}$$

Proof. Let $R' \in \mathcal{O}_{\tau n}$ be the original $\tau n \times \tau n$ sampled in Step 2 of RANDOM-ROTATION and let R be its upper $\tau n \times n$ submatrix. Let M and M' be the matrices input to RANDOM-ROTATION and computed in Step 1, respectively, as shown in Figure 19. If $M \sim N(0,1)^{\otimes m \times n}$, then it follows that $M' \sim N(0,1)^{\otimes m \times \tau n}$. Since the rows of M' are independent and distributed according to the isotropic distribution $N(0,1)^{\otimes \tau n}$, multiplication on the right by any orthogonal matrix leaves the

60

distribution of M' invariant. Therefore $M'R' \sim N(0,1)^{\otimes m \times \tau n}$ and since $M'R$ consists of the first n columns of $M'R'$, it follows that $\phi(M) = M'R \sim N(0,1)^{\otimes m \times n}$.

Now suppose that M is distributed as $\lambda \cdot uv^\top + N(0,1)^{\otimes m \times n}$ for some unit vectors u, v. Let v' be the unit vector in $\mathbb{R}^{\tau n}$ formed by appending $\tau n - n$ zeros to the end of v. It follows that M' is distributed as $\lambda \cdot uv'^\top + N(0,1)^{\otimes m \times \tau n}$. Let $M' = \lambda \cdot uv'^\top + W$ where $W \sim N(0,1)^{\otimes m \times \tau n}$. Now let $W' = WS^{-1}$ where $S \in \mathcal{O}_{\tau n}$ is sampled according to the Haar measure and independently of R'. Observe that

$$M'R' = \lambda \cdot u \cdot \left(R'^\top v' \right)^\top + W'SR'$$

Now note that conditioned on R', the product SR' is distributed according to the Haar measure on $\mathcal{O}_{\tau n}$. This implies that SR' is independent of R'. Therefore $W'SR'$ is independent of R' and distributed according to $N(0,1)^{\otimes m \times \tau n}$. This also implies that $R'^\top v$ is independent of $W'SR'$ and distributed uniformly over $\mathbb{S}^{\tau n - 1}$. Let $r \in \mathbb{R}^n$ denote the vector consisting of the first n coordinates of $R'^\top v'$ and let W'' denote the $m \times n$ matrix consisting of the first n columns of $W'SR'$. It follows that $\phi(M) = M'R = \lambda \cdot ur^\top + W''$ where r and W'' are independent. Now let $g \in \mathbb{R}^n$ be a Gaussian vector with entries i.i.d. sampled from $N(0, n^{-1})$. Also let $Z = \lambda \cdot ug^\top + W''$ and note that by Diaconis-Freedman's theorem and coupling the noise terms W'', the data processing inequality implies

$$d_{\mathrm{TV}} \left(\mathcal{L} \left(\lambda \cdot ug^\top + W'' \right), \mathcal{L} \left(\lambda \cdot ur^\top + W'' \right) \right) \leq d_{\mathrm{TV}} \left(\mathcal{L}(r), \mathcal{L}(g) \right) \leq \frac{2(n+3)}{\tau n - n - 3}$$

Now note that since the entries of g are independent, the matrix $\lambda \cdot ug^\top + W''$ has independent columns. Its ith row has jointly Gaussian entries with covariance matrix

$$\mathbb{E} \left[(\lambda \cdot ug_i + W_i)(\lambda \cdot ug_i + W_i)^\top \right] = \mathbb{E} \left[\lambda^2 \cdot uu^\top g_i^2 + \lambda \cdot g_i \cdot uW_i^\top + \lambda \cdot g_i \cdot W_i u^\top + W_i W_i^\top \right]$$
$$= \frac{\lambda^2}{\tau n} \cdot uu^\top + I_m$$

Therefore $\lambda \cdot ug^\top + W'' \sim N \left(0, I_m + \frac{\lambda^2}{\tau n} \cdot uu^\top \right)^{\otimes n}$. Combining these results yields that

$$d_{\mathrm{TV}} \left(\phi(M), N \left(0, I_m + \frac{\lambda^2}{\tau n} \cdot uu^\top \right)^{\otimes n} \right) \leq \frac{2(n+3)}{\tau n - n - 3}$$

which completes the proof of the lemma. $\qquad \square$

Applying reflection cloning to produce an instance of rank-1 submatrix and then randomly rotating to obtain an instance of sparse PCA yields a reduction from PC to SPCA as given in SPCA-High-Sparsity. This establishes tight lower bounds in the regime $k \gg \sqrt{n}$. This reduction is stated in the next lemma, which takes an instance of $\mathrm{PC}(n, k, 1/2)$ to an instance of sparse PCA with sparsity $2^\ell k$ and $\theta = \frac{\mu^2 k^2}{2\tau n}$ where τ and μ can be taken to be polylogarithmically small in n. The proof involves a simple application of the data processing and triangle inequalities and is deferred to Appendix E.

Lemma 8.3. *Suppose that n and ℓ are such that $\ell = O(\log n)$ and are sufficiently large,*

$$\mu = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$$

Algorithm SPCA-HIGH-SPARSITY

Inputs: Graph $G \in \mathcal{G}_n$, number of iterations ℓ, function $\tau : \mathbb{N} \to \mathbb{N}$ with $\tau(n) \to \infty$

1. Compute the output M of ROS-REDUCTION applied to G with ℓ iterations

2. Output the matrix returned by RANDOM-ROTATION applied with inputs M and τ

Algorithm SPCA-LOW-SPARSITY

Inputs: Graph $G \in \mathcal{G}_n$, number of iterations ℓ, function $\tau : \mathbb{N} \to \mathbb{N}$ with $\tau(n) \to \infty$

1. Compute the output M of BC-REDUCTION applied to G with ℓ iterations

2. Output the matrix returned by RANDOM-ROTATION applied with inputs M and τ

Algorithm SPCA-RECOVERY

Inputs: Graph $G \in \mathcal{G}_n$, density bias ρ, function $\tau : \mathbb{N} \to \mathbb{N}$ with $\tau(n) \to \infty$

1. Let M be the output of BC-RECOVERY applied to G with density bias ρ

2. Output the matrix returned by RANDOM-ROTATION applied with inputs M and τ

Figure 19: Reductions to SPCA_D when $k \gtrsim \sqrt{n}$ and $k \lesssim \sqrt{n}$ in Lemmas 8.3 and 8.4 and reduction to SPCA_R in Theorem 8.7.

and $\tau : \mathbb{N} \to \mathbb{N}$ is an arbitrary function with $\tau(n) \to \infty$ as $n \to \infty$. Then $\phi = \text{SPCA-HIGH-SPARSITY}$ is a randomized polynomial time computable map $\phi : \mathcal{G}_n \to \mathbb{R}^{n \times n}$ such that if G is an instance of $PC(n, k, 1/2)$ then under H_0, it holds that $\phi(G) \sim N(0,1)^{\otimes n \times n}$ and under H_1, there is a prior π such that

$$d_{TV}\left(\mathcal{L}_{H_1}(\phi(G)), \int N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right)^{\otimes n} d\pi(u)\right) \leq \frac{2(n+3)}{\tau n - n - 3} + O\left(\frac{1}{\sqrt{\log n}} + k^{-1}\right)$$

where π is supported on unit vectors in $\mathcal{V}_{n,2^\ell k}$.

The next lemma gives the guarantees of SPCA-LOW-SPARSITY, which maps from planted clique to an instance of biclustering and then to sparse PCA. This reduction shows hardness for the canonical simple vs. simple hypothesis testing formulation of sparse PCA. In particular, the output in Lemma 8.2 is close in total variation to the simple vs. simple model UBSPCA. After multiplying the rows of the matrix output in Lemma 8.2 by ± 1, each with probability $1/2$, this also yields a reduction to USPCA. The lemma can be proven with the same applications of the triangle and data processing inequalities as in Lemma 8.3 using the total variation bound in Lemma 6.5 instead of Lemma 7.2.

Before stating the lemma, we determine the parameters of the sparse PCA instance that SPCA-LOW-SPARSITY produces. Under H_1, BC-REDUCTION takes an instance of $G(n, k, 1/2)$ approximately in total variation to $2^{-\ell-1/2}\mu \cdot \mathbf{1}_S \mathbf{1}_T^\top + N(0,1)^{\otimes 2^\ell n \times 2^\ell n}$ where $S, T \subseteq [2^\ell n]$ have size $2^\ell k$ and μ is subpolynomial in n. This matrix can be rewritten as $2^{-1/2}\mu k \cdot uv^\top + N(0,1)^{\otimes 2^\ell n \times 2^\ell n}$

where u, v are $2^\ell k$-sparse unit vectors. Now RANDOM-ROTATION takes this matrix to an instance of UBSPCA$_D$ with the resulting parameters $d = n' = 2^\ell n$, $k' = 2^\ell k$ and $\theta = \frac{\mu^2 k^2}{2^{\ell+1}\tau n}$ where τ, μ are subpolynomial in n.

Lemma 8.4. *Suppose that n and ℓ are such that $\ell = O(\log n)$ and are sufficiently large,*

$$\mu = \frac{\log 2}{2\sqrt{6\log n + 2\log 2}}$$

and $\tau : \mathbb{N} \to \mathbb{N}$ is an arbitrary function with $\tau(n) \to \infty$ as $n \to \infty$. Then $\phi = \text{SPCA-LOW-SPARSITY}$ is a randomized polynomial time computable map $\phi : \mathcal{G}_n \to \mathbb{R}^{2^\ell n \times 2^\ell n}$ such that if G is an instance of $PC(n, k, 1/2)$ then under H_0, it holds that $\phi(G) \sim N(0,1)^{\otimes 2^\ell n \times 2^\ell n}$ and

$$d_{TV}\left(\mathcal{L}_{H_1}(\phi(G)), \int N\left(0, I_n + \frac{\mu^2 k^2}{2^{\ell+1}\tau n} \cdot uu^\top\right)^{\otimes n} d\pi(u)\right) \leq \frac{2(2^\ell n + 3)}{\tau \cdot 2^\ell n - 2^\ell n - 3} + O\left(\frac{1}{\sqrt{\log n}}\right)$$

where π is the uniform distribution over all $2^\ell k$-sparse unit vectors in $\mathbb{R}^{2^\ell n}$ with nonzero entries equal to $1/\sqrt{2^\ell k}$.

We now apply these reductions to deduce planted clique hardness for sparse PCA and its variants. The proofs of the next two theorems are deferred to Appendix E. Note that when $k \ll \sqrt{n}$, the lower bounds for sparse PCA are not tight. For biased sparse PCA, the lower bounds are only tight at the single point when $\theta = \tilde{\Theta}(1)$. The next theorem deduces tight hardness for SPCA$_D$ when $k \gtrsim \sqrt{n}$ with Lemma 8.3.

Theorem 8.5. *Let $\alpha > 0$ and $\beta \in (0, 1)$ be such that $\alpha > \max(1 - 2\beta, 0)$. There is a sequence $\{(N_n, K_n, d_n, \theta_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. *The parameters are in the regime $d = \Theta(N)$, $\theta = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,*

$$\lim_{n \to \infty} \frac{\log \theta_n^{-1}}{\log N_n} = \alpha \quad and \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. *For any sequence of randomized polynomial-time tests $\phi_n : \mathcal{G}_{N_n} \to \{0, 1\}$, the asymptotic Type I+II error of ϕ_n on the problems SPCA$_D(N_n, K_n, d_n, \theta_n)$ is at least 1 assuming the PC conjecture holds for $p = 1/2$.*

Similarly, varying the parameters ℓ and k in Lemma 8.4 yields the following hardness for the simple vs. simple hypothesis testing formulations of biased and ordinary sparse PCA. Since UBSPCA$_D$ and USPCA$_D$ are instances of BSPCA$_D$ and SPCA$_D$, respectively, the next theorem also implies the lower bounds when $\alpha > 1 - 2\beta$ in Theorem 8.5 when $k \lesssim \sqrt{n}$.

Theorem 8.6. *Let $\alpha > 0$ and $\beta \in (0, 1)$ be such that $\frac{1-\alpha}{2} < \beta < \frac{1+\alpha}{2}$. There is a sequence $\{(N_n, K_n, d_n, \theta_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. *The parameters are in the regime $d = \Theta(N)$, $\theta = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,*

$$\lim_{n \to \infty} \frac{\log \theta_n^{-1}}{\log N_n} = \alpha \quad and \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. *For any sequence of randomized polynomial-time tests $\phi_n : \mathcal{G}_{N_n} \to \{0, 1\}$, the asymptotic Type I+II error of ϕ_n on the problems USPCA$_D(N_n, K_n, d_n, \theta_n)$ and UBSPCA$_D(N_n, K_n, d_n, \theta_n)$ is at least 1 assuming the PC conjecture holds for $p = 1/2$.*

To conclude this section, we observe that the reduction SPCA-RECOVERY shows that recovery in UBSPCA$_R$ is hard if $\theta \ll 1$ given the PDS recovery conjecture. The proof of the following theorem follows the same structure as Lemma 8.3 and Theorem 6.8. Note that BC-RECOVERY approximately maps from PDS$_R(n, k, 1/2 + \rho, 1/2)$ to BC$_R(n, k, \mu)$ where $\mu = \frac{\log(1+2\rho)}{2\sqrt{6 \log n + 2 \log 2}} = \tilde{\Theta}(\rho)$. This map preserves the support of the planted dense subgraph in the row support of the planted matrix in BC$_R$. Then RANDOM-ROTATION approximately maps from this BC$_R$ instance to a UBSPCA$_R(n, k, n, \theta)$ instance with $\theta = \frac{k^2 \mu^2}{\tau n} = \tilde{\Theta}\left(\frac{k^2 \rho^2}{n}\right)$. This map ensures that the planted vector u is supported on the same indices as the original PDS$_R$ instance. Furthermore, the PDS conjecture is that the original PDS$_R$ instance is hard if $\rho^2 \ll \frac{n}{k^2}$ which corresponds to the barrier $\theta \ll 1$ under this reduction.

Theorem 8.7. *Let $\alpha \in \mathbb{R}$ and $\beta \in (0, 1)$. There is a sequence $\{(N_n, K_n, D_n, \theta_n)\}_{n \in \mathbb{N}}$ of parameters such that:*

1. *The parameters are in the regime $d = \Theta(N)$, $\theta = \tilde{\Theta}(N^{-\alpha})$ and $K = \tilde{\Theta}(N^\beta)$ or equivalently,*

$$\lim_{n \to \infty} \frac{\log \theta_n^{-1}}{\log N_n} = \alpha \quad and \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

2. *If $\alpha > 0$ and $\beta > \frac{1}{2}$, then the following holds. Let $\epsilon > 0$ be fixed and let X_n be an instance of UBSPCA$_R(N_n, K_n, D_n, \theta_n)$. There is no sequence of randomized polynomial-time computable functions $\phi_n : \mathbb{R}^{D_n \times N_n} \to \binom{[N_n]}{k}^2$ such that for all sufficiently large n the probability that $\phi_n(X_n)$ is exactly the pair of latent row and column supports of X_n is at least ϵ, assuming the PDS recovery conjecture.*

Therefore, given the PDS recovery conjecture, the computational boundary for UBSPCA$_R(n, k, d, \theta)$ in the parameter regime $\theta = \tilde{\Theta}(n^{-\alpha})$ and $k = \tilde{\Theta}(n^\beta)$ is $\alpha^* = 0$ when $\beta > \frac{1}{2}$.

Proof. Suppose that $\alpha > 0$ and $\beta \geq \frac{1}{2}$. Let $\gamma = \beta - \frac{1-\alpha}{2} > 0$ and define

$$K_n = k_n = \lceil n^\beta \rceil, \qquad \rho_n = n^{-\gamma}, \qquad N_n = D_n = n, \qquad \mu_n = \frac{\log(1 + 2\rho_n)}{2\sqrt{6 \log n + 2 \log 2}}, \qquad \theta_n = \frac{k_n^2 \mu_n^2}{\tau n}$$

where τ is an arbitrarily slowly growing function of n. Let $\varphi_n = $ SPCA-RECOVERY be the reduction in Figure 19. Let $G_n \sim G(n, S, 1/2 + \rho_n, 1/2)$ and $X_n = \varphi_n(G_n)$ where S is a k_n-subset of $[n]$. Let u_S denote the unit vector supported on indices in S with nonzero entries equal to $1/\sqrt{k_n}$. Lemma 6.7 and Lemma 8.2 together imply that

$$d_{\mathrm{TV}}\left(\mathcal{L}(X_n), N\left(0, I_n + \theta_n u_S u_S^\top\right)\right) \leq O\left(\frac{1}{\sqrt{\log n}}\right) + \frac{2(n+3)}{\tau n - n - 3} \to 0 \text{ as } n \to \infty$$

Let $\mathcal{L}_{n,S} = N\left(0, I_n + \theta_n u_S u_S^\top\right)$. Assume for contradiction that there is a sequence of randomized polynomial-time computable functions ϕ_n as described above. Now observe that

$$\left|\mathbb{P}_{X \sim \mathcal{L}(X_n)}\left[\phi_n(X) = S\right] - \mathbb{P}_{X \sim \mathcal{L}_{n,S}}\left[\phi_n(X) = S\right]\right| \leq d_{\mathrm{TV}}\left(\mathcal{L}(X_n), \mathcal{L}_{n,S}\right) \to 0 \text{ as } n \to \infty$$

Since $\mathbb{P}_{X \sim \mathcal{L}_{n,S}}\left[\phi_n(X) = S\right] \geq \epsilon$ for sufficiently large n, it follows that $\mathbb{P}_{X \sim \mathcal{L}(X_n)}\left[\phi_n \circ \varphi_n(G_n) = S\right] = \mathbb{P}_{X \sim \mathcal{L}(X_n)}\left[\phi_n(X) = S\right] \geq \epsilon/2$ for sufficiently large n. Furthermore observe

$$\lim_{n \to \infty} \frac{\log k_n}{\log n} = \beta \quad and \quad \lim_{n \to \infty} \log_n\left(\frac{k_n^2 \rho_n^2}{\frac{1}{4} - \rho_n^2}\right) = 2\beta - 2\gamma = 1 - \alpha < 1$$

Since the sequence of functions $\phi_n \circ \varphi_n$ can be computed in randomized polynomial time, this contradicts the PDS recovery conjecture. Therefore no such sequence of functions ϕ_n exists for the parameter sequence $\{(N_n, K_n, D_n, \theta)\}_{n \in \mathbb{N}}$ defined above. As in Theorem 6.8, $\mu_n \sim \frac{\rho_n}{\sqrt{6 \log n}}$ as $n \to \infty$. Therefore it follows that

$$\lim_{n \to \infty} \frac{\log \theta_n^{-1}}{\log N_n} = \lim_{n \to \infty} \frac{2\gamma \log n + \log(6 \log n) + \log n - 2\beta \log n}{\log n} = \alpha \quad \text{and} \quad \lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \beta$$

which completes the proof of the theorem. $\qquad \square$

As is the case for BC-RECOVERY, the reduction SPCA-RECOVERY also shows hardness for partial and weak recovery if the PDS recovery conjecture is strengthened to assume hardness of partial and weak recovery, respectively, for PDS_R.

9 Algorithms and Information-Theoretic Thresholds

In this section, we give the algorithms and information-theoretic lower bounds necessary to prove Theorem 2.1. Specifically, for each problem, we give an information-theoretic lower bound, an inefficient algorithm that achieves the information-theoretic lower bound and a polynomial-time algorithm. As the computational lower bounds and reductions previously presented are the main novel contribution of the paper, the details in this section are succinctly presented only as needed for Theorem 2.1.

Many of the problems we consider have pre-existing algorithms and information-theoretic lower bounds. In these cases, we cite the relevant literature and state the results needed for Theorem 2.1. Note that we only require algorithms and lower bounds optimal up to sub-polynomial factors for Theorem 2.1. For some problems, we only give an information-theoretic lower bound for detection and show that this implies the recovery lower bound in the next section.

9.1 Biclustering, Planted Dense Subgraph and Independent Set

Information-Theoretic Lower Bounds. The information-theoretic lower bound for BC_D was shown in [BI13]. More precisely, they showed the following theorem re-written in our notation. Note that they showed a lower bound for a composite hypothesis testing version of BC_D, but took a uniform prior over the support of the hidden submatrix, matching our formulation.

Theorem 9.1 (Theorem 2.2 in [BI13]). *Suppose that k, μ are such that as $n \to \infty$, it holds that $k/n \to 0$ and one of the following holds*

$$\frac{\mu k^2}{n} \to 0 \quad and \quad \limsup_{n \to \infty} \frac{\mu}{2\sqrt{k^{-1} \log(n/k)}} < 1$$

Then if M_n denotes an instance of $\mathrm{BC}_D(n, k, \mu)$,

$$d_{TV}(\mathcal{L}_{H_0}(M_n), \mathcal{L}_{H_1}(M_n)) \to 0 \quad as \quad n \to \infty$$

This corresponds exactly to the information-theoretic barrier of $\mu \ll \frac{1}{\sqrt{k}}$ and $\mu \ll \frac{n}{k^2}$. We remark that the information-theoretic lower bounds for biclustering can also be deduced from the information-theoretic lower bounds for planted dense subgraph with $q = 1/2$ using the reduction from Lemma 6.7 and the data-processing inequality. Tight information-theoretic lower bounds for PDS_D and PIS_D can be deduced from a mild adaptation of [HWX15]. The argument from the proof of Proposition 3 in [HWX15] yields the following lemma.

Lemma 9.2 (Proposition 3 in [HWX15]). *If G is an instance of $\mathrm{PDS}_D(n,k,p,q)$, then*

$$d_{TV}(\mathcal{L}_{H_0}(G), \mathcal{L}_{H_1}(G)) \leq \frac{1}{2}\sqrt{\mathbb{E}\left[\exp\left(\frac{(p-q)^2}{q(1-q)} \cdot H^2\right) - 1\right]}$$

where $H \sim Hypergeometric(n, k, k)$.

This lemma can be derived by combining the χ^2 computation in the proof of Proposition 3 with Cauchy-Schwarz. In [HWX15], Proposition 3 is specifically for the case when $p = cq$ where $c > 1$ and also for a mixture over $\mathrm{PDS}_D(n, K, p, q)$ where $K \sim \mathrm{Bin}(n, k/n)$. However, the first step in the proof of Proposition 3 is to condition on K and prove this bound for each fixed K. When combined with Lemma 14 from [HWX15], we obtain the desired information-theoretic lower bounds.

Lemma 9.3 (Lemma 14 in [HWX15]). *There is an increasing function $\tau : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ with $\lim_{x \to 0^+} \tau(x) = 1$ and*

$$\mathbb{E}[\exp(\lambda H^2)] \leq \tau(b)$$

where $H \sim Hypergeometric(n, k, k)$, $\lambda = b \cdot \max\left\{\frac{1}{k}\log\left(\frac{en}{k}\right), \frac{n^2}{k^4}\right\}$ and $0 < b < (16e)^{-1}$.

Combining these two lemmas and setting b as

$$b = \frac{(p-q)^2}{q(1-q)} \cdot \left(\max\left\{\frac{1}{k}\log\left(\frac{en}{k}\right), \frac{n^2}{k^4}\right\}\right)^{-1}$$

yields the following theorem on the information-theoretic lower bound for the general regime of PDS_D.

Theorem 9.4. *Suppose p, q, k are such that as $n \to \infty$, it holds that $\frac{(p-q)^2}{q(1-q)} \ll \frac{1}{k}$ and $\frac{(p-q)^2}{q(1-q)} \ll \frac{n^2}{k^4}$. Then if G_n is an instance of $\mathrm{PDS}_D(n, k, p, q)$, it follows that*

$$d_{TV}(\mathcal{L}_{H_0}(G_n), \mathcal{L}_{H_1}(G_n)) \to 0 \quad as \quad n \to \infty$$

Note that when $p = cq$ for some constant $c > 1$ or when $p = 0$, this barrier is $q \ll \frac{1}{k}$ and $q \ll \frac{n^2}{k^4}$. This recovers the information-theoretic lower bounds for PIS_D and PDS_D when $p = cq$. The information-theoretic lower bounds for the weak and strong recovery variants BC_R and PDS_R are derived in [HWX16b]. The following theorems of [HWX16b] characterize these lower bounds.

Theorem 9.5 (Corollary 2 in [HWX16b]). *Suppose k and μ are such that as $n \to \infty$,*

$$k\mu^2 \to \infty \quad and \quad \liminf_{n \to \infty} \frac{(k-1)\mu^2}{\log\frac{n}{k}} > 4 \tag{2}$$

then weak recovery in $\mathrm{BC}_R(n, k, \mu)$ is possible. If weak recovery is possible, then (2) holds as a non-strict inequality.

Theorem 9.6 (Corollary 4 in [HWX16b]). *Suppose k and μ are such that as $n \to \infty$, condition (2) holds and*

$$\liminf_{n \to \infty} \frac{k\mu^2}{\left(\sqrt{2\log n} + \sqrt{2\log k}\right)^2} > 1 \tag{3}$$

then exact recovery in $\mathrm{BC}_R(n, k, \mu)$ is possible. If exact recovery is possible, then (2) and (3) hold as a non-strict inequalities.

Theorem 9.7 (Corollary 1 in [HWX16b])**.** *Suppose p and q are such that the ratios $\log \frac{p}{q}$ and $\log \frac{1-p}{1-q}$ are bounded as $n \to \infty$. If k satisfies*

$$k \cdot d_{KL}(p, q) \to \infty \quad and \quad \liminf_{n \to \infty} \frac{k \cdot d_{KL}(p, q)}{\log \frac{n}{k}} > 2 \tag{4}$$

then there is an algorithm achieving weak recovery for $\mathrm{PDS}_R(n, k, p, q)$ is possible. If weak recovery is possible, then (4) holds as a non-strict inequality.

Theorem 9.8 (Corollary 3 in [HWX16b])**.** *Suppose p and q are such that the ratios $\log \frac{p}{q}$ and $\log \frac{1-p}{1-q}$ are bounded as $n \to \infty$. If k satisfies*

$$\tau = \frac{\log \frac{1-q}{1-p} + \frac{1}{k} \log \frac{n}{k}}{\log \frac{p(1-q)}{q(1-p)}}$$

If (4) holds and

$$\liminf_{n \to \infty} \frac{k \cdot d_{KL}(\tau, q)}{\log n} > 1 \tag{5}$$

then exact recovery in $\mathrm{PDS}_R(n, k, p, q)$ is possible. If exact recovery is possible, then (4) and (5) hold as non-strict inequalities.

These theorems show that both weak and strong recovery for $\mathrm{BC}_R(n, k, \mu)$ are information-theoretically impossible when $\mu \lesssim \frac{1}{\sqrt{k}}$ by the first condition in (2). Now note that if $p - q = O(q)$ and $q \to 0$ as $n \to \infty$, then

$$
\begin{aligned}
d_{\mathrm{KL}}(p, q) &= p \log \left(\frac{p}{q} \right) + (1 - p) \cdot \log \left(\frac{1-p}{1-q} \right) \\
&= p \cdot \left(\frac{p-q}{q} \right) - O \left(p \cdot \left(\frac{p-q}{q} \right)^2 \right) - (1 - p) \cdot \left(\frac{p-q}{1-q} \right) - O \left((1 - p) \cdot \left(\frac{p-q}{1-q} \right)^2 \right) \\
&= \frac{(p-q)^2}{q(1-q)} + O \left(p \cdot \left(\frac{p-q}{q} \right)^2 + (p-q)^2 \right) = O \left(\frac{(p-q)^2}{q(1-q)} \right)
\end{aligned}
$$

as $n \to \infty$. Therefore it follows that if $p - q = O(q)$, $q \to 0$ as $n \to \infty$ and $\log \frac{p}{q}$ and $\log \frac{1-p}{1-q}$ are bounded as $n \to \infty$ then both weak and strong recovery in $\mathrm{PDS}_R(n, k, p, q)$ are information-theoretically impossible if $\frac{(p-q)^2}{q(1-q)} \lesssim \frac{1}{k}$ by the first condition in (4). Note that these theorems of [HWX16b] do not imply the necessary information-theoretic lower bound for PIS_R since $p = 0$ violates the condition that $\log \frac{p}{q}$ is bounded. However, the genie argument in the necessary part of Theorem 1 in [HWX16b] can be mildly adapted to obtain the following theorem, the proof of which is deferred to Appendix F.

Theorem 9.9. *If $k \geq 2$, $q \ll \frac{1}{k}$ and $n - k = \Omega(n)$ as $n \to \infty$, then weak recovery in $\mathrm{PIS}_R(n, k, q)$ is impossible.*

Information-Theoretically Optimal Algorithms. A corresponding algorithm achieving the information-theoretic lower bound for BC_D was also shown in [BI13]. Their algorithm outputs the hypothesis H_1 if either the maximum sum over all $k \times k$ submatrices of the input exceeds a threshold or if the total sum of the input exceeds another threshold. The guarantees of this algorithm are summarized in the following theorem.

Theorem 9.10 (Theorem 2.1 in [BI13])**.** *Suppose that k, μ are such that as $n \to \infty$, it holds that $k/n \to 0$ and one of the following holds*

$$\frac{\mu k^2}{n} \to \infty \quad or \quad \limsup_{n \to \infty} \frac{\mu}{2\sqrt{k^{-1} \log(n/k)}} > 1$$

Then there is an algorithm solving $\mathrm{BC}_D(n, k, \mu)$ with Type I+II error tending to zero as $n \to \infty$.

A very similar algorithm is optimal for PDS_D. Generalizing the concentration bounds in the proof of Proposition 4 in [HWX15] to any p, q with $p - q = O(q)$ and $q \to 0$ yields the following theorem, the proof of which is deferred to Appendix F.

Theorem 9.11. *Suppose that p, q and k are such that $|p - q| = O(q)$, $q \to 0$ and*

$$\frac{(p-q)^2}{q(1-q)} = \omega\left(\frac{n^2}{k^4}\right) \quad or \quad \frac{(p-q)^2}{q(1-q)} = \omega\left(\frac{\log(n/k)}{k}\right)$$

as $n \to \infty$. Then there is an algorithm solving $\mathrm{PDS}_D(n, k, p, q)$ with Type I+II error tending to zero as $n \to \infty$.

This theorem gives the necessary algorithm matching the information-theoretic lower bound for PDS_D in the general regime $p - q = O(q)$, including $p = cq$ for some constant $c > 1$. The algorithm needed for PIS_D can be obtained by setting $p = 0$ in this theorem. An algorithm matching the information-theoretic lower bound for BC_R follows from Theorem 9.6, which asserts that exact recovery is possible as long as

$$\mu > (1 + \epsilon) \cdot \frac{\sqrt{2 \log n} + \sqrt{2 \log k}}{\sqrt{k}}$$

for some fixed $\epsilon > 0$. Specializing Corollary 2.4 in [CX16] to the case of $r = 1$ clusters yields an analogous algorithm for PDS_R.

Theorem 9.12 (Corollary 2.4 in [CX16])**.** *Suppose that p, q and k are such that $p > q$ and*

$$\frac{(p-q)^2}{q(1-q)} \geq \frac{C \log n}{k}, \quad q \geq \frac{C \log k}{k} \quad and \quad kq \log \frac{p}{q} \geq C \log n$$

for some sufficiently large constant $C > 0$. Then the maximum likelihood estimator for the planted dense subgraph in $\mathrm{PDS}_R(n, k, p, q)$ solves strong recovery with error probability tending to zero.

This implies that if $p > q$, $p - q = O(q)$, $q \to 0$ and $\frac{(p-q)^2}{q(1-q)} \gtrsim \frac{1}{k}$ as $n \to \infty$, then exact recovery is possible. Specializing the result to $p = 1$ and applying this algorithm to the complement graph of a PIS_R instance yields that there is an algorithm for PIS_R if $q \gtrsim \frac{1}{k}$. We remark that the necessary algorithm for PDS_R can also be deduced from Theorem 9.8. However, the constraints that $\log \frac{p}{q}$ and $\log \frac{1-p}{1-q}$ must be bounded does not yield the desired algorithm for PIS_R.

Polynomial-Time Algorithms. The polynomial time algorithm matching our planted clique lower bound for BC_D is another simple algorithm thresholding the maximum and sum of the input matrix. Given an instance M of $\mathrm{BC}_D(n, k, \mu)$, let $\max(M) = \max_{i,j \in [n]} M_{ij}$ and $\mathrm{sum}(M) = \sum_{i,j=1}^{n} M_{ij}$. Specializing Lemma 1 of [MW15b] to our setup yields the following lemma.

Lemma 9.13 (Lemma 1 in [MW15b]). *If M is an instance of $\mathrm{BC}_D(n, k, \mu)$ then*

$$\mathbb{P}_{H_0}\left[\mathrm{sum}(M) > \frac{\mu k^2}{2}\right] + \mathbb{P}_{H_1}\left[\mathrm{sum}(M) \le \frac{\mu k^2}{2}\right] \le \exp\left(-\frac{\mu^2 k^4}{8n^2}\right)$$

If $c > 0$ is any absolute constant and $\tau = \sqrt{(4 + c)\log n}$, then

$$\mathbb{P}_{H_0}\left[\max(M) > \tau\right] + \mathbb{P}_{H_1}\left[\max(M) \le \tau\right] \le n^{-c/2} + \exp\left(-\frac{1}{2}|\mu - \tau|_+\right)$$

It follows that the algorithm that outputs H_1 if $\max(M) > \sqrt{5\log n}$ or $\mathrm{sum}(M) > \frac{\mu k^2}{2}$ solves BC_D with Type I+II error tending to zero as $n \to \infty$ if either $\mu \ge \sqrt{6\log n}$ or $\mu = \omega\left(\frac{n}{k^2}\right)$. By Theorem 9.11, if $\frac{(p-q)^2}{q(1-q)} = \omega\left(\frac{n^2}{k^4}\right)$ it follows that thresholding the number of edges of an instance G of $\mathrm{PDS}_D(n, k, p, q)$ has Type I+II error tending to zero as $n \to \infty$. Setting $p = 0$ recovers the computational barrier of PIS_D. Polynomial-time algorithms for the recovery variants of these problems were given in [CX16]. The following are three theorems of [CX16] written in our notation.

Theorem 9.14 (Theorem 2.5 in [CX16]). *A polynomial-time convex relaxation of the MLE solves exact recovery in $\mathrm{PDS}_R(n, k, p, q)$ with error probability at most n^{-10} if $p > q$ and*

$$k^2(p - q) \ge C\left[p(1 - q)k\log n + q(1 - q)n\right]$$

where $C > 0$ is a fixed constant.

Theorem 9.15 (Theorem 3.3 in [CX16]). *A polynomial-time convex relaxation of the MLE solves exact recovery in $\mathrm{BC}_R(n, k, \mu)$ with error probability at most n^{-10} if*

$$\mu^2 \ge C\left[\frac{\log n}{k} + \frac{n}{k^2}\right]$$

where $C > 0$ is a fixed constant.

Theorem 9.16 (Theorem 3.3 in [CX16]). *A polynomial-time element-wise thresholding algorithm solves exact recovery in $\mathrm{BC}_R(n, k, \mu)$ with error probability at most n^{-3} if $\mu^2 \ge C\log n$ where $C > 0$ is a fixed constant.*

Since $k\log n = \tilde{O}(n)$ and $p = O(q)$ if $p - q = O(q)$, the first theorem above implies that exact recovery is possible in polynomial time for the general regime of PDS_R if $\frac{(p-q)^2}{q(1-q)} \gg \frac{n}{k^2}$. Taking the complement graph of the input and setting $p = 1$ and $q = 1 - \tilde{\Theta}(n^{-\alpha})$ in the first theorem yields that $\mathrm{PIS}_R(n, k, 1 - q)$ can be solved in polynomial time if $1 - q \gg \frac{n}{k^2}$. The second and third theorems above imply that exact recovery for BC_R is possible in polynomial time if $\mu \gg \frac{1}{\sqrt{k}}$ or $\mu \gg 1$. These polynomial-time algorithms for detection and recovery match the computational lower bounds shown in previous sections.

9.2 Rank-1 Submatrix, Sparse Spiked Wigner and Subgraph SBM

Information-Theoretic Lower Bounds. Applying a similar χ^2 computation as in information-theoretic lower bounds for sparse PCA and planted dense subgraph, we can reduce showing an information-theoretic lower bound for SROS_D to bounding an MGF. In the case of SROS_D, this MGF turns out to be that of the square of a symmetric random walk on $\mathbb{Z}$ terminated after a hypergeometric number of steps. An asymptotically tight upper bound on this MGF was obtained in [CMW15] through the following lemma.

Lemma 9.17 (Lemma 1 in [CMW15]). *Suppose that $d \in \mathbb{N}$ and $k \in [p]$. Let $B_1, B_2, \ldots, B_k$ be independent Rademacher random variables. Let the symmetric random walk on $\mathbb{Z}$ stopped at the mth step be*

$$G_m = \sum_{i=1}^{m} B_i$$

If $H \sim \text{Hypergeometric}(d, k, k)$ then there is an increasing function $g : (0, 1/36) \to (1, \infty)$ such that $\lim_{x \to 0^+} g(x) = 1$ and for any $a \in (0, 1/36)$, it holds that

$$\mathbb{E}\left[\exp\left(G_H^2 \cdot \frac{a}{k} \log \frac{ed}{k}\right)\right] \leq g(a)$$

With this bound, we obtain the following information-theoretic lower bound for SROS_D, which matches Theorem 2.1.

Theorem 9.18. *Suppose that M is an instance of $\text{SROS}_D(n, k, \mu)$ where under H_1, the planted vector v is chosen uniformly at random from all k-sparse unit vectors in $\mathbb{R}^n$ with nonzero coordinates equal to $\pm\frac{1}{\sqrt{k}}$. Suppose it holds that $\mu \leq \sqrt{\beta_0 k \log \frac{en}{k}}$ for some $0 < \beta_0 < (16e)^{-1}$. Then there is a function $w : (0, 1) \to (0, 1)$ satisfying that $\lim_{\beta_0 \to 0^+} w(\beta_0) = 0$ and*

$$d_{TV}(\mathcal{L}_{H_0}(M), \mathcal{L}_{H_1}(M)) \leq w(\beta_0)$$

Proof. Let $\mathbb{P}_0$ denote $\mathcal{L}_{H_0}(M) = N(0, 1)^{\otimes n \times n}$ and $\mathbb{P}_u$ denote $\mathcal{L}\left(\mu \cdot uu^\top + N(0, 1)^{\otimes n \times n}\right)$ where u is in the set S of k-sparse unit vectors u with nonzero entries equal to $\pm 1/\sqrt{k}$. Now let $\mathcal{P}_1$ denote $\mathcal{L}_{H_1}(M)$ which can also be written as

$$\mathbb{P}_1 = \frac{1}{|S|} \sum_{u \in S} \mathbb{P}_u$$

Given two matrices $A, B \in \mathbb{R}^{n \times n}$, let $\langle A, B \rangle = \sum_{i,j=1}^{n} A_{ij} B_{ij}$ denote their inner product. Now note that for any $X \in \mathbb{R}^{n \times n}$,

$$\frac{d\mathbb{P}_u}{d\mathbb{P}_0}(X) = \exp\left(-\frac{1}{2} \sum_{i,j=1}^{n} (X_{ij} - \mu \cdot u_i u_j)^2 + \frac{1}{2} \sum_{i,j=1}^{n} X_{ij}^2\right)$$

$$= \exp\left(\mu \cdot \langle X, uu^\top \rangle - \frac{\mu^2}{2} \|u\|_2^4\right) = \exp\left(\mu \cdot \langle X, uu^\top \rangle - \frac{\mu^2}{2}\right)$$

since $\|u\|_2 = 1$. Now observe that

$$\chi^2(\mathbb{P}_1, \mathbb{P}_0) = \mathbb{E}_{X \sim \mathbb{P}_0}\left[\left(\frac{d\mathbb{P}_1}{d\mathbb{P}_0}(X) - 1\right)^2\right] = -1 + \frac{1}{|S|^2} \sum_{u,v \in S} \mathbb{E}_{X \sim \mathbb{P}_0}\left[\frac{d\mathbb{P}_u}{d\mathbb{P}_0}(X) \cdot \frac{d\mathbb{P}_v}{d\mathbb{P}_0}(X)\right]$$

$$= -1 + \frac{1}{|S|^2} \sum_{u,v \in S} \mathbb{E}_{X \sim \mathbb{P}_0}\left[\exp\left(\mu \cdot \langle X, uu^\top + vv^\top \rangle - \mu^2\right)\right]$$

$$= -1 + \frac{1}{|S|^2} \sum_{u,v \in S} \exp\left(\frac{\mu^2}{2} \left\|uu^\top + vv^\top\right\|_F^2 - \mu^2\right)$$

$$= -1 + \frac{1}{|S|^2} \sum_{u,v \in S} \exp\left(\frac{\mu^2}{2} \langle uu^\top, uu^\top \rangle + \mu^2 \langle uu^\top, vv^\top \rangle + \frac{\mu^2}{2} \langle vv^\top, vv^\top \rangle - \mu^2\right)$$

$$= -1 + \frac{1}{|S|^2} \sum_{u,v \in S} \exp\left(\mu^2 \langle u, v \rangle^2\right) = -1 + \mathbb{E}_{u,v \sim \text{Unif}[S]}\left[\exp\left(\mu^2 \langle u, v \rangle^2\right)\right]$$

where the third inequality follows since $\mathbb{E}[\exp{(\langle t, X \rangle)}] = \exp{\left(\frac{1}{2}\|t\|_2^2\right)}$ and the last inequality follows since $\langle uu^\top, uu^\top \rangle = \|u\|_2^4 = \langle vv^\top, vv^\top \rangle = \|v\|_2^4 = 1$ and $\langle uu^\top, vv^\top \rangle = \langle u, v \rangle^2$. Let G_m denote a symmetric random walk on $\mathbb{Z}$ stopped at the mth step and $H \sim \text{Hypergeometric}(n, k, k)$ as in Lemma 9.17. Now note that if $u, v \sim \text{Unif}[S]$ are independent, then $\langle u, v \rangle$ is distributed as G_H/k. Now let $a = \mu^2 \left(k \log \frac{en}{k}\right)^{-1} \leq \beta_0$ and note that Lemma 9.17 along with Cauchy-Schwarz implies that

$$d_{\text{TV}}(\mathbb{P}_0, \mathbb{P}_1) \leq \frac{1}{2}\sqrt{\chi^2(\mathbb{P}_1, \mathbb{P}_0)} \leq \frac{1}{2}\sqrt{g(\beta_0) - 1}$$

where g is the function from Lemma 9.17. Setting $w(\beta_0) = \frac{1}{2}\sqrt{g(\beta_0) - 1}$ proves the theorem. $\qquad\square$

Note that any instance of SROS_D is also an instance of ROS_D and thus the information theoretic lower bound in Theorem 9.18 also holds for ROS_D. Symmetrizing SROS_D as in Section 7 yields that the same information-theoretic lower bound holds for SSW_D. Now consider the function $\tau : \mathbb{R}^{n \times n} \to \mathcal{G}_n$ that such that if $\tau(M) = G$ then $\{i, j\} \in E(G)$ if and only if $M_{ij} > 0$ for all $i < j$. In other words, τ thresholds the above-diagonal entries of M as in Step 3 of SSBM-REDUCTION from Lemma 7.6. Note that τ maps $N(0, 1)^{\otimes n \times n}$ to $G(n, 1/2)$ and takes $\mathcal{L}_{H_1}(M)$ from Theorem 9.18 to a distribution in $\mathcal{L}_{\text{SSBM}} \in G_B(n, k, 1/2, \rho)$ where

$$\rho = \Phi\left(\frac{\mu}{k}\right) - \frac{1}{2} = \frac{1}{\sqrt{2\pi}} \cdot \frac{\mu}{k}$$

As in the proof of Theorem 7.7 there is a method $e : \mathcal{G}_n \to \mathcal{G}_n$ that either adds or removes edges with a fixed probability mapping $G(n, 1/2)$ to $G(n, q)$ and $\mathcal{L}_{\text{SSBM}}$ to $e(\mathcal{L}_{\text{SSBM}}) \in G_B(n, k, 1/2, \rho')$ where $\rho' = \Theta(\rho)$ as long as $q = \Theta(1)$. By the data processing inequality, we now have that

$$d_{\text{TV}}\left(G(n, 1/2), \mathcal{L}'_{\text{SSBM}}\right) \leq d_{\text{TV}}\left(\mathcal{L}_{H_0}(M), \mathcal{L}_{H_1}(M)\right) \to 0 \quad \text{as } n \to \infty$$

if $\mu \ll \sqrt{k}$ which corresponds to $\rho \ll 1/\sqrt{k}$, establishing the information theoretic lower bound for SSBM_D in the regime $q = \Theta(1)$ matching Theorem 2.1.

Corresponding recovery lower bounds for these problems follow from information-theoretic lower bounds for biclustering. Note that an instance of $\text{BC}_{WR}(n, k, \mu)$ is an instance of $\text{ROS}_{WR}(n, k, \mu/k)$. By Theorem 9.5, $\text{ROS}_{WR}(n, k, \mu)$ is therefore information-theoretically impossible if $\mu \leq 2\sqrt{k \log \frac{n}{k}}$. An analogous information-theoretic lower bound is given for a symmetric variant of BC_{WR} in [HWX16b], which implies the corresponding lower bounds for SROS_{WR} and SSW_{WR}.

Information-Theoretically Optimal Algorithms. Unlike existing maximum likelihood estimators for recovery such as those for BC_R in [CX16] and [CLR15], the definition of $\mathcal{V}_{n,k}$ requires that algorithms solving ROS_R are adaptive to the sparsity level k. We introduce a modified exhaustive search algorithm ROS-SEARCH that searches over all possible sparsity levels and checks whether each resulting output is reasonable using an independent copy B of the data matrix.

We first establish the notation that will be used in this section. Given some $v \in \mathbb{R}^n$, let $\text{supp}_+(v)$ denote the set of i with $v_i > 0$ and $\text{supp}_-(v)$ denote the set of i with $v_i < 0$. If $A, B \in \mathbb{R}^{n \times n}$, let $\langle A, B \rangle = \text{Tr}(A^\top B)$. Let S_t be the set of $v \in \mathbb{R}^n$ with exactly t nonzero entries each in $\{-1, 1\}$. In order to show that ROS-SEARCH succeeds at solving ROS_R asymptotically down to its information-theoretic limit, we begin by showing the following lemma.

Lemma 9.19. *Let R and C be subsets of $[n]$ such that $|R| = k_1$ and $|C| = k_2$ where $k_1, k_2 \in [c_1 k, k]$ for some constant $c_1 \in (0, 1)$. Let $\rho > 0$ and $M \in \mathbb{R}^{n \times n}$ be a random matrix and with independent sub-Gaussian entries with sub-Gaussian norm at most 1 such that:*

Algorithm ROS-SEARCH

Inputs: Matrix $M \in \mathbb{R}^{n \times n}$, sparsity upper bound k, threshold $\rho > 0$, constant $c_1 \in (0, 1)$

1. Sample $G \sim N(0,1)^{\otimes n \times n}$ and form $A = \frac{1}{\sqrt{2}}(M + G)$ and $B = \frac{1}{\sqrt{2}}(M - G)$

2. For each pair $k_1, k_2 \in [c_1 k, k]$ do:

    a. Let S_t be the set of $v \in \mathbb{R}^n$ with exactly t nonzero entries each in $\{-1, 1\}$ and compute

$$(u, v) = \mathrm{argmax}_{(u,v) \in S_{k_1} \times S_{k_2}} \left\{ u^\top A v \right\}$$

    b. Mark the pair (u, v) if it satisfies that
- The set of i with $\sum_{j=1}^n u_i v_j B_{ij} \geq \frac{1}{2} k_2 \rho$ is exactly $\mathrm{supp}(u)$
- The set of j with $\sum_{i=1}^n u_i v_j B_{ij} \geq \frac{1}{2} k_1 \rho$ is exactly $\mathrm{supp}(v)$

3. Output $\mathrm{supp}(u), \mathrm{supp}(v)$ where (u, v) is the marked pair maximizing $|\mathrm{supp}(u)| + |\mathrm{supp}(v)|$

Figure 20: Exhaustive search algorithm for sparse rank-1 submatrix recovery in Theorem 9.19.

- $\mathbb{E}[M_{ij}] \geq \rho$ *if $(i, j) \in R \times C$; and*
- $\mathbb{E}[M_{ij}] = 0$ *if $(i, j) \notin R \times C$.*

There is an absolute constant $c_2 > 0$ such that if $k\rho^2 \geq c_2 \log n$, then

$$\mathrm{argmax}_{(u,v) \in S_{k_1} \times S_{k_2}} \left\{ u^\top M v \right\}$$

is either $(\mathbf{1}_R, \mathbf{1}_C)$ and $(-\mathbf{1}_R, -\mathbf{1}_C)$ with probability at least $1 - n^{-1}$ for sufficiently large n.

Proof. For each pair $(u, v) \in S_{k_1} \times S_{k_2}$, let $A_1(u, v)$ be the set of pairs $(i, j) \in R \times C$ with $u_i v_j = -1$, let $A_2(u, v)$ be the set of pairs $(i, j) \in \mathrm{supp}(u) \times \mathrm{supp}(v)$ that are not in $R \times C$ and let $A_3(u, v)$ be the set of $(i, j) \in R \times C$ that are not in $\mathrm{supp}(u) \times \mathrm{supp}(v)$. Now observe that

$$\mathbf{1}_R^\top M \mathbf{1}_C - u^\top M v = \langle M, \mathbf{1}_R \mathbf{1}_C^\top - uv^\top \rangle = \sum_{(i,j) \in A_1(u,v)} 2 M_{ij} - \sum_{(i,j) \in A_2(u,v)} u_i v_j M_{ij} + \sum_{(i,j) \in A_3(u,v)} M_{ij}$$

$$\geq \rho \left(2|A_1(u,v)| + |A_3(u,v)| \right) + \sum_{(i,j) \in A_1(u,v)} 2 \left(M_{ij} - \mathbb{E}[M_{ij}] \right)$$

$$- \sum_{(i,j) \in A_2(u,v)} u_i v_j M_{ij} + \sum_{(i,j) \in A_3(u,v)} \left(M_{ij} - \mathbb{E}[M_{ij}] \right)$$

Since $R \times C$ and $\mathrm{supp}(u) \times \mathrm{supp}(v)$ both have size $k_1 k_2$, it follows that $|A_2(u, v)| = |A_3(u, v)|$. Note that the random variables in the sum above are independent, zero mean and sub-Gaussian with norm at most 1. By Hoeffding's inequality for sub-Gaussian random variables as in Proposition

5.10 in [Ver10], it follows that

$$\mathbb{P}\left[\langle M, \mathbf{1}_R \mathbf{1}_C^\top - uv^\top \rangle \leq 0\right] \leq e \cdot \exp\left(-\frac{c\rho^2 \left(2|A_1(u,v)| + |A_2(u,v)|\right)^2}{4|A_1(u,v)| + |A_2(u,v)| + |A_3(u,v)|}\right)$$

$$= e \cdot \exp\left(-\frac{1}{2}c\rho^2 \left(2|A_1(u,v)| + |A_2(u,v)|\right)\right)$$

$$\leq e \cdot n^{-c_1^{-2} \cdot \frac{16}{k}\left(2|A_1(u,v)| + |A_2(u,v)|\right)}$$

for some absolute constant $c > 0$ as long as $c\rho^2 \geq 16kc_1^{-2}\log n$. Let $S(a_1, a_2, b_1, b_2)$ be the set of all pairs (u,v) such that $a_1 = |\text{supp}(u)\backslash R|$, $a_2 = |\text{supp}_-(u) \cap R|$, $b_1 = |\text{supp}(v)\backslash C|$ and $b_2 = |\text{supp}_-(v) \cap C|$. Suppose that $a_2 \leq \frac{1}{2}(k_1 - a_1)$. Note that for any $(u,v) \in S(a_1, a_2, b_1, b_2)$, we have

$$|A_1(u,v)| = a_2(k_2 - b_1 - b_2) + b_2(k_1 - a_1 - a_2) \quad \text{and} \quad |A_2(u,v)| = a_1 k_2 + b_1 k_1 - a_1 b_1$$

Note that $k_1, k_2 \geq c_1 k$, $a_1 + a_2 \leq k_1$ and $b_1 + b_2 \leq k_2$. Therefore we have that $\frac{1}{k}|A_2(u,v)| \geq a_1 \cdot \frac{k_2}{k} \geq c_1 a_1$ and $\frac{1}{k}|A_2(u,v)| \geq c_1 b_1$. This implies that $\frac{1}{k}|A_2(u,v)| \geq \frac{1}{2}c_1(a_1 + b_1)$. Now note that if $b_2 \geq c_1 a_2$, then it holds that

$$\frac{1}{k}|A_1(u,v)| \geq \frac{b_2}{k}(k_1 - a_1 - a_2) \geq \frac{1}{2k}b_2(k_1 - a_1) \geq \frac{c_1}{2} \cdot b_2 - \frac{a_1}{2} \geq \frac{c_1^2}{4}(a_2 + b_2) - \frac{a_1}{2}$$

Otherwise if $b_2 < c_1 a_2$ then it follows that $b_2 < c_1 a_2 \leq c_1 \cdot \frac{1}{2}(k_1 - a_1) \leq \frac{k_2}{2}$ since $k_2 \geq c_1 k \geq c_1 k_1$. Now we have that

$$\frac{1}{k}|A_1(u,v)| \geq \frac{a_2}{k}(k_2 - b_1 - b_2) \geq \frac{a_2}{k}(k_2 - b_2) - b_1 \geq \frac{c_1}{2}a_2 - b_1 \geq \frac{c_1}{4}(a_2 + b_2) - b_1$$

Therefore in either case it follows that

$$\frac{1}{k}|A_1(u,v)| \geq \frac{c_1^2}{4}(a_2 + b_2) - a_1 - b_1$$

Combining these inequalities and the fact that $c_1 \in (0,1)$ yields that

$$\frac{2}{k}|A_1(u,v)| + \frac{1}{k}|A_2(u,v)| \geq \frac{c_1}{4k} \cdot |A_1(u,v)| + \frac{1}{k}|A_2(u,v)|$$

$$\geq \frac{c_1^2}{4}(a_2 + b_2) + \frac{c_1}{4}(a_1 + b_1) \geq \frac{c_1^2}{4}(a_1 + a_2 + b_1 + b_2)$$

as long as $a_2 \leq \frac{1}{2}(k_1 - a_1)$. Furthermore, we have that

$$|S(a_1, a_2, b_1, b_2)| = \binom{k_1}{a_1}\binom{n-k_1}{a_1}\binom{k_1 - a_1}{a_2}\binom{k_2}{b_1}\binom{n-k_2}{b_1}\binom{k_2 - b_1}{b_2} \leq n^{2a_1 + 2b_1 + a_2 + b_2}$$

since $\binom{n}{k} \leq n^k$ and $k_1, k_2 \leq n$. Let T be the set of $(a_1, a_2, b_1, b_2) \neq (0,0,0,0)$ such that $a_1, a_2, b_1, b_2 \geq 0$, $a_1 + a_2 \leq k_1$, $b_1 + b_2 \leq k_2$ and $a_2 \leq \frac{1}{2}(k_1 - a_1)$. Now note for all $(u,v) \in S^2$, it holds that at least one of the pairs (u,v) or $(-u,-v)$ satisfies that $a_2 \leq \frac{1}{2}(k_1 - a_1)$. Since (u,v) and $(-u,-v)$

yield the same value of $u^\top M v$, we can restrict to T in the following union bound. Now note that

$$\mathbb{P}\left[\text{there is } (u,v) \in S^2 \text{ with } (u,v) \neq \pm(\mathbf{1}_R, \mathbf{1}_C) \text{ and } \langle M, \mathbf{1}_R \mathbf{1}_C^\top - uv^\top \rangle \leq 0\right]$$

$$\leq \sum_{(a_1,a_2,b_1,b_2) \in T} \left(\sum_{(u,v) \in S(a_1,a_2,b_1,b_2)} e \cdot n^{-c_1^{-2} \cdot \frac{16}{k}(2|A_1(u,v)| + |A_2(u,v)|)} \right)$$

$$\leq \sum_{(a_1,a_2,b_1,b_2) \in T} |S(a_1, a_2, b_1, b_2)| \cdot n^{-4(a_1+a_2+b_1+b_2)}$$

$$\leq \sum_{(a_1,a_2,b_1,b_2) \in T} n^{-2a_1 - 3a_2 - 2b_1 - 3b_2}$$

$$\leq -1 + \sum_{a_1,a_2,b_1,b_2 = 0}^{\infty} n^{-2a_1 - 3a_2 - 2b_1 - 3b_2}$$

$$= -1 + \left(\sum_{i=0}^{\infty} n^{-2i} \right)^2 \left(\sum_{j=0}^{\infty} n^{-3j} \right)^2$$

$$= -1 + (1 - n^{-2})^{-2}(1 - n^{-3})^{-2} = O(n^{-2})$$

which as at most n^{-1} for sufficiently large n, completing the proof of the lemma. $\qquad\square$

We now use this lemma to prove the following theorem, which shows that ROS-SEARCH solves ROS_R and SSW_R as long as $\mu \gtrsim \sqrt{k}$, asymptotically matching their information theoretic limits.

Theorem 9.20. *Suppose that $M \sim \mu \cdot rc^\top + N(0,1)^{\otimes n \times n}$ where $r, c \in \mathcal{V}_{n,k}$. There is an an absolute constant $c > 0$ such that if $\mu \geq c\sqrt{k \log n}$, then ROS-SEARCH applied with $c_1 = 1/2$ and $\rho = \mu/k$ outputs $\text{supp}(r)$ and $\text{supp}(c)$ with probability at least $1 - 4n^{-1}$ for sufficiently large n.*

Proof. Suppose that $(u,v) \in S_{k_1} \times S_{k_2}$ where $k_1, k_2 \in [c_1 k, k]$ are random vectors that are independent of B and either $\text{supp}(u) \not\subseteq \text{supp}(r)$ or $\text{supp}(v) \not\subseteq \text{supp}(c)$. Note that the definition of $\mathcal{V}_{n,k}$ is such that any fixed $c_1 \in (0,1)$ suffices for sufficiently large k. We first observe that if $\rho = \mu/k$ and $\mu \geq c\sqrt{k \log n}$ then (u,v) is not marked in Step 2b of ROS-SEARCH with probability at least $1 - n^{-3}$. If $\text{supp}(u) \not\subseteq \text{supp}(r)$, then let $i \in \text{supp}(u) \backslash \subseteq \text{supp}(r)$. It follows that $\sum_{j=1}^{n} u_i v_j B_{ij} \sim N(0, k_2)$ since $\|v\|_0 = k_2$ and by Gaussian tail bounds that

$$\mathbb{P}\left[\sum_{j=1}^{n} u_i v_j B_{ij} \geq \frac{1}{2}k_2 \rho \right] \leq \frac{1}{\sqrt{2\pi}} \cdot \frac{2}{\rho\sqrt{k_2}} \cdot \exp\left(-\frac{\rho^2 k_2}{8} \right) \leq n^{-3}$$

if $\mu^2 \geq \sqrt{3c_1 k \log n}$. This implies that if $\text{supp}(u) \not\subseteq \text{supp}(r)$, then (u,v) is marked in Step 2b of ROS-SEARCH with probability at most n^{-3}. A symmetric argument shows that the same is true if $\text{supp}(v) \not\subseteq \text{supp}(c)$. Now for each pair $k_1, k_2 \in [c_1 k, k]$, let u_{k_1} and v_{k_2} be such that

$$(u_{k_1}, v_{k_2}) = \text{argmax}_{(u,v) \in S_{k_1} \times S_{k_2}} \left\{ u^\top A v \right\}$$

if the pair is marked and let $(u_{k_1}, v_{k_2}) = (0,0)$ otherwise. By Lemma 10.2 in the next section, A and B from Step 1 of ROS-SEARCH are i.i.d. and distributed as $\frac{1}{\sqrt{2}} \cdot rc^\top + N(0,1)^{\otimes n \times n}$. In

particular, the pairs (u_{k_1}, v_{k_2}) are in the σ-algebra generated by A and hence are independent of B. By a union bound, we have

$$\mathbb{P}\left[\operatorname{supp}(u_{k_1}) \subseteq \operatorname{supp}(r) \text{ and } \operatorname{supp}(v_{k_2}) \subseteq \operatorname{supp}(c) \text{ for all } k_1, k_2 \in [c_1 k, k]\right]$$

$$\geq 1 - \sum_{k_1, k_2 \in [c_1 k, k]} \mathbb{P}\left[\operatorname{supp}(u_{k_1}) \not\subseteq \operatorname{supp}(r) \text{ or } \operatorname{supp}(v_{k_2}) \not\subseteq \operatorname{supp}(c)\right]$$

$$\geq 1 - k^2 \cdot n^{-3} \geq 1 - n^{-1}$$

Now let $k_1' = \|r\|_0$ and $k_2' = \|c\|_0$ and let r' be the vector such that $r_i' = 0$ if $v_i = 0$, $r_i' = 1$ if $v_i > 0$ and $r_i' = -1$ if $v_i < 0$. Define the map $\tau_r : \mathbb{R}^n \to \mathbb{R}^n$ such that τ_r maps v to the vector with ith entry $r_i' v_i$. Define c' and τ_c analogously. Now let M' be the matrix with (i,j)th entry $r_i' c_j' A_{ij}$. Since the entries of M are independent Gaussians, it follows that $M' \sim \mu \cdot \tau_r(r) \tau_c(c)^\top + N(0,1)^{\otimes n \times n}$. Now observe that since $S_{k_1'}$ and $S_{k_2'}$ are preserved by τ_r and τ_c, respectively, we have that $(u_{k_1'}, v_{k_2'}) = (\tau_r(u), \tau_c(v))$ where

$$(u, v) = \operatorname{argmax}_{(u,v) \in S_{k_1'} \times S_{k_2'}} \left\{ u^\top M' v \right\}$$

Now note that the mean matrix $\mu \cdot \tau_r(r) \tau_c(c)^\top$ of M' has all nonnegative entries. Furthermore, the entries in its support are at least $\mu \cdot \min_{(i,j) \in \operatorname{supp}(r) \times \operatorname{supp}(c)} |r_i c_j| \geq \frac{\mu}{k} = \rho$ by the definition of $\mathcal{V}_{n,k}$. Applying Lemma 9.19 now yields that with probability at least $1 - n^{-1}$, it follows that $(u,v) = (\mathbf{1}_{\operatorname{supp}(r)}, \mathbf{1}_{\operatorname{supp}(c)})$ or $(u,v) = (-\mathbf{1}_{\operatorname{supp}(r)}, -\mathbf{1}_{\operatorname{supp}(c)})$. This implies that $u_{k_1'} = r'$ and $v_{k_2'} = c'$ and thus are supported on all of $\operatorname{supp}(r)$ and $\operatorname{supp}(c)$, respectively.

We now will show that r' and c' are marked by the test in Step 2b with high probability. If $i \in \operatorname{supp}(r)$, then $\sum_{j=1}^n r_i' c_j' B_{ij} \sim N\left(\sum_{j=1}^n \mu \cdot |r_i| \cdot |c_j|, k_2\right)$. Since $\sum_{j=1}^n \mu \cdot |r_i| \cdot |c_j| \geq \mu \cdot \frac{k_2}{k} = k_2 \rho$ by the definition of $\mathcal{V}_{n,k}$. Therefore it follows by the same Gaussian tail bound as above that

$$\mathbb{P}\left[\sum_{j=1}^n r_i' c_j' B_{ij} < \frac{1}{2} k_2 \rho\right] \leq \mathbb{P}\left[N(0, k_2) < -\frac{1}{2} k_2 \rho\right] \leq n^{-3}$$

Furthermore, if $i \notin \operatorname{supp}(r)$ it follows that $\sum_{j=1}^n r_i' c_j' B_{ij} \sim N(0, k_2)$ and thus

$$\mathbb{P}\left[\sum_{j=1}^n r_i' c_j' B_{ij} \geq \frac{1}{2} k_2 \rho\right] = \mathbb{P}\left[N(0, k_2) \geq \frac{1}{2} k_2 \rho\right] \leq n^{-3}$$

Now a union bound yields that

$$\mathbb{P}\left[\operatorname{supp}(r) = \left\{ i \in [n] : \sum_{j=1}^n r_i' c_j' B_{ij} \geq \frac{1}{2} k_2 \rho \right\}\right] \geq 1 - \sum_{i \in \operatorname{supp}(r)} \mathbb{P}\left[\sum_{j=1}^n r_i' c_j' B_{ij} < \frac{1}{2} k_2 \rho\right]$$

$$- \sum_{i \notin \operatorname{supp}(r)} \mathbb{P}\left[\sum_{j=1}^n r_i' c_j' B_{ij} \geq \frac{1}{2} k_2 \rho\right]$$

$$\geq 1 - k_1 \cdot n^{-3} - (n - k_1) n^{-3} = 1 - n^{-2}$$

An identical argument yields that

$$\mathbb{P}\left[\operatorname{supp}(c) = \left\{ j \in [n] : \sum_{i=1}^n r_i' c_j' B_{ij} \geq \frac{1}{2} k_2 \rho \right\}\right] \geq 1 - n^{-2}$$

A union bound now yields that (r', c') is marked by the test in Step 2b with probability at least $1 - 2n^{-2}$. A further union bound now yields that with probability at least $1 - 2n^{-1} - 2n^{-2}$, the following events all hold:

- $\text{supp}(u_{k_1}) \subseteq \text{supp}(r)$ and $\text{supp}(v_{k_2}) \subseteq \text{supp}(c)$ for all $k_1, k_2 \in [c_1 k, k]$;

- $(u_{k_1'}, v_{k_2'}) = (r', c')$; and

- (r', c') is marked when input to the test in Step 2b.

These three events imply that that the vector (r', c') is marked in ROS-SEARCH and hence the maximum of $|\text{supp}(u_{k_1})| + |\text{supp}(v_{k_2})|$ over marked pairs (u_{k_1}, v_{k_2}) is $k_1' + k_2'$. Furthermore, first event implies that any marked pair (u_{k_1}, v_{k_2}) with $|\text{supp}(u_{k_1})| + |\text{supp}(v_{k_2})| = k_1' + k_2'$ must satisfy that $\text{supp}(u_{k_1}) = \text{supp}(r)$ and $\text{supp}(v_{k_2}) = \text{supp}(v)$. Thus the algorithm correctly recovers the supports of r and c with probability at least $1 - 2n^{-1} - 2n^{-2}$, proving the theorem. □

The last theorem of this section gives a simple test solving the detection problems SSBM_D, ROS_D and SSW_D asymptotically down to their information-theoretic limits. More precisely, this test solves SSBM_D if $\rho \gtrsim \frac{1}{\sqrt{k}}$ and by setting $\rho = \mu/k$, solves ROS_D and SSW_D as long as $\mu \gtrsim \sqrt{k}$.

Theorem 9.21. *Suppose that $c_1 \in (0, 1)$ is a fixed constant and let R_+ and R_- be disjoint subsets of $[n]$ with $c_1 k \leq |R_+| + |R_-| \leq k$. Let C_+ and C_- be defined similarly. Let $M \in \mathbb{R}^{n \times n}$ be a random matrix with independent sub-Gaussian entries with sub-Gaussian norm at most 1 such that:*

- $\mathbb{E}[M_{ij}] \geq \rho$ *if* $(i, j) \in R_+ \times C+$ *or* $(i, j) \in R_- \times C_-$;

- $\mathbb{E}[M_{ij}] \leq -\rho$ *if* $(i, j) \in R_+ \times C_-$ *or* $(i, j) \in R_- \times C_+$; *and*

- $\mathbb{E}[M_{ij}] = 0$ *if* $(i, j) \notin (R_+ \cup R_-) \times (C_+ \cup C_-)$.

There is a constant $c_2 > 0$ such that if $k\rho \geq c_2\sqrt{\log n}$, then $\max_{(u,v) \in S_k^2} u^\top M v \geq \frac{1}{2} c_1^2 k^2 \rho$ with probability at least $1 - en^{-1}$. If $\mathbb{E}[M_{ij}] = 0$ for all $(i, j) \in [n]^2$, then there is some constant $c_3 > 0$ such that if $k\rho^2 \geq c_2 \log n$ then $\max_{(u,v) \in S_k^2} u^\top M v < \frac{1}{2} c_1^2 k^2 \rho$ with probability at least $1 - en^{-1}$.

Proof. Let $u \in S_k$ satisfy that $u_i = 1$ for each $i \in R_+$ and $u_i = -1$ for each $i \in R_-$. Similarly let $v \in S_k$ satisfy that $v_i = 1$ for each $i \in C_+$ and $v_i = -1$ for each $i \in C_-$. It follows that

$$\sum_{i,j=1}^{n} u_i v_j \cdot \mathbb{E}[M_{ij}] \geq (|R_+| + |R_-|)(|C_+| + |C_-|)\rho \geq c_1^2 k^2 \rho$$

Now note that if $c \cdot c_1^4 k^2 \rho^2 \geq 4 \log n$, then

$$\mathbb{P}\left[u^\top M v < \frac{1}{2} c_1^2 k^2 \rho\right] = \mathbb{P}\left[\sum_{i,j=1}^{n} u_i v_j \left(M_{ij} - \mathbb{E}[M_{ij}]\right) < \frac{1}{2} c_1^2 k^2 \rho - \sum_{i,j=1}^{n} u_i v_j \cdot \mathbb{E}[M_{ij}]\right]$$

$$\leq \mathbb{P}\left[\sum_{i,j=1}^{n} u_i v_j \left(M_{ij} - \mathbb{E}[M_{ij}]\right) < -\frac{1}{2} c_1^2 k^2 \rho\right]$$

$$\leq e \cdot \exp\left(-\frac{c \cdot \left(c_1^2 k^2 \rho\right)^2}{4k^2}\right) \leq en^{-1}$$

for some constant $c > 0$ by Hoeffding's inequality for sub-Gaussian random variables as in Proposition 5.10 in [Ver10]. This proves the first claim of the theorem.

Now suppose that $\mathbb{E}[M_{ij}] = 0$ for all $(i,j) \in [n]^2$. For a fixed pair $(u,v) \in S_k^2$, we have by the same application of Hoeffding's inequality that

$$\mathbb{P}\left[u^\top M v \geq \frac{1}{2}c_1^2 k^2 \rho\right] \leq e \cdot \exp\left(-\frac{c \cdot c_1^4 k^2 \rho^2}{4}\right)$$

Now note that $|S_k| = 2^k \binom{n}{k} \leq (2n)^k$. Thus a union bound yields that

$$\mathbb{P}\left[\max_{(u,v) \in S_k^2} u^\top M v \geq \frac{1}{2}c_1^2 k^2 \rho\right] \leq |S_k| \cdot e \cdot \exp\left(-\frac{c \cdot c_1^4 k^2 \rho^2}{4}\right) \leq e \cdot \exp\left(k \log(2n) - \frac{c \cdot c_1^4 k^2 \rho^2}{4}\right) \leq e n^{-1}$$

if $c \cdot c_1^4 k^2 \rho^2 \geq 4k \log(2n) + 4 \log n$. This completes the proof of the theorem. $\qquad\square$

Polynomial-Time Algorithms. The proofs in this section are given in Appendix F. The polynomial-time algorithms achieving the tight boundary for the problems in this section are very different in the regimes $k \lesssim \sqrt{n}$ and $k \gtrsim \sqrt{n}$. When $k \lesssim \sqrt{n}$, the simple linear-time algorithm thresholding the absolute values of the entries of the data matrix matches the planted clique lower bounds in Theorem 2.1 up to polylogarithmic factors. This is captured in the following simple theorem, which shows that if $\mu \gtrsim k$ then this algorithm solves ROS_R and SSW_R. Setting $u = v = 0$ in the Theorem yields that the test outputting H_1 if $\max_{i,j\in[n]^2}|M_{ij}| > \sqrt{6\log n}$ solves the detection variants ROS_D and SSW_D if $\mu \gtrsim k$.

Theorem 9.22. *Let $M \sim \mu \cdot uv^\top + N(0,1)^{\otimes n \times n}$ where $u,v \in \mathcal{V}_{n,k}$ and suppose that $\mu \geq 2k\sqrt{6\log n}$, then the set of (i,j) with $|M_{ij}| > \sqrt{6\log n}$ is exactly $\mathrm{supp}(u) \times \mathrm{supp}(v)$ with probability at least $1 - O(n^{-1})$.*

In the regime $k \gtrsim \sqrt{n}$, the spectral projection algorithm in Figure 21 from [CLR15] achieves exact recovery in ROS_R down to the planted clique lower bounds in Theorem 2.1. This method is described in Algorithm 1 and its guarantees established in Lemma 1 of [CLR15]. Although it is stated as a recovery algorithm for a sub-Gaussian variant of BC_R, as indicated in Remark 2.1 in [CLR15], the guarantees of the algorithm extend more generally to rank one perturbations of a sub-Gaussian noise matrix. Our model of ROS_R does not exactly fit into the extended model in Remark 2.1, but the argument in Lemma 1 can be applied to show ROS-Spectral-Projection solves ROS_R. The details of this argument are show below. For brevity, we omit parts of the proof that are identical to [CLR15].

Theorem 9.23. *Suppose that $M = \mu \cdot rc^\top + N(0,1)^{\otimes n \times n}$ where $r,c \in \mathcal{V}_{n,k}$. There is a constant $C_1 > 0$ such that if $\mu \geq C_1(\sqrt{n} + \sqrt{k\log n})$ then the algorithm ROS-Spectral-Projection correctly outputs $\mathrm{supp}(r)$ and $\mathrm{supp}(c)$ with probability at least $1 - 2n^{-C_2} - 2\exp(-2C_2 n)$ for some constant $C_2 > 0$.*

A simple singular value thresholding algorithm solves ROS_D if $\mu \gtrsim \sqrt{n}$ and is comparatively simpler to analyze. As previously mentioned, this algorithm also solves SSW_D since any instance of SSW_D is an instance of ROS_D. Let $\sigma_1(M)$ denote the largest singular value of the matrix M.

Theorem 9.24. *Suppose that M is an instance of $\mathrm{ROS}_D(n,k,\mu)$. There is a constant $C_1 > 0$ such that if $\mu > 4\sqrt{n} + 2\sqrt{2\log n}$ then the algorithm that outputs H_1 if $\sigma_1(M) \geq \frac{1}{2}\mu$ and H_0 otherwise has Type I+II error tending to zero as $n \to \infty$.*

77

Algorithm ROS-SPECTRAL-PROJECTION

Inputs: Matrix $M \in \mathbb{R}^{n \times n}$

1. Let $G \sim N(0,1)^{\otimes n \times n}$ and let $A = \frac{1}{\sqrt{2}}(M + G)$ and $B = \frac{1}{\sqrt{2}}(M - G)$

2. Compute the top left and right singular vectors U and V of A

3. Sort the n entries of $U^\top B$ in decreasing order and separate the entries into two clusters R and $[n]\backslash R$ at the largest gap between consecutive values

4. Sort the n entries of BV in decreasing order and separate the entries into two clusters C and $[n]\backslash C$ at the largest gap between consecutive values

5. Output R and C

Figure 21: Algorithm for sparse rank-1 submatrix recovery from [CLR15] and in Theorem 9.23.

To complete this section, we give a simple spectral thresholding algorithm for SSBM_D if $\rho \gtrsim \frac{\sqrt{n}}{k}$. Let $\lambda_1(X)$ denote the largest eigenvalue of X where X is a square symmetric matrix.

Theorem 9.25. *Let G be an instance of $\text{SSBM}_D(n, k, q, \rho)$ where $q = \Theta(1)$, $k = \Omega(\sqrt{n})$ and $\rho \geq \frac{6\sqrt{n}}{k}$. Let $A \in \mathbb{R}^{n \times n}$ be the adjacency matrix of G and $J \in \mathbb{R}^{n \times n}$ be the matrix with zeros on its diagonal and ones elsewhere. Then the algorithm that outputs H_1 if $\lambda_1(A - qJ) \geq 2\sqrt{n}$ and H_0 otherwise has Type I+II error tending to zero as $n \to \infty$.*

9.3 Sparse PCA and Biased Sparse PCA

Information-Theoretic Lower Bounds. Other than for Theorem 9.31, the proofs in these sections on sparse PCA and biased sparse PCA are given in Appendix F. In [BR13b], information-theoretic lower bounds for $\text{SPCA}_D(n, k, d, \theta)$ were considered and it was shown that if

$$\theta \leq \min \left\{ \frac{1}{\sqrt{2}}, \sqrt{\frac{k \log(1 + o(d/k^2))}{n}} \right\}$$

then the optimal Type I+II error of any algorithm for SPCA_D tends to 1 as $n \to \infty$. The proof of this information-theoretic lower bound follows a similar χ^2 and MGF argument as in the previous section. If $d \ll k^2$, then this bound degrades to $\theta = o\left(\frac{d}{kn}\right)$. When $d = \Theta(n)$, there is a gap between this information-theoretic lower bound and the best known algorithm based on thresholding the k-sparse eigenvalue of the empirical covariance matrix, which only requires $\theta \gtrsim \sqrt{k/n}$. This information-theoretic lower bound was improved in [CMW15] to match the algorithmic upper bound. The following is their theorem in our notation.

Theorem 9.26 (Proposition 2 in [CMW15]). *Let $\beta_0 \in (0, 1/36)$ be a constant. Suppose that $X = (X_1, X_2, \ldots, X_n)$ is an instance of $\text{SPCA}_D(n, k, d, \theta)$ where under H_1, the planted vector v is chosen uniformly at random from all k-sparse unit vectors with nonzero coordinates equal to $\pm \frac{1}{\sqrt{k}}$. If it holds that*

$$\theta \leq \min \left\{ 1, \sqrt{\frac{\beta_0 k}{n} \log\left(\frac{ed}{k}\right)} \right\}$$

then there is a function $w : (0, 1/36) \to (0, 1)$ satisfying that $\lim_{\beta_0 \to 0^+} w(\beta_0) = 0$ and

$$d_{TV}(\mathcal{L}_{H_0}(X), \mathcal{L}_{H_1}(X)) \leq w(\beta_0)$$

In particular if $\theta \ll \sqrt{k/n}$, then this inequality eventually applies for every $\beta_0 > 0$ and $d_{\mathrm{TV}}(\mathcal{L}_{H_0}(X), \mathcal{L}_{H_1}(X)) \to 0$, establishing the desired information-theoretic lower bound for SPCA_D. We now show that BSPCA_D satisfies a weaker lower bound. The prior on the planted vector v used in Theorem 5.1 of [BR13b] to derive the suboptimal lower bound for SPCA_D shown above placed entries equal to $1/\sqrt{k}$ on a random k-subset of the d coordinates of v. Therefore their bound also applies to BSPCA_D. In order strengthen this to the optimal information-theoretic lower bound for BSPCA_D, we need apply Lemma 9.3 in place of the weaker hypergeometric squared MGF bounds used in [BR13b]. To simplify the proof, we will use the following lemma from [CMW15].

Lemma 9.27 (Lemma 7 in [CMW15]). *Let v be a distribution on $d \times d$ symmetric random matrices M such that $\|M\| \leq 1$ almost surely. If $\mathbb{E}_v[N(0, I_d + M)^{\otimes n}] = \int N(0, I_d + M)^{\otimes n} dv(M)$, then*

$$\chi^2\left(\mathbb{E}_v[N(0, I_d + M)^{\otimes n}], N(0, I_d)^{\otimes n}\right) + 1 = \mathbb{E}\left[\det(I_d - M_1 M_2)^{-n/2}\right]$$

where M_1 and M_2 are independently drawn from v.

Applying this lemma with the hypergeometric squared MGF bounds in Lemma 9.3 yields the following information-theoretic lower bound for BSPCA_D.

Theorem 9.28. *Suppose that $X = (X_1, X_2, \ldots, X_n)$ is an instance of $\mathrm{BSPCA}_D(n, k, d, \theta)$ where under H_1, the planted vector v is chosen uniformly at random from all k-sparse unit vectors in $\mathbb{R}^d$ with nonzero coordinates equal to $\frac{1}{\sqrt{k}}$. Suppose it holds that $\theta \leq 1/\sqrt{2}$ and*

$$\theta \leq \min\left\{\sqrt{\frac{\beta_0 k}{n} \log\left(\frac{ed}{k}\right)}, \sqrt{\frac{\beta_0 d^2}{nk^2}}\right\}$$

for some $0 < \beta_0 < (16e)^{-1}$. Then there is a function $w : (0, 1) \to (0, 1)$ satisfying that $\lim_{\beta_0 \to 0^+} w(\beta_0) = 0$ and

$$d_{TV}(\mathcal{L}_{H_0}(X), \mathcal{L}_{H_1}(X)) \leq w(\beta_0)$$

We remark that this same proof technique applied to the ensemble of k-sparse unit vectors v chosen uniformly at random from those with nonzero coordinates equal to $\pm 1/\sqrt{k}$ with Lemma 9.17 proves Theorem 9.26. This difference in the lower bounds resulting from these two choices of ensembles illustrates the information-theoretic difference between SPCA_D and BSPCA_D. We now will show information-theoretic lower bounds for the weak recovery problems SPCA_{WR} and BSCPA_{WR}. The argument presented here is similar to the proof of Theorem 3 in [WBS16]. For this argument, we will need a variant of the Gilbert-Varshamov lemma and generalized Fano's lemma as in [WBS16]. Given two $u, v \in \mathbb{R}^d$, let $d_H(u, v)$ denote the Hamming distance between u and v.

Lemma 9.29 (Gilbert-Varshamov, Lemma 4.10 in [Mas07]). *Suppose that $\alpha, \beta \in (0, 1)$ and $k \leq \alpha\beta d$. Let*

$$\rho = \frac{\alpha}{\log(\alpha\beta)^{-1}}(\beta - \log\beta - 1)$$

Then there is a subset S of $\{v \in \{0, 1\}^d : \|v\|_0 = k\}$ of size at least $\left(\frac{d}{k}\right)^{\rho k}$ such that for any two $u, v \in S$ with $u \neq v$, it holds that $d_H(u, v) \geq 2(1 - \alpha)k$.

Lemma 9.30 (Generalized Fano's Lemma, Lemma 3 in [Yu97]). *Let $P_1, P_2, \ldots, P_M$ be probability distributions on a measurable space $(\mathcal{X}, \mathcal{B})$ and assume that $d_{KL}(P_i, P_j) \leq \beta$ for all $i \neq j$. Any measurable function $\phi : \mathcal{X} \to \{1, 2, \ldots, M\}$ satisfies that*

$$\max_{1 \leq i \leq M} P_i(\phi \neq i) \geq 1 - \frac{\beta + \log 2}{\log M}$$

With these two lemmas, we now will show that the weak recovery problems SPCA_{WR} and BSCPA_{WR} cannot be solved if $\theta \lesssim \sqrt{k/n}$, matching Theorem 2.1. Note that in the theorem below, $\phi(X)$ and $\text{supp}(v)$ have size k for all $X \in \mathbb{R}^{d \times n}$ and $v \in S$. Therefore it holds that $|\phi(X)\Delta\text{supp}(v)| = 2k - 2|\phi(X) \cap \text{supp}(v)|$ for all such X and v.

Theorem 9.31. *Fix positive integers n, k, d and real numbers $\theta > 0$ and a constant $\epsilon \in (0, 1)$ such that $k \leq \epsilon d/4$. Let P_v denote the distribution $N(0, I_d + \theta v v^\top)$ and let S be the set of all k-sparse unit vectors with nonzero entries equal to $1/\sqrt{k}$. If*

$$\frac{n\theta^2}{2(1 + \theta)} + \log 2 \leq \frac{\epsilon^2}{2 \log 4\epsilon^{-1}} \cdot k \log\left(\frac{d}{k}\right)$$

then for any function $\phi : \mathbb{R}^{d \times n} \to \binom{[n]}{k}$, it holds that

$$\min_{v \in S} \mathbb{E}_{X \sim P_v^{\otimes n}} [|\phi(X) \cap \text{supp}(v)|] \leq \left(\frac{1}{2} + \epsilon\right) k$$

Proof. Note that $d_{\text{KL}}(N(0, \Sigma_0), N(0, \Sigma_1)) = \frac{1}{2}\left[\text{Tr}\left(\Sigma_1^{-1}\Sigma_0\right) - d + \ln\frac{\det(\Sigma_1)}{\det(\Sigma_0)}\right]$ for any positive semidefinite $\Sigma_0, \Sigma_1 \in \mathbb{R}^{d \times d}$. Therefore for any $\|u\|_2 = \|v\|_2 = 1$,

$$d_{\text{KL}}(P_u^{\otimes n}, P_v^{\otimes n}) = n \cdot d_{\text{KL}}(P_u, P_v) = \frac{n}{2} \cdot \text{Tr}\left(\left(I_d + \theta u u^\top\right)^{-1}\left(I_d + \theta v v^\top\right) - I_d\right)$$

$$= \frac{n\theta}{2} \cdot \text{Tr}\left(\left(I_d + \theta u u^\top\right)^{-1}\left(v v^\top - u u^\top\right)\right)$$

$$= \frac{n\theta}{2} \cdot \text{Tr}\left(\left(I_d - \frac{\theta}{1 + \theta} \cdot u u^\top\right)\left(v v^\top - u u^\top\right)\right)$$

$$= \frac{n\theta}{2} \cdot \text{Tr}\left(v v^\top - u u^\top - \frac{\theta}{1 + \theta} \cdot \langle u, v \rangle \cdot u v^\top + \frac{\theta}{1 + \theta} \cdot u u^\top\right)$$

$$= \frac{n\theta^2}{4(1 + \theta)} \cdot \|u - v\|_2^2 \leq \frac{n\theta^2}{2(1 + \theta)}$$

since $\det(I_d + \theta u u^\top) = \det(I_d + \theta v v^\top) = 1 + \theta$. Let S_0 be the subset from Gilbert-Varshamov's lemma applied with $\alpha = \epsilon$ and $\beta = \frac{1}{4}$. It follows that

$$\log|S_0| \geq \frac{\epsilon}{2 \log 4\epsilon^{-1}} \cdot k \log\left(\frac{d}{k}\right)$$

For each $u \in S$, let $\hat{u}$ be an element of S_0 such that $|\text{supp}(u) \cap \text{supp}(\hat{u})|$ is maximal. Let $\hat{\phi}$ denote the function that maps $X \in \mathbb{R}^{d \times n}$ to $\text{supp}(\hat{u})$ where $u = \mathbf{1}_{\phi(X)}$. Suppose that $v \in S_0$ is such that $v \neq \hat{u}$. Observe by the triangle inequality that

$$4k - 2|\phi(X) \cap \text{supp}(v)| - 2|\hat{\phi}(X) \cap \phi(X)| = d_H(\hat{u}, u) + d_H(u, v) \geq d_H(\hat{u}, v) \geq 2(1 - \epsilon)k$$

Rearranging and using the fact that $|\phi(X) \cap \text{supp}(v)| \le |\hat{\phi}(X) \cap \phi(X)|$ yields that

$$k - |\phi(X) \cap \text{supp}(v)| \ge \frac{1}{2}(1 - \epsilon)k$$

Note that $k - |\phi(X) \cap \text{supp}(v)| \ge 0$ is true for all v. Therefore if $v \in S_0$, we have

$$\mathbb{E}_{X \sim P_v^{\otimes n}}[k - |\phi(X) \cap \text{supp}(v)|] \ge \mathbb{P}_{X \sim P_v^{\otimes n}}\left[\hat{\phi}(X) \ne v\right] \cdot \frac{1}{2}(1 - \epsilon)k$$

Now observe by generalized Fano's Lemma,

$$
\begin{aligned}
\max_{v \in S} \mathbb{E}_{X \sim P_v^{\otimes n}}[k - |\phi(X) \cap \text{supp}(v)|] &\ge \max_{v \in S_0} \mathbb{E}_{X \sim P_v^{\otimes n}}[k - |\phi(X) \cap \text{supp}(v)|] \\
&\ge \frac{1}{2}(1 - \epsilon)k \cdot \max_{v \in S_0} \mathbb{P}_{X \sim P_v^{\otimes n}}\left[\hat{\phi}(X) \ne v\right] \\
&\ge \frac{1}{2}(1 - \epsilon)k \cdot \left[1 - \frac{\frac{n\theta^2}{2(1+\theta)} + \log 2}{\log |S_0|}\right] \\
&\ge \frac{1}{2}(1 - \epsilon)k \cdot (1 - \epsilon) \ge \left(\frac{1}{2} - \epsilon\right)k
\end{aligned}
$$

Rearranging completes the proof of the lemma. $\qquad\square$

Information-Theoretically Optimal Algorithms. Given a positive semidefinite matrix $M \in \mathbb{R}^{d \times d}$, let the k-sparse maximum eigenvalue and eigenvector be

$$\lambda_{\max}^k(M) = \max_{\|u\|_2 = 1, \|u\|_0 = k} u^\top M u \quad \text{and} \quad v_{\max}^k(M) = \arg\max_{\|u\|_2 = 1, \|u\|_0 = k} u^\top M u$$

Note that $\lambda_{\max}^k(M)$ and $v_{\max}^k(M)$ can be computed by searching over all principal $k \times k$ minors for the maximum eigenvalue and corresponding eigenvector. In [BR13a] and [WBS16], the k-sparse maximum eigenvalue and eigenvector of the empirical covariance matrix are shown to solve sparse PCA detection and estimation in the ℓ_2 norm under general distributions satisfying a restricted covariance concentration condition. Specializing these results to Gaussian formulations of sparse PCA yields the following theorems. Let $L(u, v) = \sqrt{1 - \langle u, v \rangle}$ for $u, v \in \mathbb{R}^d$ with $\|u\|_2 = \|v\|_2 = 1$. The statement of Theorem 2 in [BR13a] is for $k \ll \sqrt{d}$, but the same argument also shows the result for $k \ll d$. Note that the following theorem applies to BSPCA_D since any instance of BSPCA_D is also an instance of SPCA_D.

Theorem 9.32 (Theorem 2 in [BR13a]). *Suppose that $X = (X_1, X_2, \ldots, X_n)$ is an instance of $\text{SPCA}_D(n, k, d, \theta)$ and let $\hat{\Sigma}$ be the empirical covariance matrix of X. If $\theta, \delta \in (0, 1)$ are such that*

$$\theta > 15\sqrt{\frac{k \log\left(\frac{3ed}{k\delta}\right)}{n}}$$

then the algorithm that outputs H_1 if $\lambda_{\max}^k(\hat{\Sigma}) > 1 + 8\sqrt{\frac{k \log\left(\frac{3ed}{k\delta}\right)}{n}}$ and H_0 otherwise has Type I+II error at most δ.

Theorem 9.33 (Theorem 2 in [WBS16]). *Suppose that k, d and n are such that $2k \log d \le n$. Let P_v denote the distribution $N(0, I_d + \theta v v^\top)$ and given some $X = (X_1, X_2, \ldots, X_n) \sim P_v^{\otimes n}$, let $\hat{\Sigma}(X)$ be the empirical covariance matrix of X. It follows that*

$$\sup_{v \in \mathcal{V}_{d,k}} \mathbb{E}_{X \sim P_v} L\left(v_{\max}^k\left(\hat{\Sigma}(X)\right), v\right) \le 7\sqrt{\frac{k \log d}{n\theta^2}}$$

81

The latter result on estimation in the ℓ_2 norm yields a weak recovery algorithm for SPCA_{WR} and BSPCA_{WR} by thresholding the entries of $v_{\max}^k(\hat{\Sigma})$, as in the following theorem. If $\frac{k \log d}{n\theta^2} \to 0$ then this algorithm achieves weak recovery.

Theorem 9.34. *Suppose that k, d and n are such that $2k \log d \leq n$. Let $S(X) \subseteq [d]$ be the set of coordinates of $v_{\max}^k \left(\hat{\Sigma}(X) \right)$ with magnitude at least $\frac{1}{2\sqrt{k}}$. It follows that*

$$\sup_{v \in \mathcal{V}_{d,k}} \mathbb{E}_{X \sim P_v} \left[\frac{1}{k} \left| S(X) \Delta \mathrm{supp}(v) \right| \right] \leq 56 \sqrt{\frac{2k \log d}{n\theta^2}}$$

We next analyze an algorithm thresholding the sum of the entries of the empirical covariance matrix. For instances of BSPCA_D, this sum test solves the detection problem when $\theta \gtrsim \frac{\sqrt{n}}{k}$ in the regime $d = \Theta(n)$ and becomes optimal when $k \gtrsim n^{2/3}$. Recall that $\delta = \delta_{\mathrm{BSPCA}} > 0$ is the constant in the definition of $\mathcal{BV}_{d,k}$.

Theorem 9.35. *Suppose that $X = (X_1, X_2, \ldots, X_n)$ is an instance of $\mathrm{BSPCA}_D(n, k, d, \theta)$ and let $\hat{\Sigma}(X)$ be the empirical covariance matrix of X. Suppose that $2\delta^2 k\theta \leq d$ and $\frac{nk^2\theta^2}{d^2} \to \infty$ as $n \to \infty$. Then the test that outputs H_1 if $\mathbf{1}^\top \hat{\Sigma}(X)\mathbf{1} > d + 2\delta^2 k\theta$ and H_0 otherwise has Type I+II error tending to zero as $n \to \infty$.*

Polynomial-Time Algorithms. As shown in [BR13a], SPCA_D and BSPCA_D can be solved with a semidefinite program in the regime $k \lesssim \sqrt{n}$. Their algorithm computes a semidefinite relaxation of the maximum k-sparse eigenvalue, first forming the empirical covariance matrix $\hat{\Sigma}(X)$ and solving the convex program

$$\mathrm{SDP}(X) = \max_Z \quad \mathrm{Tr} \left(\hat{\Sigma}(X)Z \right)$$
$$\text{s.t.} \quad \mathrm{Tr}(Z) = 1, |Z|_1 \leq k, Z \succeq 0$$

Thresholding $\mathrm{SDP}(X)$ yields a detection algorithm for SPCA_D with guarantees captured in the following theorem. In [BR13a], a more general model is considered with sub-Gaussian noise. We specialize the more general theorem in [BR13a] to our setup.

Theorem 9.36 (Theorem 5 in [BR13a]). *Suppose that $\delta \in (0, 1)$. Let $X = (X_1, X_2, \ldots, X_n)$ be an instance of $\mathrm{SPCA}_D(n, k, d, \theta)$ and suppose that $\theta \in [0, 1]$ satisfies that*

$$\theta \geq 23 \sqrt{\frac{k^2 \log(d^2/\delta)}{n}}$$

Then the algorithm that outputs H_1 if $\mathrm{SDP}(X) \geq 16\sqrt{\frac{k^2 \log(d^2/\delta)}{n}} + \frac{1}{\sqrt{n}}$ and H_0 otherwise has Type I+II error at most δ.

In [WBS16], a semidefinite programming approach is shown to solve the sparse PCA estimation task under the ℓ_2 norm. As in the proof of Theorem 9.34, this yields a weak recovery algorithm for SPCA_{WR} and BSPCA_{WR}, achieving the tight barrier when $k \lesssim \sqrt{n}$. The SDP algorithm in [WBS16] is shown in Figure 22 and its guarantees specialized to the case of Gaussian data are in the following theorem.

Algorithm SPCA-SDP

Inputs: $X = (X_1, X_2, \ldots, X_n) \in \mathbb{R}^{d \times n}$

1. Compute the empirical covariance matrix $\hat{\Sigma} = \hat{\Sigma}(X)$ and set

$$\epsilon = \frac{\log d}{4n} \quad \text{and} \quad \lambda = 4\sqrt{\frac{\log d}{n}}$$

2. Let $f : \mathbb{R}^{d \times d} \to \mathbb{R}$ be

$$f(M) = \text{Tr}\left(\hat{\Sigma} M\right) - \lambda \|M\|_1$$

and compute an ϵ-maximizer $\hat{M}^\epsilon$ of $f(M)$ subject to the constraints that M is symmetric, $M \succeq 0$ and $\text{Tr}(M) = 1$

3. Output $v_{\text{SDP}} = \arg\max_{\|u\|_2 = 1} u^\top \hat{M}^\epsilon u$

Figure 22: SDP algorithm for weak ℓ_2 estimation in sparse PCA from [WBS16] and in Theorem 9.37.

Theorem 9.37 (Theorem 5 in [WBS16]). *Suppose that k, d and n are such that $4 \log d \leq n \leq k^2 d^2 \theta^{-2} \log d$ and $0 < \theta \leq k$. Let P_v denote the distribution $N(0, I_d + \theta vv^\top)$ and given some $X = (X_1, X_2, \ldots, X_n) \sim P_v^{\otimes n}$, let $v_{\text{SDP}}(X)$ be the output of SPCA-SDP applied to X. Then*

$$\sup_{v \in \mathcal{V}_{d,k}} \mathbb{E}_{X \sim P_v} L\left(v_{\text{SDP}}(X), v\right) \leq \min\left((16\sqrt{2} + 2)\sqrt{\frac{k^2 \log d}{n\theta^2}}, 1\right)$$

Using an identical argument to Theorem 9.34, we obtain the following theorem.

Theorem 9.38. *Suppose that k, d and n are such that $4 \log d \leq n \leq k^2 d^2 \theta^{-2} \log d$ and $0 < \theta \leq k$. Let $S(X) \subseteq [d]$ be the set of coordinates of $v_{\text{SDP}}(X)$ with magnitude at least $\frac{1}{2\sqrt{k}}$. It follows that*

$$\sup_{v \in \mathcal{V}_{d,k}} \mathbb{E}_{X \sim P_v}\left[\frac{1}{k}|S(X)\Delta\text{supp}(v)|\right] \leq 8\sqrt{2} \cdot \min\left((16\sqrt{2} + 2)\sqrt{\frac{k^2 \log d}{n\theta^2}}, 1\right)$$

To establish the polynomial-time upper bounds for SPCA and BSPCA in Theorem 2.1, it now suffices to consider $k \gtrsim \sqrt{n}$. First consider the detection problems SPCA_D and BSPCA_D. The following theorem establishes that a spectral algorithm directly applied to the empirical covariance matrix solves SPCA_D when $\theta \gtrsim 1$.

Theorem 9.39. *Suppose that $X = (X_1, X_2, \ldots, X_n)$ is an instance of $\text{SPCA}_D(n, k, d, \theta)$ and let $\hat{\Sigma}(X)$ be the empirical covariance matrix of X. Suppose that $d \leq cn$ for some constant $c > 0$, $d \to \infty$ and $n(1 + \theta)^{-2} \to \infty$ as $n \to \infty$ and it holds that $\theta > 4\sqrt{c}$. Then the test that outputs H_1 if $\lambda_1(\hat{\Sigma}(X)) > 1 + 2\sqrt{c}$ and H_0 otherwise has Type I+II error tending to zero as $n \to \infty$.*

The algorithm summing the entries of the empirical covariance matrix in Theorem 9.35 runs in polynomial time and shows that BSPCA_D can be solved in polynomial time as long as $\theta \gtrsim \frac{\sqrt{n}}{k}$.

Note that this algorithm gives an upper bound matching Theorem 2.1 and can detect smaller signal levels θ in BSPCA_D than the spectral algorithm.

For the recovery problem when $k \gtrsim \sqrt{n}$, the spectral algorithm considered in Theorem 1.1 of [KNV15] achieves the upper bound in Theorem 2.1 for the exact recovery problems SPCA_R and BSPCA_R. As given in [KNV15], this spectral algorithm is not adaptive to the support size of the planted vector and assumes that the planted sparse vector has nonzero entries of the form $\pm 1/\sqrt{k}$. We mildly adapt this algorithm to only require that the planted vector is in $\mathcal{V}_{d,k}$. The proof of its correctness follows a similar argument as Theorem 1.1 in [KNV15]. We omit details that are identical for brevity.

Theorem 9.40. *Suppose that k, d and n are such that $k, d \to \infty$, $\frac{d}{n} \to c$ and $\frac{k \log d}{n} \to 0$ as $n \to \infty$ for some constant $c > 0$. Let $X = (X_1, X_2, \ldots, X_n) \sim P_v^{\otimes n}$ and let $\hat{v}$ be the leading eigenvector of $\hat{\Sigma}(X)$. Let $S \subseteq [d]$ be the set of i such that $|\hat{v}_i|^4 \geq \frac{\log d}{kd}$. If $\theta > \sqrt{c}$ is fixed then $S = \text{supp}(v)$ with probability tending to one as $n \to \infty$.*

Note that the theorem statement assumes that $d/n \to c$ where $c > 1$. The algorithm can more generally accommodate inputs with $d = \Theta(n)$ by padding the input X with i.i.d. $N(0,1)$ entries so that $d/n \to c$ where $c > 1$.

10 Detection-Recovery Reductions

In this section, we show that our computational lower bounds for detection problems imply corresponding lower bounds for the recovery. The idea is to produce two instances of each problem that are coupled to have the same planted sparse structure but are conditionally independent given this structure. If there is a polynomial-time recovery algorithm that outputs a set S containing a constant fraction of the planted sparse structure, then we restrict to the indices in S yields an instance with a planted structure of linear size and apply a detection algorithm from Section 9. For each of the problems that we consider, this solves detection within a sub-polynomial factor of when detection first becomes information-theoretically possible. The reduction also always runs in polynomial time. Therefore our detection lower bounds also imply recovery lower bounds.

To carry out this argument, we first require methods of creating two such instances. We first give such a cloning method for PDS instances. This method is similar to the cloning map in PC-LIFTING from Section 4 but produces two copies rather than four. The proof is given in Appendix G.

Lemma 10.1. *Suppose that $S \subseteq [n]$ and $p, q \in (0, 1]$ are such that $p > q$. Also suppose that $P, Q \in [0, 1]$ are such that $Q \neq 0, 1$ and the quotients $\frac{P}{Q}$ and $\frac{1-P}{1-Q}$ are both between $\sqrt{\frac{1-p}{1-q}}$ and $\sqrt{\frac{p}{q}}$. If $G \sim G(n, S, p, q)$ and (G^1, G^2) is the output of PDS-CLONING applied to G with parameters p, q, P and Q, then $(G^1, G^2) \sim G(n, S, P, Q)^{\otimes 2}$. Furthermore, if $G \sim G(n, q)$ then $(G^1, G^2) \sim G(n, Q)^{\otimes 2}$.*

A similar argument as in Lemma 6.3 on GAUSSIAN-LIFTING yields the following lemma. The proof is also deferred to Appendix G.

Lemma 10.2. *If $M \sim \mathcal{L}\left(A + N(0,1)^{\otimes n \times n}\right)$ for any fixed matrix $A \in \mathbb{R}^{n \times n}$ and (M^1, M^2) is the output of GAUSSIAN-CLONING applied to M, then $(M^1, M^2) \sim \mathcal{L}\left(\frac{1}{\sqrt{2}} A + N(0,1)^{\otimes n \times n}\right)^{\otimes 2}$.*

With these two lemmas, we now overview how detection lower bounds imply partial recovery lower bounds for each problem that we consider. In the cases of sparse PCA and biased sparse PCA, rather than give a direct detection-recovery reduction, we outline modifications to our detection reductions that yield recovery reductions from planted clique.

Algorithm PDS-CLONING

Inputs: Graph $G \in \mathcal{G}_n$, parameters $p, q \in (0, 1]$ with $p > q$

1. Set $P = 1 - \sqrt{1 - p}$ and $Q = 1 - \sqrt{1 - q}$

2. For each pair $i, j \in [n]$ with $i < j$, independently generate $x^{ij} \in \{0, 1\}^2$ such that

   - If $\{i, j\} \in E(G)$, then generate x^{ij} from

     $$\mathbb{P}[x^{ij} = v] = \frac{1 - q}{p - q} \cdot P^{|v|_1}(1 - P)^{2 - |v|_1} - \frac{1 - p}{p - q} \cdot Q^{|v|_1}(1 - Q)^{2 - |v|_1}$$

   - If $\{i, j\} \notin E(G)$, then generate x^{ij} from

     $$\mathbb{P}[x^{ij} = v] = \frac{p}{p - q} \cdot Q^{|v|_1}(1 - Q)^{2 - |v|_1} - \frac{q}{p - q} \cdot P^{|v|_1}(1 - P)^{2 - |v|_1}$$

3. Construct the graphs G^1 and G^2 such that $\{i, j\} \in E(G^k)$ if and only if $x^{ij}_k = 1$

4. Output (G^1, G^2)

Algorithm GAUSSIAN-CLONING

Inputs: Matrix $M \in \mathbb{R}^{n \times n}$

1. Generate a matrix $G \sim N(0, 1)^{\otimes n \times n}$ with independent Gaussian entries

2. Compute the two matrices

   $$M^1 = \frac{1}{\sqrt{2}}(M + G) \quad \text{and} \quad M^2 = \frac{1}{\sqrt{2}}(M - G)$$

3. Output (M^1, M^2)

Figure 23: Cloning procedures in Lemmas 10.1 and 10.2.

Biclustering. Suppose that there is a randomized algorithm ϕ that solves BC_{PR}. More precisely, if M is an instance of $\mathrm{BC}_{PR}(n, k, \mu)$ with latent row and column supports S and T, then $\mathbb{E}[|\phi_1(M) \cap S|] = \Omega(k)$ and $\mathbb{E}[|\phi_2(M) \cap T|] = \Omega(k)$. Consider the detection algorithm for $\mathrm{BC}_D(n, k, \mu\sqrt{2})$ that applies GAUSSIAN-CLONING to the input to produce (M^1, M^2) and then takes the sum of the entries of M^2 restricted to the indices in $\phi_1(M^1) \times \phi_2(M^1)$. The algorithm then outputs H_1 if this sum is at least $k \cdot \tau(k)$ where $\tau(k) \to \infty$ arbitrarily slowly as $k \to \infty$. Since M^1 and M^2 are independent, it follows that M^2 is independent of $\phi(M^1)$. Under H_0, the sum is distributed as $N(0, k^2)$ which is less than $k \cdot \tau(k)$ with probability tending to 1 as $k \to \infty$. Under H_1, let S and T denote the latent row and column supports of the planted submatrix. If $k_1 = |\phi_1(M^1) \cap S|$ and $k_2 = |\phi_2(M^1) \cap T|$, then the sum is distributed as $N(\mu \cdot k_1 k_2, k^2)$. If $\mu \cdot k_1 k_2 \geq 2k \cdot \tau(k)$, then the algorithm outputs H_1 with probability tending to 1 as $k \to \infty$. If $\mu \geq \frac{2\tau(k)^3}{k}$ then $\mu \cdot k_1 k_2 < 2k \cdot \tau(k)$ is only possible

85

if either $k_1 \cdot \tau(k) < k$ or $k_2 \cdot \tau(k) < k$. By Markov's inequality, each of these events occurs with probability tending to zero as $k \to \infty$ over the randomness of ϕ. Therefore this algorithm has Type I+II error tending to zero as $k \to \infty$ if $\mu \geq \frac{2\tau(k)^3}{k}$, which is true for some τ as long as BC_R is information-theoretically possible.

In summary, if ϕ solves $\mathrm{BC}_{PR}(n, k, \mu)$ then there is a polynomial-time algorithm using ϕ as a blackbox that solves $\mathrm{BC}_D(n, k, \mu\sqrt{2})$. Hence our computational and information-theoretic lower bounds for BC_D imply partial recovery lower bounds in the same parameter regimes. We now give similar detection-recovery reductions using an initial cloning step for other problems.

Sparse Spiked Wigner and Rank-1 Submatrix. Similarly to biclustering, suppose that ϕ solves $\mathrm{ROS}_{PR}(n, k, \mu)$. Consider the detection algorithm for $\mathrm{ROS}_D(n, k, \mu\sqrt{2})$ that applies GAUSSIAN-CLONING to the input to produce (M^1, M^2), forms the $k \times k$ matrix W given by M^2 restricted to indices in $\phi_1(M^1) \times \phi_2(M^1)$ and outputs H_1 if and only if $\sigma_1(W) > 2\sqrt{k} + \sqrt{2\log k}$ where $\sigma_1(W)$ is the largest singular value of W. By Corollary 5.35 in [Ver10], the algorithm outputs H_0 under H_0 with probability at least $1 - 2k^{-1}$. Under H_1, let k_1 and k_2 be the sizes of the intersections of $\phi_1(M^1)$ and $\phi_2(M^2)$ with the row and column supports, respectively, of the rank-1 spike. By the definition of $\mathcal{V}_{n,k}$, it follows that the rank-1 spike has largest singular value at least $\mu k^{-1}\sqrt{k_1 k_2}$. By Weyl's interlacing inequality, it follows that $\sigma_1(W) \geq \mu k^{-1}\sqrt{k_1 k_2} - 2\sqrt{k} - \sqrt{2\log k}$. Suppose that $\mu \geq \tau(k)\sqrt{k}$ where $\tau(k) \to \infty$ as $k \to \infty$ arbitrarily slowly. It follows that $\mu k^{-1}\sqrt{k_1 k_2} < 4\sqrt{k} + 2\sqrt{2\log k} < 8\sqrt{k}$ implies that either $\tau(k) \cdot k_1 < 64k$ or $\tau(k) \cdot k_2 < 64k$. Both of these events occur with probability tending to zero as $k \to \infty$. It follows that this algorithm has Type I+II error tending to zero as $k \to \infty$ if $\mu \geq \tau(k)\sqrt{k}$, which aligns with the information theoretic lower bound on ROS_R of $\mu \gtrsim \sqrt{k}$. Similar reductions yield an analogous result for SROS and therefore also SSW.

Planted Independent Set and Planted Dense Subgraph. We first give a detection-recovery algorithm for PIS. Suppose that ϕ solves $\mathrm{PIS}_{PR}(n, k, q)$. Consider the detection algorithm that takes the complement graph of an input $\mathrm{PIS}_D(n, k, q)$, applies PDS-CLONING with $P = p' = 1$, $q' = 1 - q$ and $Q = 1 - q/2$ to produce (G^1, G^2) and then outputs H_1 if and only if $\overline{G^2}$ restricted to the vertices in $\phi(\overline{G^1})$ contains at most $\binom{k}{2}q - k\sqrt{q(1-q)\log k}$ edges. Here $\overline{G}$ denotes the complement of the graph G. First note that these inputs to PDS-CLONING are valid since $1 - P = 1 - p' = 0$ and
$$\frac{P}{Q} = \frac{1}{1 - q/2} = \frac{2}{1 + q'} \leq \sqrt{\frac{1}{q'}} = \sqrt{\frac{1}{1 - q}}$$

Under H_0, it follows that $\overline{G^1}$ and $\overline{G^2}$ are independent and distributed as $G(n, q/2)$. By the same applications of Bernstein's inequality as in Theorem 9.11, the algorithm outputs H_1 with probability tending to zero as $k \to \infty$. Under H_1, it follows that $\overline{G^1}$ and $\overline{G^2}$ are independent and distributed as $G_I(n, k, q/2)$. Let k_1 be the size of the intersection between $\phi(\overline{G^1})$ and the latent support of the planted independent set. The number of edges in $\overline{G^2}$ restricted to the vertices in $\phi(\overline{G^1})$ is distributed as $\mathrm{Bin}(\binom{k}{2} - \binom{k_1}{2}, q)$. If $\binom{k_1}{2}q \geq 2k\sqrt{q(1-q)\log k}$, then with high probability the algorithm outputs H_1. If $q \geq \frac{\tau(k)\log k}{k^2}$ where $\tau(k) \to \infty$ then it would have to hold that $\tau(k) \cdot k_1 \leq 4k$ for this inequality not to be true. However, this event occurs with probability tending to zero by Markov's inequality. This gives a reduction from $\mathrm{PIS}_{PR}(n, k, q)$ to $\mathrm{PIS}_D(n, k, 2q)$.

Similar detection-recovery reductions apply for planted dense subgraph. Fix some constant $w \in (0, 1/2)$. For an instance of $\mathrm{PDS}_D(n, k, p, q)$ with $p > q$, consider PDS-CLONING with $P =$

$\frac{wp+(1-w)q}{2}$ and $Q = q/2$. Note that these are valid inputs to PDS-CLONING when

$$\sqrt{\frac{1-p}{1-q}} \leq 1 - \frac{p-q}{2(1-q)} \leq 1 - \frac{w(p-q)}{2-q} = \frac{1-P}{1-Q} \quad \text{and} \quad \frac{P}{Q} = \frac{wp+(1-w)q}{q} = 1 + \frac{w(p-q)}{q} \leq \sqrt{\frac{p}{q}}$$

where the second inequality holds as long as $\frac{p-q}{q} \leq \frac{1-2w}{w^2}$. Taking w to be sufficient small covers the entire parameter space $p - q = O(q)$. Producing two independent copies of planted dense subgraph and thresholding the edge count now yields a detection-recovery reduction by the same argument as for planted independent set.

Sparse PCA and Biased Sparse PCA. Rather than give a generic reduction between detection and recovery for variants of sparse PCA, we modify our existing reductions to produce two copies. Note that the reductions SPCA-HIGH-SPARSITY and SPCA-LOW-SPARSITY approximately produce instances of ROS_D and BC_D, respectively, as intermediates. Consider the reduction that applies GAUSSIAN-CLONING to these intermediates and then the second steps of the reductions to both copies. Under H_1, this yields two independent copies of $N\left(0, I_n + \theta u u^\top\right)$ with a common latent spike u. Furthermore the resulting parameter θ is only affected up to a constant factor.

Now given two independent samples from $\text{SPCA}_D(n, k, n, \theta)$ constrained to have the same hidden vector under H_1 and a randomized algorithm ϕ that solves the weak recovery problem, we will show that a spectral algorithm solves detection. Let (X^1, X^2) denote the $n \times n$ data matrices for the two independent samples. Consider the algorithm that computes the $k \times k$ empirical covariance matrix $\hat{\Sigma}$ using the columns in X^2 restricted to the indices in $\phi(X^1)$ and then outputs H_1 if and only if $\lambda_1(\hat{\Sigma}) \geq 1 + 2\sqrt{k/n}$. Under H_0, the same argument in Theorem 9.39 implies that $\lambda_1(\hat{\Sigma}) < 1 + 2\sqrt{k/n}$ with probability at least $1 - 2e^{-k/2}$. Under H_1, let k_1 denote the size of the intersection between $\phi(X^1)$ and the latent support of the hidden vector. It follows that each column of X^2 restricted to the indices in $S = \phi(X^1)$ is distributed as $N\left(0, I_k + \theta u_S u_S^\top\right)$ where u is the hidden vector. Now note that $\|u_S\|_2 \geq \sqrt{k_1/k}$ and by the argument in Theorem 9.39, it follows that $u_S^\top \hat{\Sigma} u_S \geq 1 + \frac{\theta}{2k} \cdot \sqrt{k_1 k_2}$ with probability tending to 1 as $k \to \infty$. Therefore if $\theta \geq \tau(k)\sqrt{k/n}$ for some $\tau(k) \to \infty$, then this algorithm has Type I+II error tending to zero as $k \to \infty$, which aligns with the information-theoretic lower bound on SPCA_D and SPCA_R.

11 Future Directions

A general direction for future work is to add more problems to the web of reductions established here. This work also has left number of specific problems open, including the following.

1. Collisions between support elements in REFLECTION-CLONING causes our formulations of $\text{SSW}_D, \text{ROS}_D, \text{SSBM}_D$ and SPCA_D to be as composite hypothesis testing problems rather than the canonical simple hypothesis testing formulations. Is there an alternative reduction from planted clique that can strengthen these lower bounds to hold for the simple hypothesis testing analogues?

2. All previous planted clique lower bounds for SPCA_D are not tight over the parameter regime $k \ll \sqrt{n}$. Is there a reduction from planted clique yielding tight computational lower bounds for SPCA_D in this highly sparse regime?

3. Is there a polynomial time algorithm for the recovery variant of SSBM matching the computational barrier for the detection variant?

4. Can the PDS recovery conjecture be shown to follow from the planted clique conjecture?

Acknowledgements

References

[AAK+07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 496–505. ACM, 2007.

[AAM+11] Noga Alon, Sanjeev Arora, Rajsekar Manokaran, Dana Moshkovitz, and Omri Weinstein. Inapproximability of densest κ-subgraph from average case hardness. *Unpublished manuscript*, 1, 2011.

[AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[Abb17] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *arXiv preprint arXiv:1703.10146*, 2017.

[ABBG11] Sanjeev Arora, Boaz Barak, Markus Brunnermeier, and Rong Ge. Computational complexity and information asymmetry in financial products. *Communications of the ACM*, 54(5):101–107, 2011.

[ABC13] Per Austrin, Mark Braverman, and Eden Chlamtác. Inapproximability of np-complete variants of nash equilibrium. *Theory of Computing*, 9(3):117–142, 2013.

[ABDR+18] Albert Atserias, Ilario Bonacina, Susanna De Rezende, Massimo Lauria, Jakob Nordström, and Alexander Razborov. Clique is hard on average for regular resolution. 2018.

[ABH16] Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2016.

[ABW10] Benny Applebaum, Boaz Barak, and Avi Wigderson. Public-key cryptography from different assumptions. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 171–180. ACM, 2010.

[ACLR15] Pranjal Awasthi, Moses Charikar, Kevin A Lai, and Andrej Risteski. Label optimal regret bounds for online local learning. pages 150–166, 2015.

[ACV+14] Ery Arias-Castro, Nicolas Verzelen, et al. Community detection in dense random networks. *The Annals of Statistics*, 42(3):940–969, 2014.

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures and Algorithms*, 13(3-4):457–466, 1998.

[AS15] Emmanuel Abbe and Colin Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic bp, and the information-computation gap. *arXiv preprint arXiv:1512.09080*, 2015.

[AV11] Brendan PW Ames and Stephen A Vavasis. Nuclear norm minimization for the planted clique and biclique problems. *Mathematical programming*, 129(1):69–89, 2011.

[AW09] Arash A Amini and Martin J Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. *The Annals of Statistics*, 37(5B):2877–2921, 2009.

[Bar17] Boaz Barak. *The Complexity of Public-Key Cryptography*, pages 45–77. Springer International Publishing, Cham, 2017.

[BB18] Nicolai Baldin and Quentin Berthet. Optimal link prediction with matrix logistic regression. *arXiv preprint arXiv:1803.07054*, 2018.

[BBB+13] Maria-Florina Balcan, Christian Borgs, Mark Braverman, Jennifer Chayes, and Shang-Hua Teng. Finding endogenously formed communities. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 767–783. Society for Industrial and Applied Mathematics, 2013.

[BCC+10] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an $o(n^{1/4})$ approximation for densest k-subgraph. *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 201–210, 2010.

[BCKS16] Umang Bhaskar, Yu Cheng, Young Kun Ko, and Chaitanya Swamy. Hardness results for signaling in bayesian zero-sum and network routing games. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 479–496. ACM, 2016.

[BDLS17] Sivaraman Balakrishnan, Simon S Du, Jerry Li, and Aarti Singh. Computationally efficient robust sparse estimation in high dimensions. pages 169–212, 2017.

[BGL17] Vijay Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee. Sum-of-squares certificates for maxima of random tensors on the sphere. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 81. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[BGN11] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

[BHK+16] Boaz Barak, Samuel B Hopkins, Jonathan Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 428–437. IEEE, 2016.

[BI13] Cristina Butucea and Yuri I Ingster. Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli*, 19(5B):2652–2688, 2013.

[BJNP13] Aharon Birnbaum, Iain M Johnstone, Boaz Nadler, and Debashis Paul. Minimax bounds for sparse pca with noisy high-dimensional data. *Annals of statistics*, 41(3):1055, 2013.

[BKR+11] Sivaraman Balakrishnan, Mladen Kolar, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. Statistical and computational tradeoffs in biclustering. In *NIPS 2011 workshop on computational trade-offs in statistical learning*, volume 4, 2011.

[BMV+18] 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*, 2018.

[BPW18] Afonso S Bandeira, Amelia Perry, and Alexander S Wein. Notes on computational-to-statistical gaps: predictions using statistical physics. *arXiv preprint arXiv:1803.11132*, 2018.

[BR13a] Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *COLT*, pages 1046–1066, 2013.

[BR13b] Quentin Berthet and Philippe Rigollet. Optimal detection of sparse principal components in high dimension. *The Annals of Statistics*, 41(4):1780–1815, 2013.

[BT06a] Andrej Bogdanov and Luca Trevisan. On worst-case to average-case reductions for np problems. *SIAM Journal on Computing*, 36(4):1119–1159, 2006.

[BT+06b] Andrej Bogdanov, Luca Trevisan, et al. Average-case complexity. *Foundations and Trends® in Theoretical Computer Science*, 2(1):1–106, 2006.

[CC18] Utkan Onur Candogan and Venkat Chandrasekaran. Finding planted subgraphs with few eigenvalues using the schur–horn relaxation. *SIAM Journal on Optimization*, 28(1):735–759, 2018.

[CDK12] Eden Chlamtac, Michael Dinitz, and Robert Krauthgamer. Everywhere-sparse spanners via dense subgraphs. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 758–767. IEEE, 2012.

[CDM17] Eden Chlamtáč, Michael Dinitz, and Yury Makarychev. Minimizing the union: Tight approximations for small set bipartite vertex expansion. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 881–899. SIAM, 2017.

[CDMF+09] Mireille Capitaine, Catherine Donati-Martin, Delphine Féral, et al. The largest eigenvalues of finite rank deformation of large wigner matrices: convergence and nonuniversality of the fluctuations. *The Annals of Probability*, 37(1):1–47, 2009.

[Che15] Yudong Chen. Incoherence-optimal matrix completion. *IEEE Transactions on Information Theory*, 61(5):2909–2923, 2015.

[CJ13] Venkat Chandrasekaran and Michael I Jordan. Computational and statistical tradeoffs via convex relaxation. *Proceedings of the National Academy of Sciences*, 110(13):E1181–E1190, 2013.

[CLR15] Tony Cai, Tengyuan Liang, and Alexander Rakhlin. Computational and statistical boundaries for submatrix localization in a large noisy matrix. *arXiv preprint arXiv:1502.01988*, 2015.

[CM18] E. Chlamtáč and P. Manurangsi. Sherali-adams integrality gaps matching the log-density threshold. *arXiv preprint arXiv:1804.07842*, 2018.

[CMW+13] T Tony Cai, Zongming Ma, Yihong Wu, et al. Sparse pca: Optimal rates and adaptive estimation. *The Annals of Statistics*, 41(6):3074–3110, 2013.

[CMW15] Tony Cai, Zongming Ma, and Yihong Wu. Optimal estimation and rank detection for sparse spiked covariance matrices. *Probability theory and related fields*, 161(3-4):781–815, 2015.

[CNW18] Moses Charikar, Yonatan Naamad, and Jimmy Wu. On finding dense common subgraphs. *arXiv preprint arXiv:1802.06361*, 2018.

[CO03] Amin Coja-Oghlan. Finding large independent sets in polynomial expected time. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 511–522. Springer, 2003.

[COE15] Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Structures & Algorithms*, 47(3):436–486, 2015.

[CPR16] Siu On Chan, Dimitris Papailliopoulos, and Aviad Rubinstein. On the approximability of sparse pca. In *Conference on Learning Theory*, pages 623–646, 2016.

[CW18] T. Tony Cai and Yihong Wu. Statistical and computational limits for sparse matrix detection. *arXiv preprint arXiv:1801.00518*, 2018.

[CX16] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *Journal of Machine Learning Research*, 17(27):1–57, 2016.

[DF87] Persi Diaconis and David Freedman. A dozen de finetti-style results in search of a theory. 23(S2):397–423, 1987.

[DGGP14] Yael Dekel, Ori Gurel-Gurevich, and Yuval Peres. Finding hidden cliques in linear time with high probability. *Combinatorics, Probability and Computing*, 23(1):29–49, 2014.

[DKS16] Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. *arXiv preprint arXiv:1611.03473*, 2016.

[DLSS14] Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. From average case complexity to improper learning complexity. pages 441–448, 2014.

[DM14] Yash Deshpande and Andrea Montanari. Sparse pca via covariance thresholding. In *Advances in Neural Information Processing Systems*, pages 334–342, 2014.

[DM15a] Yash Deshpande and Andrea Montanari. Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Foundations of Computational Mathematics*, 15(4):1069–1128, 2015.

[DM15b] Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *COLT*, pages 523–562, 2015.

[DSS16] Amit Daniely and Shai Shalev-Shwartz. Complexity theoretic limitations on learning dnfs. pages 815–830, 2016.

[Dug14] Shaddin Dughmi. On the hardness of signaling. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 354–363. IEEE, 2014.

[EHV12] Kord Eickmeyer, Kristoffer Arnsfelt Hansen, and Elad Verbin. Approximating the minmax value of 3-player games within a constant is as hard as detecting planted cliques. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 19, page 25, 2012.

[Fei02] Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 534–543. ACM, 2002.

[FGR⁺12] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh Vempala, and Ying Xiao. Statistical algorithms and a lower bound for planted clique. *arXiv preprint arXiv:1201.1214*, 2012.

[FGR⁺13] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 655–664. ACM, 2013.

[FK00] Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures and Algorithms*, 16(2):195–208, 2000.

[FK03] Uriel Feige and Robert Krauthgamer. The probable value of the lovász–schrijver relaxations for maximum independent set. *SIAM Journal on Computing*, 32(2):345–370, 2003.

[FO05] Uriel Feige and Eran Ofek. Finding a maximum independent set in a sparse random graph. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 282–293. Springer, 2005.

[FP07] Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large wigner matrices. *Communications in mathematical physics*, 272(1):185–228, 2007.

[FPV15] Vitaly Feldman, Will Perkins, and Santosh Vempala. On the complexity of random satisfiability problems with planted solutions. pages 77–86, 2015.

[FR10] Uriel Feige and Dorit Ron. Finding hidden cliques in linear time. In *21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10)*, pages 189–204. Discrete Mathematics and Theoretical Computer Science, 2010.

[GMZ17] Chao Gao, Zongming Ma, and Harrison H Zhou. Sparse cca: Adaptive estimation and computational barriers. *The Annals of Statistics*, 45(5):2074–2101, 2017.

[GS14] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 369–376. ACM, 2014.

[HK11] Elad Hazan and Robert Krauthgamer. How hard is it to approximate the best nash equilibrium? *SIAM Journal on Computing*, 40(1):79–91, 2011.

[HKP$^+$16] Samuel B Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm. On the integrality gap of degree-4 sum of squares for planted clique. pages 1079–1095, 2016.

[HKP$^+$17] 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. *Proceedings of the fifty-eighth IEEE Foundations of Computer Science*, pages 720–731, 2017.

[HMRW14] Moritz Hardt, Raghu Meka, Prasad Raghavendra, and Benjamin Weitz. Computational limits for matrix completion. In *COLT*, pages 703–725, 2014.

[HS17a] Shuichi Hirahara and Rahul Santhanam. On the average-case complexity of mcsp and its variants. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 79. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[HS17b] Samuel B Hopkins and David Steurer. Efficient bayesian estimation from few samples: community detection and related problems. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 379–390. IEEE, 2017.

[HWX15] Bruce E Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. In *COLT*, pages 899–928, 2015.

[HWX16a] Bruce Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016.

[HWX16b] Bruce Hajek, Yihong Wu, and Jiaming Xu. Information limits for recovering a hidden community. pages 1894–1898, 2016.

[Jer92] Mark Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992.

[JL04] Iain M Johnstone and Arthur Yu Lu. Sparse principal components analysis. *Unpublished manuscript*, 2004.

[JP00] Ari Juels and Marcus Peinado. Hiding cliques for cryptographic security. *Designs, Codes and Cryptography*, 20(3):269–280, 2000.

[KBRS11] Mladen Kolar, Sivaraman Balakrishnan, Alessandro Rinaldo, and Aarti Singh. Minimax localization of structural information in large noisy matrices. In *Advances in Neural Information Processing Systems*, pages 909–917, 2011.

[KMOW17] Pravesh K Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any csp. *arXiv preprint arXiv:1701.04521*, 2017.

[KNV15] Robert Krauthgamer, Boaz Nadler, and Dan Vilenchik. Do semidefinite relaxations solve sparse pca up to the information limit? *The Annals of Statistics*, 43(3):1300–1322, 2015.

[KZ14] Pascal Koiran and Anastasios Zouzias. Hidden cliques and the certification of the restricted isometry property. *IEEE Transactions on Information Theory*, 60(8):4999–5006, 2014.

[Lev86] Leonid A Levin. Average case complete problems. *SIAM Journal on Computing*, 15(1):285–286, 1986.

[Li17] Jerry Li. Robust sparse estimation tasks in high dimensions. *arXiv preprint arXiv:1702.05860*, 2017.

[Ma13] Zongming Ma. Sparse principal component analysis and iterative thresholding. *The Annals of Statistics*, 41(2):772–801, 2013.

[Mas07] Pascal Massart. *Concentration inequalities and model selection*, volume 6. Springer, 2007.

[Mas14] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703. ACM, 2014.

[McS01] Frank McSherry. Spectral partitioning of random graphs. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 529–537. IEEE, 2001.

[MNS12] Elchanan Mossel, Joe Neeman, and Allan Sly. Stochastic block models and reconstruction. *arXiv preprint arXiv:1202.1499*, 2012.

[MNS13] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, pages 1–44, 2013.

[MNS14] Elchanan Mossel, Joe Neeman, and Allan Sly. Consistency thresholds for binary symmetric block models. *arXiv preprint arXiv:1407.1591*, 2014.

[Mon15] Andrea Montanari. Finding one community in a sparse graph. *Journal of Statistical Physics*, 161(2):273–299, 2015.

[MRZ15] 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.

[MV09] Lorenz Minder and Dan Vilenchik. Small clique detection and approximate nash equilibria. pages 673–685, 2009.

[MW15a] Tengyu Ma and Avi Wigderson. Sum-of-squares lower bounds for sparse pca. In *Advances in Neural Information Processing Systems*, pages 1612–1620, 2015.

[MW15b] Zongming Ma and Yihong Wu. Computational barriers in minimax submatrix detection. *The Annals of Statistics*, 43(3):1089–1116, 2015.

[Pau07]   Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[Péc06]   Sandrine Péché. The largest eigenvalue of small rank perturbations of hermitian random matrices. *Probability Theory and Related Fields*, 134(1):127–173, 2006.

[Pit97]   Jim Pitman. Some probabilistic aspects of set partitions. *The American mathematical monthly*, 104(3):201–209, 1997.

[PWB16]   Amelia Perry, Alexander S Wein, and Afonso S Bandeira. Statistical limits of spiked tensor models. *arXiv preprint arXiv:1612.07728*, 2016.

[PWBM16]   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.

[RS15]   Prasad Raghavendra and Tselil Schramm. Tight lower bounds for planted clique in the degree-4 sos program. *arXiv preprint arXiv:1507.05136*, 2015.

[RV+17]   Mustazee Rahman, Balint Virag, et al. Local algorithms for independent sets are half-optimal. *The Annals of Probability*, 45(3):1543–1577, 2017.

[SBW16]   Nihar B Shah, Sivaraman Balakrishnan, and Martin J Wainwright. Feeling the bern: Adaptive estimators for bernoulli probabilities of pairwise comparisons. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 1153–1157. IEEE, 2016.

[SSM13]   Dan Shen, Haipeng Shen, and James Stephen Marron. Consistency of sparse pca in high dimension, low sample size contexts. *Journal of Multivariate Analysis*, 115:317–333, 2013.

[SWP+09]   Andrey A Shabalin, Victor J Weigman, Charles M Perou, Andrew B Nobel, et al. Finding large average submatrices in high dimensional data. *The Annals of Applied Statistics*, 3(3):985–1012, 2009.

[VAC+15]   Nicolas Verzelen, Ery Arias-Castro, et al. Community detection in sparse random networks. *The Annals of Applied Probability*, 25(6):3465–3510, 2015.

[Ver10]   Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

[VL12]   Vincent Q Vu and Jing Lei. Minimax rates of estimation for sparse pca in high dimensions. In *AISTATS*, volume 15, pages 1278–1286, 2012.

[Vu05]   Van H Vu. Spectral norm of random matrices. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 423–430. ACM, 2005.

[WBP16]   Tengyao Wang, Quentin Berthet, and Yaniv Plan. Average-case hardness of rip certification. In *Advances in Neural Information Processing Systems*, pages 3819–3827, 2016.

[WBS16]   Tengyao Wang, Quentin Berthet, and Richard J Samworth. Statistical and computational trade-offs in estimation of sparse principal components. *The Annals of Statistics*, 44(5):1896–1930, 2016.

[Yu97] Bin Yu. Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997.

[ZWJ14] Yuchen Zhang, Martin J Wainwright, and Michael I Jordan. Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. In *COLT*, pages 921–948, 2014.

[ZX17] Anru Zhang and Dong Xia. Tensor svd: Statistical and computational limits. *arXiv preprint arXiv:1703.02724*, 2017.

A Deferred Proofs from Section 3

Proof of Lemma 3.4. Let B be any event in $\sigma\{Y\}$. It follows that

$$\mathbb{P}[Y \in B | Y \in A] - \mathbb{P}[Y \in B] = \mathbb{P}[Y \in B | Y \in A] \cdot (1 - \mathbb{P}[Y \in A]) - \mathbb{P}[Y \in B \cap A^c]$$

Since $\mathbb{P}[Y \in B \cap A^c] \le \mathbb{P}[Y \in A^c] = 1 - \mathbb{P}[Y \in A]$ and $\mathbb{P}[Y \in B | Y \in A] \in [0, 1]$, we have that the quantity above is between $-\mathbb{P}[Y \in A^c]$ and $\mathbb{P}[Y \in A^c]$. Therefore, by the definition of total variation distance

$$d_{\mathrm{TV}}\left(\mathcal{L}(Y|A), \mathcal{L}(Y)\right) = \sup_{B \in \sigma\{Y\}} |\mathbb{P}[Y \in B | Y \in A] - \mathbb{P}[Y \in B]| = \mathbb{P}[Y \in A^c]$$

where the equality case is achieved by setting $B = A$. This proves the lemma. $\qquad\square$

B Deferred Proofs from Section 5

Proof of Lemma 5.2. Let $f_X(m)$ and $g_X(m)$ be the PMFs of $\mathrm{Pois}(c\lambda)$ and $\mathrm{Pois}(\lambda)$, respectively. Note that

$$f_X(m) = \frac{e^{-c\lambda}(c\lambda)^m}{m!} \quad \text{and} \quad g_X(m) = \frac{e^{-\lambda}(\lambda)^m}{m!}$$

can be computed and sampled in $O(1)$ operations. Therefore Lemma 5.1 implies that $\mathrm{RKP1}$ can be computed in $O(N) = O(\log n)$ time. The rest of the proof entails bounding Δ in Lemma 5.1 with $p = 1$. Let the set S be as defined in Lemma 5.1, let $M = \log_c q^{-1} \ge 3\epsilon^{-1}$ and define the set $S' = \{m \in \mathbb{Z}_{\ge 0} : m \le M\}$. Now note that if $m \in S'$ then

$$\frac{f_X(m)}{g_X(m)} = \frac{\frac{e^{-c\lambda}(c\lambda)^m}{m!}}{\frac{e^{-\lambda}(\lambda)^m}{m!}} = e^{-(c-1)\lambda} c^m \le c^M \le q^{-1}$$

and therefore it follows that $S' \subseteq S$. For sufficiently large n, we have that $M = \log_c q^{-1} \ge cn^{-\epsilon} \ge c\lambda > \lambda$ and therefore a standard Poisson tail bound yields that

$$\mathbb{P}_{X \sim g_X}[X \notin S] \le \mathbb{P}_{X \sim g_X}[X > M] \le e^{-\lambda} \left(\frac{e\lambda}{M}\right)^M \le \left(\frac{e}{M}\right)^M \lambda^{3\epsilon^{-1}} \le \left(\frac{e}{M}\right)^M n^{-3}$$

since $\lambda \le n^{-\epsilon}$. Similarly, we have that $\mathbb{P}_{X \sim f_X}[X \notin S] \le \left(\frac{ce}{M}\right)^M n^{-3}$. Now note that for sufficiently large n, we have that

$$
\begin{aligned}
\frac{2 \cdot \mathbb{P}_{X \sim f_X}[X \notin S]}{1 - q} + (\mathbb{P}_{X \sim g_X}[X \notin S] + q)^N &\le \frac{2\left(\frac{ce}{M}\right)^M n^{-3}}{1 - q} + \left(\left(\frac{e}{M}\right)^M n^{-3} + q\right)^N \\
&\le \frac{2\left(\frac{ce}{M}\right)^M n^{-3}}{1 - q} + \left(q^{1/2}\right)^N \\
&\le \left(2(1 - q)^{-1}\left(\frac{ce}{M}\right)^M + 1\right) n^{-3}
\end{aligned}
$$

By similar reasoning, we have that for sufficiently large n

$$\frac{2 \cdot \mathbb{P}_{X \sim g_X}[X \notin S]}{1 - q} + (\mathbb{P}_{X \sim f_X}[X \notin S])^N \le \left(2(1 - q)^{-1}\left(\frac{e}{M}\right)^M + 1\right) n^{-3}$$

Therefore $\Delta \le \left(2(1 - q)^{-1}\left(\frac{ce}{M}\right)^M + 1\right) n^{-3}$ for sufficiently large n and applying Lemma 5.1 proves the lemma. $\qquad\square$

Proof of Lemma 5.3. As in Lemma 5.2, let $f_X(m)$ and $g_X(m)$ be the PMFs of $\text{Pois}(c\lambda)$ and $\text{Pois}(\lambda)$ and note that they can be computed and sampled in $O(1)$ operations. Lemma 5.1 implies that RK_{P2} can be computed in $O(N) = O(n^K \log n) = \text{poly}(n)$ time. Let the set S be as defined in Lemma 5.1, let $M = \log_c(1 + 2\rho) \geq (K + 3)\epsilon^{-1}$ and define the set $S' = \{m \in \mathbb{Z}_{\geq 0} : m \leq M\}$. Now note that if $n > 1$, then $\lambda \leq n^{-\epsilon} < 1$ and it follows that

$$e^{-(c-1)\lambda} \geq 1 - (c-1)\lambda > 2 - c \geq 2 - (1+2\rho)^{\epsilon/(K+3)} > 1 - 2\rho$$

since $\epsilon \in (0, 1)$. Therefore if $m \in S'$, then

$$1 - 2\rho < e^{-(c-1)\lambda} \leq \frac{f_X(m)}{g_X(m)} = e^{-(c-1)\lambda}c^m \leq c^M \leq 1 + 2\rho$$

and it follows that $S' \subseteq S$. By the same Poisson tail bounds as in Lemma 5.2, we have that for sufficiently large n

$$\mathbb{P}_{X \sim g_X}[X \notin S] \leq \left(\frac{e}{M}\right)^M n^{-K-3} \quad \text{and} \quad \mathbb{P}_{X \sim f_X}[X \notin S] \leq \left(\frac{ce}{M}\right)^M n^{-K-3}$$

Now note that for sufficiently large n, we have that $\rho^{-1} \leq n^K$ and $\left(\frac{e}{M}\right)^M n^{-K-3} \leq \frac{1}{2}n^{-K} \leq \frac{1}{2}\rho$ since $M = O_n(1)$. Therefore

$$2\rho^{-1} \cdot \mathbb{P}_{X \sim f_X}[X \notin S] + \left(\mathbb{P}_{X \sim g_X}[X \notin S] + \frac{1}{1+2\rho}\right)^N \leq 2\rho^{-1}\left(\frac{ce}{M}\right)^M n^{-K-3}$$

$$+ \left(\left(\frac{e}{M}\right)^M n^{-K-3} + 1 - \rho\right)^N$$

$$\leq 2\left(\frac{ce}{M}\right)^M n^{-3} + \left(1 - \frac{\rho}{2}\right)^N$$

$$\leq 2\left(\frac{ce}{M}\right)^M n^{-3} + e^{-\rho N/2}$$

$$\leq \left(2\left(\frac{ce}{M}\right)^M + 1\right)n^{-3}$$

By similar reasoning, we have that for sufficiently large n,

$$2\rho^{-1} \cdot \mathbb{P}_{X \sim g_X}[X \notin S] + (\mathbb{P}_{X \sim f_X}[X \notin S] + 1 - 2\rho)^N \leq \left(2\left(\frac{e}{M}\right)^M + 1\right)n^{-3}$$

Therefore $\Delta \leq \left(2\left(\frac{ce}{M}\right)^M + 1\right)n^{-3}$ for sufficiently large n and applying Lemma 5.1 proves the lemma. $\square$

Proof of Lemma 5.4. Let $f_X(x)$ and $g_X(x)$ be the PDFs of $N(\mu, 1)$ and $N(0, 1)$, respectively, given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-(x-\mu)^2/2} \quad \text{and} \quad g_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

which can be computed and sampled in $O(1)$ operations in the given computational model. To bound N, note that since $\log(1 + x) \geq x/2$ for $x \in (0, 1)$, we have that

$$\log\left(\frac{p}{q}\right) \geq \frac{p-q}{2q} \geq \frac{1}{2}(p - q) \geq \frac{1}{2}n^{-O_n(1)}$$

and similarly that $\log\left(\frac{1-q}{1-p}\right) \geq \frac{p-q}{2(1-p)} \geq \frac{1}{2}(p-q) \geq \frac{1}{2}n^{-O_n(1)}$. Therefore $N = \text{poly}(n)$ and Lemma 5.1 implies that RK_G can be computed in $\text{poly}(n)$ time. Let the set S be as defined in Lemma 5.1, let $M = \sqrt{6\log n + 2\log(p-q)^{-1}}$ and define the set $S' = \{x \in \mathbb{R} : |x| \leq M\}$. Since $M\mu \leq \delta/2$, we have for any $x \in S'$ that

$$\frac{1-p}{1-q} \leq \exp\left(-2M\mu\right) \leq \exp\left(-M\mu - \frac{\mu^2}{2}\right) \leq \frac{f_X(x)}{g_X(x)} = \exp\left(x\mu - \frac{\mu^2}{2}\right) \leq \exp\left(M\mu\right) \leq \frac{p}{q}$$

for sufficiently large n since $M \to \infty$ as $n \to \infty$ and $\mu \in (0,1)$. This implies that $S' \subseteq S$. Using the bound $1 - \Phi(t) \leq \frac{1}{\sqrt{2\pi}} \cdot t^{-1}e^{-t^2/2}$ for $t \geq 1$, we have that

$$\mathbb{P}_{X \sim g_X}[X \notin S] \leq \mathbb{P}_{X \sim g_X}[X \notin S'] = 2\left(1 - \Phi(M)\right) \leq \frac{2}{\sqrt{2\pi}} \cdot M^{-1}e^{-M^2/2}$$

Similarly since $M/(M-\mu) \leq 2$, we have for sufficiently large n that

$$\begin{aligned}
\mathbb{P}_{X \sim f_X}[X \notin S] \leq \mathbb{P}_{X \sim f_X}[X \notin S'] &= (1 - \Phi(M-\mu)) + (1 - \Phi(M+\mu)) \\
&\leq \frac{1}{\sqrt{2\pi}} \cdot (M-\mu)^{-1}e^{-(M-\mu)^2/2} + \frac{1}{\sqrt{2\pi}} \cdot (M+\mu)^{-1}e^{-(M+\mu)^2/2} \\
&\leq \frac{1}{\sqrt{2\pi}} \cdot M^{-1}e^{-(\mu^2+M^2)/2}\left[\frac{M}{M-\mu} \cdot e^{M\mu} + 1\right] \\
&\leq \frac{1}{\sqrt{2\pi}}\left(1 + 2\sqrt{\frac{p}{q}}\right) \cdot M^{-1}e^{-M^2/2}
\end{aligned}$$

Now note that $M^{-1}e^{-M^2/2} \leq e^{-M^2/2} \leq (p-q)n^{-3}$ for sufficiently large n. If n is large enough then $\frac{2}{\sqrt{2\pi}} \cdot n^{-3} \leq \frac{q^{1/2}}{p(p^{1/2}+q^{1/2})} = \Omega_n(1)$ since $q = \Omega_n(1)$. Rearranging yields that $\frac{2}{\sqrt{2\pi}}(p-q)n^{-3} \leq \sqrt{\frac{q}{p}} - \frac{q}{p}$. Together with the previous two equations and the fact that $N \geq 6\delta^{-1}\log n$, this implies

$$\begin{aligned}
\frac{2 \cdot \mathbb{P}_{X \sim f_X}[X \notin S]}{p-q} + \left(\mathbb{P}_{X \sim g_X}[X \notin S] + \frac{q}{p}\right)^N &\leq \frac{2}{\sqrt{2\pi}}\left(1 + 2\sqrt{\frac{p}{q}}\right)n^{-3} + \left(\frac{2}{\sqrt{2\pi}}(p-q)n^{-3} + \frac{q}{p}\right)^N \\
&\leq \frac{2}{\sqrt{2\pi}}\left(1 + 2\sqrt{\frac{p}{q}}\right)n^{-3} + \left(\frac{q}{p}\right)^{N/2} \\
&\leq \left(1 + \frac{2}{\sqrt{2\pi}} + \frac{4}{\sqrt{2\pi}}\sqrt{\frac{p}{q}}\right)n^{-3} = O_n(n^{-3})
\end{aligned}$$

Now note that if $\frac{1}{\sqrt{2\pi}}\left(1 + 2\sqrt{\frac{p}{q}}\right)n^{-3} > \frac{(1-p)^{1/2}}{(1-q)((1-q)^{1/2}+(1-p)^{1/2})}$ then it follows that

$$\frac{1-p}{1-q} \leq \left(\frac{(1-q)}{\sqrt{2\pi}}\left(1 + 2\sqrt{\frac{p}{q}}\right) \cdot n^3 - 1\right)^{-1} \leq Cn^{-3}$$

for some constant $C > 0$ if n is sufficiently large, since $1 - q = \Omega_n(1)$. Otherwise, the same manipulation as above implies that $\frac{1}{\sqrt{2\pi}}\left(1 + 2\sqrt{\frac{p}{q}}\right)n^{-3} \leq \sqrt{\frac{1-p}{1-q}} - \frac{1-p}{1-q}$. Therefore we have in either case that

$$\frac{1}{\sqrt{2\pi}}\left(1 + 2\sqrt{\frac{p}{q}}\right)n^{-3} + \frac{1-p}{1-q} \leq \max\left\{\sqrt{\frac{1-p}{1-q}}, \frac{1}{\sqrt{2\pi}}\left(1 + 2\sqrt{\frac{p}{q}}\right)n^{-3} + Cn^{-3}\right\}$$

For sufficiently large n, the second term in the maximum above is at most n^{-2}. Therefore

$$\frac{2 \cdot \mathbb{P}_{X \sim g_X}[X \notin S]}{p - q} + \left(\mathbb{P}_{X \sim f_X}[X \notin S] + \frac{1-p}{1-q} \right)^N \leq \left(\frac{4}{\sqrt{2\pi}} \right) n^{-3} + \left(\frac{1}{\sqrt{2\pi}} \left(1 + 2\sqrt{\frac{p}{q}} \right) n^{-3} + \frac{1-p}{1-q} \right)^N$$

$$\leq \left(\frac{4}{\sqrt{2\pi}} \right) n^{-3} + \max\left\{ \left(\frac{1-p}{1-q} \right)^{N/2}, n^{-2N} \right\}$$

$$= O_n(n^{-3})$$

Therefore $\Delta = O_n(n^{-3})$ for sufficiently large n since $q = \Omega_n(1)$. Now applying Lemma 5.1 proves the lemma. $\qquad\square$

C Deferred Proofs from Section 6

Proof of Theorem 6.2. If $\beta < \alpha$ then PDS in this regime is information-theoretically impossible. Thus we may assume that $\beta \geq \alpha$. Take $\epsilon > 0$ to be a small enough constant so that

$$\beta + \epsilon(1 - \beta) < \frac{1}{2} + \frac{\alpha}{4}$$

Now let $\gamma = \frac{2\beta - \alpha + \epsilon(1-\beta)}{2-\alpha}$. Rearranging the inequality above yields that $\gamma \in (0, 1/2)$. Now set

$$\ell_n = \left\lceil \frac{(\alpha - \epsilon) \log_2 n}{2 - \alpha} \right\rceil, \qquad k_n = \lceil n^\gamma \rceil, \qquad N_n = 2^{\ell_n} n \qquad K_n = 2^{\ell_n} k_n,$$

$$p_n = 1 - e^{-4^{-\ell_n} c n^{-\epsilon}}, \qquad q_n = 1 - e^{-4^{-\ell_n} n^{-\epsilon}}$$

Take p to be a small enough constant so that $p < \frac{1}{2} c^{-3\epsilon^{-1}}$. By Lemma 6.1, there is a randomized polynomial time algorithm mapping $\mathrm{PC}_D(n, k_n, p)$ to $\mathrm{PDS}_D(N_n, K_n, p_n, q_n)$ with total variation converging to zero as $n \to \infty$. This map with Lemma 3.1 now implies that property 2 above holds. We now verify property 1. Note that

$$\lim_{n \to \infty} \frac{\log K_n}{\log N_n} = \lim_{n \to \infty} \frac{\left\lceil \frac{(\alpha-\epsilon) \log_2 n}{2-\alpha} \right\rceil \cdot \log 2 + \left(\frac{2\beta-\alpha+\epsilon(1-\beta)}{2-\alpha} \right) \log n}{\left\lceil \frac{(\alpha-\epsilon) \log_2 n}{2-\alpha} \right\rceil \cdot \log 2 + \log n} = \frac{\frac{\alpha-\epsilon}{2-\alpha} + \frac{2\beta-\alpha+\epsilon(1-\beta)}{2-\alpha}}{\frac{\alpha-\epsilon}{2-\alpha} + 1} = \beta$$

Note that as $n \to \infty$, it follows that since $4^{-\ell_n} n^{-\epsilon} \to 0$,

$$q_n = 1 - e^{-4^{-\ell_n} n^{-\epsilon}} \sim 4^{-\ell_n} n^{-\epsilon}$$

Similarly $p_n \sim 4^{-\ell_n} c n^{-\epsilon}$ and thus $\frac{p_n}{q_n} \to c$. Note that

$$\lim_{n \to \infty} \frac{\log q_n^{-1}}{\log N_n} = \lim_{n \to \infty} \frac{2 \left\lceil \frac{(\alpha-\epsilon) \log_2 n}{2-\alpha} \right\rceil \log 2 + \epsilon \log n}{\left\lceil \frac{(\alpha-\epsilon) \log_2 n}{2-\alpha} \right\rceil \cdot \log 2 + \log n} = \frac{\frac{2(\alpha-\epsilon)}{2-\alpha} + \epsilon}{\frac{\alpha-\epsilon}{2-\alpha} + 1} = \alpha$$

which completes the proof. $\qquad\square$

Proof of Theorem 6.4. If $\beta < 2\alpha$ then PDS is information-theoretically impossible. Thus we may assume that $\beta \geq 2\alpha$. Let $\gamma = \frac{\beta-\alpha}{1-\alpha}$ and note that $\gamma \in (0, 1/2)$. Now set

$$\ell_n = \left\lceil \frac{\alpha \log_2 n}{1-\alpha} \right\rceil, \qquad k_n = \lceil n^\gamma \rceil, \qquad N_n = 2^{\ell_n} n \qquad K_n = 2^{\ell_n} k_n,$$

$$p_n = \Phi \left(2^{-\ell_n} \cdot \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}} \right)$$

By Lemma 6.1, there is a randomized polynomial time algorithm mapping $\mathrm{PC}_D(n, k_n, 1/2)$ to the detection problem $\mathrm{PDS}_D(N_n, K_n, p_n, 1/2)$ with total variation converging to zero as $n \to \infty$. This map with Lemma 3.1 now implies that property 2 above holds. We now verify property 1. Note that

$$\lim_{n\to\infty} \frac{\log K_n}{\log N_n} = \lim_{n\to\infty} \frac{\left\lceil \frac{\alpha \log_2 n}{1-\alpha} \right\rceil \cdot \log 2 + \left(\frac{\beta-\alpha}{1-\alpha} \right) \log n}{\left\lceil \frac{\alpha \log_2 n}{1-\alpha} \right\rceil \cdot \log 2 + \log n} = \frac{\frac{\alpha}{1-\alpha} + \frac{\beta-\alpha}{1-\alpha}}{\frac{\alpha}{1-\alpha} + 1} = \beta$$

Let $\mu = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$ and note that as $n \to \infty$, we have that

$$\lim_{n\to\infty} \frac{\Phi \left(2^{-\ell_n} \mu \right) - \frac{1}{2}}{2^{-\ell_n} \mu} = \lim_{\tau\to 0} \left(\frac{1}{\tau\sqrt{2\pi}} \int_0^\tau e^{-x^2/2} dx \right) = \frac{1}{\sqrt{2\pi}}$$

Therefore $p_n - q_n \sim \frac{2^{-\ell_n} \mu}{\sqrt{2\pi}}$ as $n \to \infty$. This implies

$$\lim_{n\to\infty} \frac{\log(p_n - q_n)^{-1}}{\log N_n} = \lim_{n\to\infty} \frac{2 \left\lceil \frac{\alpha \log_2 n}{1-\alpha} \right\rceil \cdot \log 2 - \log \mu}{\left\lceil \frac{\alpha \log_2 n}{1-\alpha} \right\rceil \cdot \log 2 + \log n} = \frac{\frac{2\alpha}{1-\alpha}}{\frac{\alpha}{1-\alpha} + 1} = \alpha$$

which completes the proof. $\qquad\square$

Proof of Lemma 6.5. Let $\phi = $ BC-REDUCTION be as in Figure 13. Let ϕ'_ℓ denote GAUSSIAN-LIFTING applied to G that outputs W after ℓ iterations without applying the thresholding in Step 4 of DISTRIBUTIONAL-LIFTING. Lemmas 5.5 and 6.3 imply GAUSSIAN-LIFTING ensures that

$$d_{\mathrm{TV}} \left(\phi'_\ell(G(n, 1/2)), M_{2^\ell n}(N(0, 1)) \right) = O \left(\frac{1}{\sqrt{\log n}} \right)$$

Now suppose that $W \sim M_{2^\ell n}(N(0, 1))$ and let $W' = \phi_{2\text{-}3}(W)$ denote the value of W after applying Steps 2 and 3 in Figure 13 to W. Note that the diagonal entries of W are i.i.d. $N(0, 1)$ since the diagonal entries of A are zero. If $i < j$, then it follows that

$$W'_{ij} = \frac{1}{\sqrt{2}} (W_{ij} + G_{ij}) \quad \text{and} \quad W'_{ji} = \frac{1}{\sqrt{2}} (W_{ij} - G_{ij})$$

Since W_{ij} and G_{ij} are independent and distributed as $N(0, 1)$, it follows that W'_{ij} and W'_{ji} are jointly Gaussian and uncorrelated, which implies that they are independent. Furthermore, W'_{ij} and W'_{ji} are both in the σ-algebra $\sigma\{W_{ij}, G_{ij}\}$ and collection of σ-algebras $\sigma\{W_{ij}, G_{ij}\}$ with $i < j$ is independent. Thus it follows that both W' and $(W')^{\mathrm{id}, \sigma}$ are distributed as $N(0, 1)^{\otimes 2^\ell n \times 2^\ell n}$. It follows by the data processing inequality that

$$d_{\mathrm{TV}} \left(\phi(G(n, 1/2)), N(0, 1)^{\otimes 2^\ell n \times 2^\ell n} \right) \leq d_{\mathrm{TV}} \left(\phi'_\ell(G(n, 1/2)), \mathcal{L}(W) \right) = O \left(\frac{1}{\sqrt{\log n}} \right)$$

101

Now consider $G \sim G(n, k, 1/2)$ and note that GAUSSIAN-LIFTING ensures that

$$d_{\mathrm{TV}}\left(\phi'_\ell(G), M_{2^\ell n}(2^\ell k, N(2^{-\ell}\mu, 1), N(0, 1))\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

Now let $W' \sim M_{2^\ell n}(S, N(2^{-\ell}\mu, 1), N(0, 1))$ where S is a subset of $[2^\ell n]$ of size $2^\ell k$ and let W be the matrix formed by applying Steps 1 and 2 above to W' in place of $\phi'_\ell(G)$. By the same jointly Gaussian independence argument above, it follows that the entries of W are independent and distributed as:

- $W_{ij} \sim N(2^{-\ell-1/2}\mu, 1)$ if $(i, j) \in S \times S$ and $i \neq j$; and

- $W_{ij} \sim N(0, 1)$ if $(i, j) \notin S \times S$ or $i = j$.

Now consider the matrix $(W')^{\mathrm{id},\sigma}$ conditioned on the permutation σ. Its entries are independent and identically distributed to the corresponding entries of $2^{-\ell-1/2}\mu \cdot \mathbf{1}_S \mathbf{1}_T^\top + N(0, 1)^{\otimes 2^\ell n \times 2^\ell n}$ where $T = \sigma(S)$ other than at the indices $(i, \sigma(i))$ for $i \in S$. Marginalizing to condition only on $\sigma(S) = T$ and coupling all entries with indices outside of $S \times T$ yields that

$$d_{\mathrm{TV}}\left(\mathcal{L}((W')^{\mathrm{id},\sigma}|\sigma(S) = T),\ \mathcal{L}\left(2^{-\ell-1/2}\mu \cdot \mathbf{1}_S \mathbf{1}_T^\top + N(0, 1)^{\otimes 2^\ell n \times 2^\ell n}\right)\right)$$

$$= d_{\mathrm{TV}}\left(\mathcal{L}\left((W')^{\mathrm{id},\sigma}[S \times T]|\sigma(S) = T\right), N(2^{-\ell-1/2}\mu, 1)^{\otimes 2^\ell k \times 2^\ell k}\right)$$

$$\leq \sqrt{\frac{1}{2} \cdot \chi^2\left(N(0, 1), N(2^{-\ell-1/2}\mu, 1)\right)}$$

$$\leq 2^{-\ell}\mu \leq \frac{\log 2}{2^{\ell+1}\sqrt{6\log n + 2\log 2}} = O\left(\frac{1}{2^\ell \sqrt{\log n}}\right)$$

by applying Lemma 4.1 and the χ^2 upper bound shown in Lemma 6.3. Note that Lemma 4.1 applies because $(W')^{\mathrm{id},\sigma}[S \times T]$ conditioned only on $\sigma(S) = T$ is distributed as a $2^\ell k \times 2^\ell k$ matrix with i.i.d. entries $N(2^{-\ell-1/2}\mu, 1)$, other than its diagonal entries which are i.i.d. $N(0, 1)$, and with its columns randomly permuted. Letting $\sigma(S) = T$ be chosen uniformly at random over all ordered pairs of size $2^\ell k$ subsets of $[2^\ell n]$ yields by the triangle inequality that

$$d_{\mathrm{TV}}\left(\mathcal{L}((W')^{\mathrm{id},\sigma}), \int \mathcal{L}\left(2^{-\ell-1/2}\mu \cdot \mathbf{1}_S \mathbf{1}_T^\top + N(0, 1)^{\otimes 2^\ell n \times 2^\ell n}\right) d\pi'(T)\right) = O\left(\frac{1}{2^\ell \sqrt{\log n}}\right)$$

where π' is the uniform distribution on all size $2^\ell k$ subsets of $[2^\ell n]$. Let $\pi(S, T)$ be the uniform distribution on all pairs of subsets of size $2^\ell k$ of $[2^\ell n]$. Taking S to be also chosen uniformly at random yields by the data processing and triangle inequalities that

$$d_{\mathrm{TV}}\left(\mathcal{L}(\phi(G)), \int \mathcal{L}\left(2^{-\ell-1/2}\mu \cdot \mathbf{1}_S \mathbf{1}_T^\top + N(0, 1)^{\otimes 2^\ell n \times 2^\ell n}\right) d\pi(S, T)\right)$$

$$\leq d_{\mathrm{TV}}\left(\phi'_\ell(G), M_{2^\ell n}(2^\ell k, N(2^{-\ell}\mu, 1), N(0, 1))\right)$$

$$+ \mathbb{E}_S\left[d_{\mathrm{TV}}\left(\mathcal{L}((W')^{\mathrm{id},\sigma}), \int \mathcal{L}\left(2^{-\ell-1/2}\mu \cdot \mathbf{1}_S \mathbf{1}_T^\top + N(0, 1)^{\otimes 2^\ell n \times 2^\ell n}\right) d\pi'(T)\right)\right]$$

$$= O\left(\frac{1}{\sqrt{\log n}}\right)$$

which completes the proof of the lemma. $\qquad\square$

Proof of Lemma 6.9. Let $\phi = \text{GENERAL-PDS-REDUCTION}$ be as in Figure 13. Let $\phi_1(G)$ denote the map in Step 1 applying GAUSSIAN-LIFTING for ℓ_1 iterations and let $\phi_2(G)$ denote the map in Step 2 applying the modified version of POISSON-LIFTING for ℓ_2 iterations. If $G \sim G(n, k, 1/2)$ then Lemma 6.3 implies that

$$d_{\text{TV}}\left(\phi_1(G), G\left(2^{\ell_1}n, 2^{\ell_1}k, \Phi\left(2^{-\ell_1}\mu\right), 1/2\right)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

For the values of λ, ρ and c in Step 2, we have that $c < 2^{1/4}$ and $\log_c(1 + 2\rho) = 4\epsilon^{-1} = O(1)$. Also observe that as $n \to \infty$,

$$\rho = \Phi\left(2^{-\ell_1}\mu\right) - 1/2 \sim \frac{1}{\sqrt{2\pi}} \cdot 2^{-\ell_1}\mu = \Omega\left(\frac{1}{2^{\ell_1}\sqrt{\log n}}\right) = \omega\left(\frac{1}{2^{\ell_1}n}\right)$$

Therefore the values $K = 1, \epsilon, \lambda, \rho, c$ and natural parameter $2^{\ell_1}n$ satisfy the preconditions to apply Lemma 5.4. It follows that

$$d_{\text{TV}}\left(\text{RKP2}(\text{Bern}(1/2 + \rho)), \text{Pois}(c\lambda)\right) \leq O\left(2^{-3\ell_1}n^{-3}\right), \quad \text{and}$$

$$d_{\text{TV}}\left(\text{RKP2}(\text{Bern}(1/2)), \text{Pois}(\lambda)\right) \leq O\left(2^{-3\ell_1}n^{-3}\right)$$

Now let H' be a sample from $G\left(2^{\ell_1}n, 2^{\ell_1}k, \Phi\left(2^{-\ell_1}\mu\right), 1/2\right)$. The argument in Lemma 6.1 applied with these total variation bounds yields that

$$d_{\text{TV}}\left(\phi_2(H'), G\left(2^{\ell}n, 2^{\ell}k, p_{\ell_1,\ell_2}, q_{\ell_1,\ell_2}\right)\right) = O\left(2^{-\ell_1\epsilon/2}n^{-\epsilon/2}\right)$$

Applying the triangle inequality and data processing inequality yields that

$$d_{\text{TV}}\left(\phi(G(n, k, 1/2)), G\left(2^{\ell}n, 2^{\ell}k, p_{\ell_1,\ell_2}, q_{\ell_1,\ell_2}\right)\right)$$

$$\leq d_{\text{TV}}\left(\phi_2 \circ \phi_1(G), \phi_2(H')\right) + d_{\text{TV}}\left(\phi_2(H'), G\left(2^{\ell}n, 2^{\ell}k, p_{\ell_1,\ell_2}, q_{\ell_1,\ell_2}\right)\right)$$

$$= O\left(\frac{1}{\sqrt{\log n}} + 2^{-\ell_1\epsilon/2}n^{-\epsilon/2}\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

By the same argument, if $G \sim G(n, 1/2)$ then

$$d_{\text{TV}}\left(\phi(G(n, 1/2)), G\left(2^{\ell}n, q_{\ell_1,\ell_2}\right)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

which completes the proof of the lemma. $\square$

Proof of Theorem 6.10. If $\beta < 2\gamma - \alpha$ then PDS in this regime is information-theoretically impossible. Thus we may assume that $1 > \beta \geq 2\gamma - \alpha$. Now let

$$\eta = 1 - (1 - \beta) \cdot \frac{2 - \epsilon}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)}$$

Note that the given condition on α, β and γ rearranges to $\frac{1-\beta}{2-\alpha-2(\gamma-\alpha)} > \frac{1}{4}$. Therefore taking $\epsilon > 0$ to be small enough ensures that $2 - \alpha - (2 - \epsilon)(\gamma - \alpha) > 0$, $\eta \in (0, 1/2)$ and $\alpha > \epsilon$. Now set

$$\ell_n^1 = \left\lceil \frac{(\gamma - \alpha)(2 - \epsilon)\log_2 n}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} \right\rceil, \qquad \ell_n^2 = \left\lceil \frac{(\alpha - \epsilon)\log_2 n}{2 - \alpha - (2 - \epsilon)(\gamma - \alpha)} \right\rceil,$$

$$k_n = \lceil n^\eta \rceil, \qquad N_n = 2^{\ell_n^1 + \ell_n^2} n, \qquad K_n = 2^{\ell_n^1 + \ell_n^2} k_n,$$

$$p_n = 1 - \exp\left(4^{-\ell_n^2}\left(2^{\ell_n^1}n\right)^{-\epsilon} \cdot \left(2\Phi\left(2^{-\ell_n^1}\mu\right)\right)^{\epsilon/4}\right), \qquad q_n = 1 - \exp\left(4^{-\ell_n^2}\left(2^{\ell_n^1}n\right)^{-\epsilon}\right)$$

where $\mu = \frac{\log 2}{2\sqrt{6\log n + 2\log 2}}$. By Lemma 6.9, there is a randomized polynomial time algorithm mapping $\mathrm{PC}_D(n, k_n, 1/2)$ to $\mathrm{PDS}_D(N_n, K_n, p_n, q_n)$ with total variation converging to zero as $n \to \infty$. This map with Lemma 3.1 now implies that property 2 above holds. We now verify property 1. Note that

$$\lim_{n\to\infty} \frac{\log K_n}{\log N_n} = \frac{\frac{(\gamma-\alpha)(2-\epsilon)}{2-\alpha-(2-\epsilon)(\gamma-\alpha)} + \frac{\alpha-\epsilon}{2-\alpha-(2-\epsilon)(\gamma-\alpha)} + 1 - \frac{(1-\beta)(2-\epsilon)}{2-\alpha-(2-\epsilon)(\gamma-\alpha)}}{\frac{(\gamma-\alpha)(2-\epsilon)}{2-\alpha-(2-\epsilon)(\gamma-\alpha)} + \frac{\alpha-\epsilon}{2-\alpha-(2-\epsilon)(\gamma-\alpha)} + 1}$$

$$= \frac{\frac{2-\epsilon}{2-\alpha-(2-\epsilon)(\gamma-\alpha)} - \frac{(1-\beta)(2-\epsilon)}{2-\alpha-(2-\epsilon)(\gamma-\alpha)}}{\frac{2-\epsilon}{2-\alpha-(2-\epsilon)(\gamma-\alpha)}} = \beta$$

Using the approximations in the proof of Theorem 6.4, we obtain as $n \to \infty$

$$q_n \sim 4^{-\ell_n^2}\left(2^{\ell_n^1}n\right)^{-\epsilon}$$

$$p_n - q_n \sim 4^{-\ell_n^2}\left(2^{\ell_n^1}n\right)^{-\epsilon}\left[\left(2\Phi\left(2^{-\ell_n^1}\mu\right)\right)^{\epsilon/4} - 1\right] \sim 4^{-\ell_n^2}\left(2^{\ell_n^1}n\right)^{-\epsilon} \cdot \frac{\epsilon}{2\sqrt{2\pi}} \cdot 2^{-\ell_n^1}\mu$$

Now it follows that

$$\lim_{n\to\infty} \frac{\log q_n^{-1}}{\log N_n} = \frac{2 \cdot \frac{\alpha-\epsilon}{2-\alpha-(2-\epsilon)(\gamma-\alpha)} + \epsilon \cdot \frac{(\gamma-\alpha)(2-\epsilon)}{2-\alpha-(2-\epsilon)(\gamma-\alpha)} + \epsilon}{\frac{2-\epsilon}{2-\alpha-(2-\epsilon)(\gamma-\alpha)}} = \alpha$$

$$\lim_{n\to\infty} \frac{\log(p_n - q_n)^{-1}}{\log N_n} = \frac{2 \cdot \frac{\alpha-\epsilon}{2-\alpha-(2-\epsilon)(\gamma-\alpha)} + (1+\epsilon) \cdot \frac{(\gamma-\alpha)(2-\epsilon)}{2-\alpha-(2-\epsilon)(\gamma-\alpha)} + \epsilon}{\frac{2-\epsilon}{2-\alpha-(2-\epsilon)(\gamma-\alpha)}} = \gamma$$

which completes the proof. $\qquad\square$

D   Deferred Proofs from Section 7

Proof of Lemma 7.2. Let $\phi = $ ROS-REDUCTION be as in Figure 15. Let $\phi_1 : \mathcal{G}_n \to \mathbb{R}^{n\times n}$ and $\phi_2 : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ denote the maps in Steps 1 and 2, respectively. By Lemma 6.5, it holds that

$$d_{\mathrm{TV}}\left(\phi_1(G(n, 1/2)), N(0, 1)^{\otimes n\times n}\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

$$d_{\mathrm{TV}}\left(\phi_1(G(n, k, 1/2)), \int \mathcal{L}\left(\frac{\mu}{\sqrt{2}} \cdot \mathbf{1}_S\mathbf{1}_T^\top + N(0, 1)^{\otimes n\times n}\right) d\pi'(S, T)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

where π' is the uniform distribution over pairs of k-subsets $S, T \subseteq [n]$. By Lemma 7.1, it holds that $\phi_2\left(N(0, 1)^{\otimes n\times n}\right) \sim N(0, 1)^{\otimes n\times n}$. By the data processing inequality, we have that

$$d_{\mathrm{TV}}\left(\mathcal{L}_{H_0}(\phi(G)), N(0, 1)^{\otimes n\times n}\right) = d_{\mathrm{TV}}\left(\phi_2 \circ \phi_1(G(n, 1/2)), \phi_2\left(N(0, 1)^{\otimes n\times n}\right)\right)$$

$$\leq d_{\mathrm{TV}}\left(\phi_1(G(n, 1/2)), N(0, 1)^{\otimes n\times n}\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

Let M be a the matrix distributed as $\frac{\mu}{\sqrt{2}} \cdot \mathbf{1}_S \mathbf{1}_T^\top + N(0,1)^{\otimes n \times n}$ where S and T are k-element subsets of $[n]$ chosen uniformly at random. It follows that the distribution of $\phi_2(M)$ conditioned on the sets S and T is given by

$$\mathcal{L}\left(\phi_2\left(M\right)|S,T\right) \sim \int \mathcal{L}\left(\frac{\mu}{2^\ell \sqrt{2}} \cdot rc^\top + N(0,1)^{\otimes n \times n}\right) d\pi(r,c)$$

where π is the prior in Lemma 7.1. As shown in Lemma 7.1, it follows that each pair (r,c) in the support of π satisfies that $\|r\|_2^2 = 2^\ell \|\mathbf{1}_S\|_2^2 = 2^\ell k$ and $\|c\|_2^2 = 2^\ell \|\mathbf{1}_T\|_2^2 = 2^\ell k$ and that $\|r\|_0, \|c\|_0 \le 2^\ell k$. Now consider the prior $\pi_{S,T}$ which is π conditioned on the event that the following inequalities hold

$$\|r\|_0, \|c\|_0 \ge 2^\ell k \left(1 - \max\left(\frac{2C\ell \cdot \log(2^\ell k)}{k}, \frac{2^\ell k}{n}\right)\right)$$

As shown in Lemma 7.1, this event occurs with probability at least $1 - 8/k$. Since $2^\ell k < \frac{n}{\log k}$, it follows that if $u = \frac{1}{\sqrt{2^\ell k}} \cdot r$ and $v = \frac{1}{\sqrt{2^\ell k}} \cdot c$ then $\pi_{S,T}$ induces a prior over pairs (u,v) in $\mathcal{V}_{n,2^\ell k}$. This follows from the fact that $r, c \in \mathbb{Z}^n$ implies that u and v have nonzero entries with magnitudes at least $1/\sqrt{2^\ell k}$. Applying Lemma 3.4 yields that since $2^{-\ell}\mu \cdot rc^\top = \mu k \cdot uv^\top$,

$$d_{\mathrm{TV}}\left(\mathcal{L}\left(\phi_2\left(M\right)|S,T\right), \int \mathcal{L}\left(\frac{\mu k}{\sqrt{2}} \cdot uv^\top + N(0,1)^{\otimes n \times n}\right) d\pi_{S,T}(u,v)\right) \le \frac{8}{k}$$

Let $\pi(u,v) = \mathbb{E}_{S,T}[\pi_{S,T}(u,v)]$ be the prior formed by marginalizing over S and T. Note that π is also supported on pairs of unit vectors in $\mathcal{V}_{n,2^\ell k}$. By the triangle inequality, it follows that

$$d_{\mathrm{TV}}\left(\mathcal{L}\left(\phi_2\left(M\right)\right), \int \mathcal{L}\left(\frac{\mu k}{\sqrt{2}} \cdot uv^\top + N(0,1)^{\otimes n \times n}\right) d\pi(u,v)\right)$$
$$\le \mathbb{E}_{S,T}\left[d_{\mathrm{TV}}\left(\mathcal{L}\left(\phi_2\left(M\right)|S,T\right), \int \mathcal{L}\left(\frac{\mu k}{\sqrt{2}} \cdot uv^\top + N(0,1)^{\otimes n \times n}\right) d\pi_{S,T}(u,v)\right)\right] \le \frac{8}{k}$$

By the triangle inequality and data processing inequality, we now have that

$$d_{\mathrm{TV}}\left(\mathcal{L}_{H_1}(\phi(G)), \int \mathcal{L}\left(\frac{\mu k}{\sqrt{2}} \cdot uv^\top + N(0,1)^{\otimes n \times n}\right) d\pi(u,v)\right)$$
$$\le d_{\mathrm{TV}}\left(\mathcal{L}_{H_1}(\phi_2 \circ \phi_1(G)), \mathcal{L}(\phi_2(M))\right) + d_{\mathrm{TV}}\left(\mathcal{L}\left(\phi_2\left(M\right)\right), \int \mathcal{L}\left(\frac{\mu k}{\sqrt{2}} \cdot uv^\top + N(0,1)^{\otimes n \times n}\right) d\pi(u,v)\right)$$
$$= O\left(\frac{1}{\sqrt{\log n}}\right) + \frac{8}{k} = O\left(\frac{1}{\sqrt{\log n}} + k^{-1}\right)$$

since M is a sample from the mixture $\int \mathcal{L}\left(\frac{\mu}{\sqrt{2}} \cdot \mathbf{1}_S \mathbf{1}_T^\top + N(0,1)^{\otimes n \times n}\right) d\pi'(S,T)$. This completes the proof of the lemma. $\square$

Proof of Theorem 7.3. If $\beta < 2\alpha$ then ROS_D is information-theoretically impossible. Thus we may assume that $\beta \ge 2\alpha$. Let $\gamma = \beta - \alpha$ and note that $\gamma \in (0, 1/2)$. Now set

$$\ell_n = \lceil \alpha \log_2 n \rceil, \qquad k_n = \lceil n^\gamma \rceil, \qquad N_n = 2n, \qquad K_n = 2^{\ell_n} k_n, \qquad \mu_n = \frac{\mu k_n}{\sqrt{2}}$$

where $\mu = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$. By Lemma 7.2, there is a randomized polynomial time algorithm mapping $\mathrm{PC}_D(2n, k_n, 1/2)$ to the detection problem $\mathrm{ROS}_D(N_n, K_n, \mu_n)$ under H_0 and to a prior over

H_1 with total variation converging to zero as $n \to \infty$. This map with Lemma 3.1 now implies that property 2 above holds. We now verify property 1. Note that

$$\lim_{n\to\infty} \frac{\log(K_n \mu_n^{-1})}{\log N_n} = \lim_{n\to\infty} \frac{\lceil \alpha \log_2 n \rceil \cdot \log 2 - \log(\mu/\sqrt{2})}{\log n + \log 2} = \alpha$$

$$\lim_{n\to\infty} \frac{\log K_n}{\log N_n} = \lim_{n\to\infty} \frac{\lceil \alpha \log_2 n \rceil \cdot \log 2 + \log k_n}{\log n + \log 2} = \alpha + (\beta - \alpha) = \beta$$

which completes the proof. $\qquad\square$

Proof of Theorem 7.5. When $2\alpha > \beta$, it holds that SROS_D is information-theoretically impossible. Therefore we may assume that $2\alpha \le \beta$ and in particular that $\alpha < \frac{1}{2}$. Now suppose that $\beta < \frac{1}{2} + \alpha$. First consider the case when $\beta \ge \frac{1}{2}$ and let $\epsilon = \frac{1}{2}\left(\alpha + \frac{1}{2} - \beta\right) \in \left(0, \frac{1}{2}\right)$. Now set

$$\ell_n = \left\lceil \left(\beta - \frac{1}{2} + \epsilon\right) \log_2 n \right\rceil, \qquad k_n = \left\lceil n^{\frac{1}{2}-\epsilon} \right\rceil, \qquad N_n = 2n, \qquad K_n = 2^{\ell_n} k_n, \qquad \mu_n = \frac{\mu k_n(k_n - 1)}{2\sqrt{n-1}}$$

where $\mu = \frac{\log 2}{2\sqrt{6\log n + 2\log 2}}$. By Lemma 7.4, there is a randomized polynomial time algorithm mapping $\mathrm{PC}_D(2n, k_n, 1/2)$ to the detection problem $\mathrm{SROS}_D(N_n, K_n, \mu_n)$ under H_0 and to a prior over H_1 with total variation converging to zero as $n \to \infty$. This map with Lemma 3.1 now implies that property 2 above holds. We now verify property 1. Note that

$$\lim_{n\to\infty} \frac{\log(K_n \mu_n^{-1})}{\log N_n} = \lim_{n\to\infty} \frac{\left[\left(\beta - \frac{1}{2} + \epsilon\right)\log_2 n\right] \cdot \log 2 - \log(k_n - 1) - \log(\mu/2) + \frac{1}{2}\log(n-1)}{\log n + \log 2}$$

$$= \left(\beta - \frac{1}{2} + \epsilon\right) - \left(\frac{1}{2} - \epsilon\right) + \frac{1}{2} = \alpha$$

$$\lim_{n\to\infty} \frac{\log K_n}{\log N_n} = \lim_{n\to\infty} \frac{\left[\left(\beta - \frac{1}{2} + \epsilon\right)\log_2 n\right] \cdot \log 2 + \log k_n}{\log n + \log 2} = \left(\beta - \frac{1}{2} + \epsilon\right) + \frac{1}{2} - \epsilon = \beta$$

Now consider the case when $\beta < \frac{1}{2}$ and $\alpha > 0$. In this case, set $\ell_n = 0$, $k_n = \left\lceil n^{\frac{1}{2}-\epsilon} \right\rceil$, $K_n = k_n$ and

$$N_n = 2\left\lceil n^{\frac{1}{\beta}\left(\frac{1}{2}-\epsilon\right)} \right\rceil \quad \text{and} \quad \mu_n = \frac{\mu' k_n(k_n - 1)}{2\sqrt{n-1}}$$

where $\epsilon = \min\left(\frac{\alpha}{2(\alpha+\beta)}, \frac{1}{2} - \beta\right)$ and

$$\mu' = \frac{\log 2}{2\sqrt{6\log n + 2\log 2}} \cdot n^{\epsilon - \frac{\alpha}{\beta}\left(\frac{1}{2}-\epsilon\right)}$$

Note that $\epsilon \le \frac{\alpha}{2(\alpha+\beta)}$ implies that $\epsilon - \frac{\alpha}{\beta}\left(\frac{1}{2} - \epsilon\right) \le 0$. By Lemma 7.4, there is a randomized polynomial time algorithm mapping $\mathrm{PC}_D(2n, k_n, 1/2)$ to the detection problem $\mathrm{SROS}_D(2n, K_n, \mu_n)$ under H_0 and to a prior over H_1 with total variation converging to zero as $n \to \infty$. Now consider the map that pads the resulting instance with i.i.d. $N(0,1)$ random variables until it is $N_n \times N_n$. Note that since $\epsilon \le \frac{1}{2} - \beta$, we have that $N_n \ge 2n$. By the data processing and triangle inequalities, it follows that this map takes $\mathrm{PC}_D(2n, k_n, 1/2)$ to $\mathrm{SSW}_D(N_n, K_n, \mu_n)$ with total variation converging

106

to zero as $n \to \infty$. This implies that property 2 above holds and we now verify property 1. Note

$$\lim_{n\to\infty} \frac{\log(K_n\mu_n^{-1})}{\log N_n} = \lim_{n\to\infty} \frac{\log 2 + \frac{1}{2}\log(n-1) - \log(k_n-1) - \log\mu'}{\log 2 + \frac{1}{\beta}\left(\frac{1}{2} - \epsilon\right)\log n}$$

$$= \frac{\frac{1}{2} - \left(\frac{1}{2} - \epsilon\right) - \left(\epsilon - \frac{\alpha}{\beta}\left(\frac{1}{2} - \epsilon\right)\right)}{\frac{1}{\beta}\left(\frac{1}{2} - \epsilon\right)} = \alpha$$

$$\lim_{n\to\infty} \frac{\log K_n}{\log N_n} = \lim_{n\to\infty} \frac{\left(\frac{1}{2} - \epsilon\right)\log n}{\log 2 + \frac{1}{\beta}\left(\frac{1}{2} - \epsilon\right)\log n} = \beta$$

which completes the proof of the theorem. $\qquad\qquad\square$

Proof of Theorem 7.7. When $2\alpha > \beta$, it holds that SSBM_D is information-theoretically impossible. Therefore we may assume that $2\alpha \le \beta$ and in particular that $\alpha < \frac{1}{2}$. Now suppose that $\beta < \frac{1}{2} + \alpha$. We consider the cases $\beta \ge \frac{1}{2}$ and $\beta < \frac{1}{2}$ and $q \le \frac{1}{2}$ and $q > \frac{1}{2}$, separately. First consider the case when $\beta \ge \frac{1}{2}$ and $q \le \frac{1}{2}$. Let $\epsilon = \frac{1}{2}\left(\alpha + \frac{1}{2} - \beta\right) \in \left(0, \frac{1}{2}\right)$ and set parameters similarly to Theorem 7.5 with

$$\ell_n = \left\lceil \left(\beta - \frac{1}{2} + \epsilon\right)\log_2 n \right\rceil, \qquad k_n = \left\lceil n^{\frac{1}{2}-\epsilon} \right\rceil, \qquad N_n = 2n, \qquad K_n = 2^{\ell_n}k_n, \qquad q_n = q$$

$$\rho_n = 2q \cdot \Phi\left(\frac{\mu(k_n-1)}{2^{\ell_n+1}\sqrt{n-1}}\right) - q$$

where $\mu = \frac{\log 2}{2\sqrt{6\log n} + 2\log 2}$. By Lemma 7.6, there is a randomized polynomial time algorithm mapping $\text{PC}_D(2n, k_n, 1/2)$ to the detection problem $\text{SSBM}_D(N_n, K_n, 1/2, (2q)^{-1}\rho_n)$ under H_0 and to a prior over H_1 with total variation converging to zero as $n \to \infty$. Now consider the algorithm that post-processes the resulting graph by keeping each edge independently with probability $2q$. This maps any instance of $\text{SSBM}_D(N_n, K_n, 1/2, (2q)^{-1}\rho_n)$ exactly to an instance of $\text{SSBM}_D(N_n, K_n, q, \rho_n)$. Therefore the data processing inequality implies that these two steps together yield a reduction that combined with Lemma 3.1 implies property 2 holds. Now observe that as $n \to \infty$,

$$\rho_n = 2q \cdot \Phi\left(\frac{\mu(k_n-1)}{2^{\ell_n+1}\sqrt{n-1}}\right) - q \sim 2q \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{\mu(k_n-1)}{2^{\ell_n+1}\sqrt{n-1}}$$

The same limit computations as in Theorem 7.5 show that property 1 above holds. If $q > 1/2$, then instead set

$$\rho_n = 2(1-q) \cdot \Phi\left(\frac{\mu(k_n-1)}{2^{\ell_n+1}\sqrt{n-1}}\right) - (1-q)$$

and post-process the graph resulting from the reduction in Lemma 7.6 by adding each absent edge with probability $2q - 1$. By a similar argument, the resulting reduction shows properties 1 and 2.

Now consider the case when $\beta < \frac{1}{2}$, $\alpha > 0$ and $q \le \frac{1}{2}$. Set $\ell_n = 0$, $k_n = \left\lceil n^{\frac{1}{2}-\epsilon} \right\rceil$, $K_n = k_n$ and

$$N_n = 2\left\lceil n^{\frac{1}{\beta}\left(\frac{1}{2}-\epsilon\right)} \right\rceil \quad \text{and} \quad \rho_n = 2q \cdot \Phi\left(\frac{\mu'(k-1)}{2\sqrt{n-1}}\right) - q$$

where $\epsilon = \min\left(\frac{\alpha}{2(\alpha+\beta)}, \frac{1}{2} - \beta\right)$ and $\mu' = \frac{\log 2}{2\sqrt{6\log n} + 2\log 2} \cdot n^{\epsilon - \frac{\alpha}{\beta}\left(\frac{1}{2}-\epsilon\right)}$ as in the proof of Theorem 7.5. Note that since $\epsilon \le \frac{1}{2} - \beta$, it follows that $N_n \ge 2n$. By Lemma 7.6, there is a randomized polynomial

time algorithm mapping $\mathrm{PC}_D(2n, k_n, 1/2)$ to the detection problem $\mathrm{SSBM}_D(2n, K_n, 1/2, (2q)^{-1}\rho_n)$ under H_0 and to a prior over H_1 with total variation converging to zero as $n \to \infty$. Now consider the map that post-processes the graph by:

1. keeping each edge independently with probability $2q$;

2. adding $N_n - 2n$ vertices to the resulting graph and includes each edge incident to these new vertices independently with probability q; and

3. randomly permuting the vertices of the resulting N_n-vertex graph.

Note that this maps any instance of $\mathrm{SSBM}_D(2n, K_n, 1/2, (2q)^{-1}\rho_n)$ exactly to an instance of $\mathrm{SSBM}_D(N_n, K_n, q, \rho_n)$. Thus the data processing inequality implies that this post-processing together with the reduction of Lemma 7.6 yields a reduction showing property 2. The same limit computations as in Theorem 7.5 show that property 1 above holds. The same adaptation as in the case $\beta \geq \frac{1}{2}$ also handles $q < \frac{1}{2}$. This completes the proof of the theorem. $\qquad\square$

E   Deferred Proofs from Section 8

Proof of Lemma 8.3. Let M' be the matrix output in Step 2. Under H_0, Lemma 7.2 implies that $M \sim N(0,1)^{\otimes n \times n}$ and Lemma 8.2 implies that we also have $M' \sim N(0,1)^{\otimes n \times n}$. It suffices to consider the case of H_1. By Lemma 7.2,

$$d_{\mathrm{TV}}\left(\mathcal{L}_{H_1}(M), \int \mathcal{L}\left(\frac{\mu k}{\sqrt{2}} \cdot uv^\top + N(0,1)^{\otimes n \times n}\right) d\pi'(u,v)\right) = O\left(\frac{1}{\sqrt{\log n}} + k^{-1}\right)$$

where $\pi'(u,v)$ is a prior supported on pairs of unit vectors in $\mathcal{V}_{n,2^\ell k}$. Let $W \sim \frac{\mu k}{\sqrt{2}} \cdot uv^\top + N(0,1)^{\otimes n \times n}$ where (u,v) is distributed according to π' and let φ denote the map in Step 2 taking A to RANDOM-ROTATION(A, τ). Conditioning on (u,v) yields by Lemma 8.2 that

$$d_{\mathrm{TV}}\left(\mathcal{L}\left(\varphi(W)|u,v\right), N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right)^{\otimes n}\right) \leq \frac{2(n+3)}{\tau n - n - 3}$$

Now consider the measure $\pi(u) = \mathbb{E}_v \pi'(u,v)$ given by marginalizing over v in π'. The triangle inequality implies that

$$d_{\mathrm{TV}}\left(\mathcal{L}(\varphi(W)), \int N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right)^{\otimes n} d\pi(u)\right)$$
$$\leq \mathbb{E}_{u,v}\left[d_{\mathrm{TV}}\left(\mathcal{L}\left(\varphi(W)|u,v\right), N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right)^{\otimes n}\right)\right] \leq \frac{2(n+3)}{\tau n - n - 3}$$

By the data processing inequality and triangle inequality, we now have that

$$d_{\mathrm{TV}}\left(\mathcal{L}_{H_1}(\phi(G)), \int N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right)^{\otimes n} d\pi(u)\right)$$
$$\leq d_{\mathrm{TV}}\left(\mathcal{L}_{H_1}(M), \mathcal{L}(W)\right) + d_{\mathrm{TV}}\left(\mathcal{L}(\varphi(W)), \int N\left(0, I_n + \frac{\mu^2 k^2}{2\tau n} \cdot uu^\top\right)^{\otimes n} d\pi(u)\right)$$
$$\leq \frac{2(n+3)}{\tau n - n - 3} + O\left(\frac{1}{\sqrt{\log n}} + k^{-1}\right)$$

since $\mathcal{L}_{H_1}(\phi(G)) \sim \mathcal{L}_{H_1}(\varphi(M))$. This completes the proof of the lemma. $\qquad\square$

Proof of Theorem 8.5. Note that if $\alpha \geq 1$, then sparse PCA is information theoretically impossible. Thus we may assume that $\alpha \in (0,1)$ and $\beta > \frac{1-\alpha}{2}$. Let $\gamma = \frac{1-\alpha}{2} \in (0,1/2)$. Now set $N_n = d_n = n$

$$\ell_n = \left\lceil \left(\beta - \frac{1-\alpha}{2} \right) \log_2 n \right\rceil, \qquad k_n = \lceil n^\gamma \rceil, \qquad K_n = 2^{\ell_n} k_n, \qquad \theta_n = \frac{\mu^2 k_n^2}{2\tau n}$$

where $\mu = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$ and τ is a sub-polynomially growing function of n. By Lemma 8.3, there is a randomized polynomial time algorithm mapping $\mathrm{PC}_D(n, k_n, 1/2)$ to the detection problem $\mathrm{SPCA}_D(N_n, K_n, d_n, \theta_n)$, exactly under H_0 and to a prior over H_1, with total variation converging to zero as $n \to \infty$. This map with Lemma 3.1 now implies that property 2 above holds. We now verify property 1. Note that

$$\lim_{n\to\infty} \frac{\log K_n}{\log N_n} = \lim_{n\to\infty} \frac{\left\lceil \left(\beta - \frac{1-\alpha}{2} \right) \log_2 n \right\rceil \cdot \log 2 + \left(\frac{1-\alpha}{2} \right) \log n}{\log n} = \beta$$

$$\lim_{n\to\infty} \frac{\log \theta_n^{-1}}{\log N_n} = \lim_{n\to\infty} \frac{(1 - 2\gamma) \log n - 2 \log \mu + \log(2\tau)}{\log n} = \alpha$$

which completes the proof. $\square$

Proof of Theorem 8.6. Note that if $\alpha \geq 1$, then USPCA_D and UBSPCA_D are information theoretically impossible. Thus we may assume that $\alpha \in (0,1)$. Now observe that $\beta \in (0,1)$ and $\frac{1-\alpha}{2} < \beta < \frac{1+\alpha}{2}$ imply that $\gamma = \frac{1-\alpha}{3-\alpha-2\beta} \in (0,1/2)$, that $\alpha + 2\beta - 1 > 0$ and $3 - \alpha - 2\beta > 0$. Now set

$$N_n = d_n = 2^{\ell_n} n, \qquad \ell_n = \left\lceil \left(\frac{\alpha + 2\beta - 1}{3 - \alpha - 2\beta} \right) \log_2 n \right\rceil, \qquad k_n = \lceil n^\gamma \rceil,$$

$$K_n = 2^{\ell_n} k_n, \qquad \theta_n = \frac{\mu^2 k_n^2}{2^{\ell_n+1} \tau n}$$

where $\mu = \frac{\log 2}{2\sqrt{6 \log n + 2 \log 2}}$ and τ is a sub-polynomially growing function of n. By Lemma 8.4, there is a randomized polynomial time algorithm mapping $\mathrm{PC}_D(n, k_n, 1/2)$ to the detection problem $\mathrm{UBSPCA}_D(N_n, K_n, d_n, \theta_n)$, exactly under H_0 and to a prior over H_1, with total variation converging to zero as $n \to \infty$. This map with Lemma 3.1 now implies that property 2 above holds. We now verify property 1. Note that

$$\lim_{n\to\infty} \frac{\log K_n}{\log N_n} = \lim_{n\to\infty} \frac{\left\lceil \left(\frac{\alpha+2\beta-1}{3-\alpha-2\beta} \right) \log_2 n \right\rceil \cdot \log 2 + \left(\frac{1-\alpha}{3-\alpha-2\beta} \right) \log n}{\left\lceil \left(\frac{\alpha+2\beta-1}{3-\alpha-2\beta} \right) \log_2 n \right\rceil \cdot \log 2 + \log n} = \frac{\frac{\alpha+2\beta-1}{3-\alpha-2\beta} + \frac{1-\alpha}{3-\alpha-2\beta}}{\frac{\alpha+2\beta-1}{3-\alpha-2\beta} + 1} = \beta$$

$$\lim_{n\to\infty} \frac{\log \theta_n^{-1}}{\log N_n} = \lim_{n\to\infty} \frac{\left\lceil \left(\frac{\alpha+2\beta-1}{3-\alpha-2\beta} \right) \log_2 n \right\rceil \cdot \log 2 + (1 - 2\gamma) \log n - 2 \log \mu + \log(2\tau)}{\left\lceil \left(\frac{\alpha+2\beta-1}{3-\alpha-2\beta} \right) \log_2 n \right\rceil \cdot \log 2 + \log n}$$

$$= \frac{\frac{\alpha+2\beta-1}{3-\alpha-2\beta} - 2 \cdot \frac{1-\alpha}{3-\alpha-2\beta} + 1}{\frac{\alpha+2\beta-1}{3-\alpha-2\beta} + 1} = \frac{\alpha + 2\beta - 1 - 2(1-\alpha) + 3 - \alpha - 2\beta}{2} = \alpha$$

which completes the proof. As described previously, the corresponding lower bound for USPCA_D follows by randomly signing the rows of the data matrix of the resulting UBSPCA_D instance. $\square$

F Deferred Proofs from Section 9

Proof of Theorem 9.9. Let $G \sim G_I(n, k, q)$ where $S \subseteq [n]$ denotes the indices of the planted independent set in G and satisfies $|S| = k$. Fix a vertex $i \in [n]$ and consider a random $J \in [n]$ where J is chosen uniformly at random from S if $i \notin S$ and J is chosen uniformly at random from S^C if $i \in S$. Now consider the binary hypothesis testing problem with observations $(G, J, S \backslash \{i, J\})$ and the task of distinguishing between $H_0 : i \notin S$ and $H_1 : i \in S$.

Since S is chosen uniformly at random, it follows that the identity of the vertex J is uniformly at random chosen from $[n] \backslash \{i\}$ and is independent of the event $\{i \in S\}$. It also holds that conditioned on J and the outcome of the event $\{i \in S\}$, the set $S \backslash \{i, J\}$ is uniformly distributed on $(k-1)$-subsets of $[n] \backslash \{i, J\}$. Therefore for any $J \in [n] \backslash \{i\}$ and $(k-1)$ subset $S \backslash \{i, J\}$ of $[n] \backslash \{i, J\}$, it holds that

$$
\frac{\mathbb{P}[G, J, S \backslash \{i, J\} | i \in S]}{\mathbb{P}[G, J, S \backslash \{i, J\} | i \notin S]} = \frac{\mathbb{P}[G, S \backslash \{i, J\} | i \in S, J]}{\mathbb{P}[G, S \backslash \{i, J\} | i \notin S, J]}
$$

$$
= \frac{\mathbb{P}[G | i \in S, J, S \backslash \{i, J\}]}{\mathbb{P}[G | i \notin S, J, S \backslash \{i, J\}]} = \prod_{k \in S \backslash \{i, J\}} \frac{(1 - A_{ik}) \cdot q^{A_{Jk}} (1 - q)^{1 - A_{Jk}}}{(1 - A_{Jk}) \cdot q^{A_{ik}} (1 - q)^{1 - A_{ik}}}
$$

where $A = A(G)$ is the adjacency matrix of G. From this factorization, it follows that the vector v of values A_{ik} and A_{Jk} for all $k \in S \backslash \{i, J\}$ is therefore a sufficient statistic for this binary hypothesis testing problem. Furthermore, if $i \in S$ then v has its first $k - 1$ coordinates equal to zero and its last $k - 1$ coordinates distributed as $\text{Bern}(q)^{\otimes(k-1)}$. If $i \notin S$, then v has its first $k - 1$ coordinates distributed as $\text{Bern}(q)^{\otimes(k-1)}$ and its last $k - 1$ coordinates equal to zero. Thus the given hypothesis testing problem is equivalent to testing between these two distributions. Note that

$$
\mathbb{P}_{H_0}[v = 0] = \mathbb{P}_{H_1}[v = 0] = (1 - q)^{k-1} \geq 1 - (k-1)q \to 1 \quad \text{as } n \to \infty
$$

by Bernoulli's inequality if $k \geq 2$. Taking any coupling of $\mathcal{L}_{H_0}(v)$ and $\mathcal{L}_{H_1}(v)$ such that the events $\{v = 0\}$ under H_0 and H_1 coincide yields that

$$
d_{\text{TV}} \left(\mathcal{L}_{H_0}(v), \mathcal{L}_{H_1}(v) \right) \leq 1 - (1 - q)^{k-1} \to 0 \quad \text{as } n \to \infty
$$

Now let p_{E1} and p_{E2} be the optimal Type I and Type II error probabilities. Note that the prior on the hypotheses $H_0 : i \in S$ and $H_1 : i \notin S$ is $\mathbb{P}[i \in S] = k/n$. Let $\mathcal{E}$ be the optimal average probability of testing error under this prior. Also note that $p_{\text{E1}} + p_{\text{E2}} = 1 - d_{\text{TV}} \left(\mathcal{L}_{H_0}(v), \mathcal{L}_{H_1}(v) \right) \to 1$ as $n \to \infty$.

Now assume for contradiction that there is some algorithm $A : \mathcal{G}_n \to \binom{[n]}{k}$ that achieves weak recovery with $\mathbb{E}[|S \cap A(G)|] = k - o(k)$ as $n \to \infty$. It follows that

$$
k - \mathbb{E}[|S \cap A(G)|] = \sum_{i=1}^{n} \mathbb{P}\left[\mathbf{1}_{\{i \in A(G)\}} \neq \mathbf{1}_{\{i \in S\}} \right] \geq \sum_{i=1}^{n} \min_{\phi_i(G)} \mathbb{P}\left[\phi_i(G) \neq \mathbf{1}_{\{i \in S\}} \right]
$$

$$
\geq \sum_{i=1}^{n} \min_{\phi_i(G, J, S \backslash \{i, J\})} \mathbb{P}\left[\phi_i(G, J, S \backslash \{i, J\}) \neq \mathbf{1}_{\{i \in S\}} \right] = n\mathcal{E}
$$

The first minimum is over all functions $\phi_i : \mathcal{G}_n \to \{0, 1\}$ that only observe the graph G, while the second minimum is over all functions ϕ_i that also observe J and $S \backslash \{i, J\}$. From these inequalities, it must follow that $\mathcal{E} = o(k/n)$ which implies the test achieving $\mathcal{E}$ must have Type I and Type II errors that are $o(1)$ as $n \to \infty$ since $1 - \frac{k}{n} = \Omega(1)$. This implies that $p_{\text{E1}} + p_{\text{E2}} = o(1)$, which is a contradiction. $\square$

Proof of Theorem 9.11. First suppose that $p > q$. Let G be an instance of $\text{PDS}_D(n, k, p, q)$. Note that under H_0, the edge count $|E(G)| \sim \text{Bin}(\binom{n}{2}, q)$ and under H_1, $|E(G)|$ is the independent sum of $\text{Bin}(\binom{n}{2} - \binom{k}{2}, q)$ and $\text{Bin}(\binom{k}{2}, p)$. By Bernstein's inequality, we have that

$$\mathbb{P}_{H_0}\left[|E(G)| > \binom{n}{2}q + \binom{k}{2} \cdot \frac{p-q}{2}\right] \leq \exp\left(-\frac{\binom{k}{2}^2(p-q)^2/4}{2\binom{n}{2}q + \binom{k}{2} \cdot (p-q)/3}\right)$$
$$= \exp\left(-\Omega\left(\frac{k^4}{n^2} \cdot \frac{(p-q)^2}{q(1-q)}\right)\right)$$

By the multiplicative Chernoff bound, it follows that

$$\mathbb{P}_{H_1}\left[|E(G)| \leq \binom{n}{2}q + \binom{k}{2} \cdot \frac{p-q}{2}\right] \leq \exp\left(-\frac{\binom{k}{2}^2(p-q)^2/4}{2\binom{n}{2}q + 2\binom{k}{2}(p-q)}\right)$$
$$= \exp\left(-\Omega\left(\frac{k^4}{n^2} \cdot \frac{(p-q)^2}{q(1-q)}\right)\right)$$

Now let X be the maximum number of edges over all subgraphs of G on k vertices. By a union bound and Bernstein's inequality

$$\mathbb{P}_{H_0}\left[X \geq \binom{k}{2} \cdot \frac{p+q}{2}\right] \leq \sum_{R \in \binom{[n]}{k}} \mathbb{P}_{H_0}\left[|E(G[R])| \geq \binom{k}{2} \cdot \frac{p+q}{2}\right]$$
$$\leq \left(\frac{en}{k}\right)^k \exp\left(-\frac{\binom{k}{2}^2(p-q)^2/4}{2\binom{k}{2}q + \binom{k}{2} \cdot (p-q)/3}\right)$$
$$= \exp\left(k\log(en/k) - \Omega\left(k^2 \cdot \frac{(p-q)^2}{q(1-q)}\right)\right)$$

where the second inequality uses the fact that for any fixed R, $|E(G[R])| \sim \text{Bin}(\binom{k}{2}, q)$. Under H_1, it holds that if S is the vertex set of the latent planted dense subgraph then $|E(G[S])| \sim \text{Bin}(\binom{k}{2}, p)$. By the multiplicative Chernoff bound, it follows that

$$\mathbb{P}_{H_1}\left[X < \binom{k}{2} \cdot \frac{p+q}{2}\right] \leq \mathbb{P}_{H_1}\left[|E(G[S])| < \binom{k}{2} \cdot \frac{p+q}{2}\right]$$
$$\leq \exp\left(-\frac{\binom{k}{2}^2(p-q)^2/4}{2\binom{k}{2}p}\right)$$
$$= \exp\left(-\Omega\left(k^2 \cdot \frac{(p-q)^2}{q(1-q)}\right)\right)$$

Therefore the test that outputs H_1 if $|E(G)| > \binom{n}{2}q + \binom{k}{2} \cdot \frac{p-q}{2}$ or $X \geq \binom{k}{2} \cdot \frac{p+q}{2}$ and H_0 otherwise has Type I+II error tending to zero as $n \to \infty$ if one of the two given conditions holds. In the case when $p < q$, this test with inequalities reversed can be shown to have Type I+II error tending to zero as $n \to \infty$ by analogous concentration bounds. $\square$

Proof of Theorem 9.22. This theorem follows from the Gaussian tail bound $1 - \Phi(t) \leq \frac{1}{\sqrt{2\pi}} \cdot t^{-1} e^{-t^2/2}$ for all $t \geq 1$ and a union bound. Now observe that if $(i, j) \notin \text{supp}(u) \times \text{supp}(v)$, then $M_{ij} \sim N(0, 1)$ and thus

$$\mathbb{P}\left[|M_{ij}| > \sqrt{6\log n}\right] = 2\left(1 - \Phi(\sqrt{6\log n})\right) \leq \frac{2}{\sqrt{2\pi}} \cdot e^{-3\log n} = O(n^{-3})$$

If $(i,j) \notin \operatorname{supp}(u) \times \operatorname{supp}(v)$, then $M_{ij} \sim N(\mu \cdot u_i v_j, 1)$ where $|\mu \cdot u_i v_j| \geq \sqrt{6 \log n}$ since $|u_i|, |v_j| \geq 1/\sqrt{k}$ by the definition of $\mathcal{V}_{n,k}$. This implies that

$$\mathbb{P}\left[|M_{ij}| \leq \sqrt{6 \log n}\right] \leq \left(1 - \Phi(\sqrt{6 \log n})\right) + \left(1 - \Phi(3\sqrt{6 \log n})\right) \leq \frac{2}{\sqrt{2\pi}} \cdot e^{-3 \log n} = O(n^{-3})$$

Now the probability that the set of (i,j) with $|M_{ij}| > \sqrt{6 \log n}$ is not exactly $\operatorname{supp}(u) \times \operatorname{supp}(v)$ is, by a union bound, at most

$$\sum_{(i,j) \in \operatorname{supp}(u) \times \operatorname{supp}(v)} \mathbb{P}\left[|M_{ij}| \leq \sqrt{6 \log n}\right] + \sum_{(i,j) \notin \operatorname{supp}(u) \times \operatorname{supp}(v)} \mathbb{P}\left[|M_{ij}| > \sqrt{6 \log n}\right] = O(n^{-1})$$

which completes the proof of the theorem. $\qquad\square$

Proof of Theorem 9.23. Let $\mathcal{P}_u$ denote the projection operator onto the vector u and let $\|r\|_0 = k_1$ and $\|c\|_0 = k_2$. By the definition of $\mathcal{V}_{n,k}$, it follows that

$$k\left(1 - \frac{1}{\log k}\right) \leq k_1, k_2 \leq k$$

By the argument in Lemma 1 of [CLR15], there are constants $C_2, C_3 > 0$ such that

$$\|\mathcal{P}_U B_{\cdot j} - \mu c_j r\|_2 \leq C_3 \sqrt{\log n} + C_3 \sqrt{\frac{n}{k_1}}$$

$$\|\mathcal{P}_V B_{i \cdot}^\top - \mu r_i c\|_2 \leq C_3 \sqrt{\log n} + C_3 \sqrt{\frac{n}{k_2}}$$

hold for all $1 \leq i, j \leq n$ with probability at least $1 - 2n^{-C_2} - 2\exp(-2C_2 n)$. Now note that if $j \in \operatorname{supp}(c)$ and $j' \notin \operatorname{supp}(c)$, then it follows that

$$\|\mu c_j r - \mu c_{j'} r\|_2 = \mu \cdot |c_j| \geq \frac{\mu}{\sqrt{k}}$$

by the definition of $\mathcal{V}_{n,k}$. Similarly, if $i \in \operatorname{supp}(r)$ and $i' \notin \operatorname{supp}(r)$ then $\|\mu r_i c - \mu r_{i'} c\|_2 \geq \mu/\sqrt{k}$. Therefore if for both $i = 1, 2$

$$\frac{\mu}{\sqrt{k}} \geq 6C_3\left(\sqrt{\log n} + \sqrt{\frac{n}{k_i}}\right)$$

then it holds that

$$2 \max_{j,j' \in \operatorname{supp}(c)} \left\|\mathcal{P}_U B_{\cdot j} - \mathcal{P}_U B_{\cdot j'}\right\|_2 \leq \max_{j \in \operatorname{supp}(c), j' \notin \operatorname{supp}(c)} \left\|\mathcal{P}_U B_{\cdot j} - \mathcal{P}_U B_{\cdot j'}\right\|_2$$

$$2 \max_{i,i' \in \operatorname{supp}(c)} \left\|\mathcal{P}_V B_{i \cdot}^\top - \mathcal{P}_U B_{i' \cdot}^\top\right\|_2 \leq \max_{i \in \operatorname{supp}(c), i' \notin \operatorname{supp}(c)} \left\|\mathcal{P}_V B_{i \cdot}^\top - \mathcal{P}_U B_{i' \cdot}^\top\right\|_2$$

and Steps 3 and 4 succeed in recovering $\operatorname{supp}(r)$ and $\operatorname{supp}(c)$. $\qquad\square$

Proof of Theorem 9.24. Under H_0, it holds that $M \sim N(0,1)^{\otimes n \times n}$. By Corollary 5.35 in [Ver10], we have

$$\sigma_1(M) \leq 2\sqrt{n} + \sqrt{2 \log n}$$

with probability at least $1 - 2n^{-1}$. Now consider the case of H_1 and suppose that $M = \mu \cdot rc^\top + Z$ where $r, c \in \mathcal{V}_{n,k}$ and $Z \sim N(0,1)^{\otimes n \times n}$. By Weyl's interlacing inequality, it follows that

$$|\mu - \sigma_1(M)| = |\sigma_1(\mu \cdot rc^\top) - \sigma_1(M)| \leq \sigma_1(Z) \leq 2\sqrt{n} + \sqrt{2 \log n}$$

with probability at least $1 - 2n^{-1}$. If $\mu > 4\sqrt{n} + 2\sqrt{2 \log n}$ then the Type I+II error of the algorithm is at most $4n^{-1}$, proving the theorem. $\qquad\square$

112

Proof of Theorem 9.25. Suppose that G is drawn from some distribution in H_1 and that the two hidden communities have index sets $S_1, S_2 \subseteq [n]$ where $k_1 = |S_1|$ and $k_2 = |S_2|$. For the remainder of the analysis of H_1, consider A and G conditioned on S_1 and S_2. Now let v be the vector

$$v_i = \begin{cases} \frac{1}{\sqrt{k_1+k_2}} & \text{if } i \in S_1 \\ -\frac{1}{\sqrt{k_1+k_2}} & \text{if } i \in S_2 \\ 0 & \text{otherwise} \end{cases}$$

for each $i \in [n]$. Now observe that

$$v^\top (A - qJ)v = \frac{2}{k_1 + k_2} \left(\sum_{(i,j) \in S_1^2 \cup S_2^2 : i < j} \left(\mathbf{1}_{\{i,j\} \in E(G)} - q \right) + \sum_{(i,j) \in S_1 \times S_2} \left(q - \mathbf{1}_{\{i,j\} \in E(G)} \right) \right)$$

By the definition of H_1 in SSBM$_D$, the expression above is the sum of $\binom{k_1+k_2}{2}$ independent shifted Bernoulli random variables each with expectation at least ρ. Therefore it follows that

$$\mathbb{E}\left[v^\top (A - qJ)v \right] \geq \frac{2}{k_1 + k_2} \cdot \binom{k_1 + k_2}{2} \cdot \rho = (k_1 + k_2 - 1)\rho \geq 3\sqrt{n}$$

since $k_1 + k_2 - 1 \geq k - 2k^{1-\delta_{\text{SSBM}}} - 1 \geq \frac{k}{2}$ for sufficiently large k, as defined in Section 2.2. Now note that each of the centered random variables $\mathbf{1}_{\{i,j\} \in E(G)} - \mathbb{E}[\mathbf{1}_{\{i,j\} \in E(G)}]$ and $\mathbb{E}[\mathbf{1}_{\{i,j\} \in E(G)}] - \mathbf{1}_{\{i,j\} \in E(G)}$ are bounded in $[-1,1]$ and therefore Bernstein's inequality implies that for all $t > 0$,

$$\mathbb{P}\left[v^\top (A - qJ)v < \mathbb{E}[v^\top (A - qJ)v] - \frac{2t}{k_1 + k_2} \right] \leq \exp\left(-\frac{\frac{1}{2}t^2}{n + \frac{1}{3}t} \right)$$

Note that v is a unit vector and thus $\lambda_1(A - qJ) \geq v^\top (A - qJ)v$. Setting $t = \frac{1}{2}(k_1 + k_2)\sqrt{n}$ now yields that

$$\mathbb{P}\left[\lambda_1(A - qJ) < 2\sqrt{n} \right] \leq \exp\left(-\frac{\frac{1}{8}(k_1 + k_2)^2 n}{n + \frac{1}{6}(k_1 + k_2)\sqrt{n}} \right) = \exp\left(-\Omega(n) \right)$$

since $k_1 + k_2 \geq \frac{k}{2} = \Omega(\sqrt{n})$ for sufficiently large k. Now suppose that G is drawn from $G(n,q)$ as in H_0. By Theorem 1.5 in [Vu05], it follows that with probability $1 - o_n(1)$, we have that

$$\lambda_1(A - qJ) \leq 2\sqrt{q(1-q)n} + C(q - q^2)^{1/4} n^{1/4} \log n$$

for some constant $C > 0$. Therefore $\lambda_1(A - qJ)$ is less than $2\sqrt{n}$ for sufficiently large values of n since $q(1-q) \leq 1/4$. Therefore the Type I+II error of this algorithm on SSBM$_D$ is $o_n(1) + \exp(-\Omega(n)) = o_n(1)$ as $n \to \infty$. $\qquad \square$

Proof of Theorem 9.28. Let u_S denote the d-dimensional unit vector with entries in S equal to $1/\sqrt{k}$ and all other entries equal to zero where S is some k-subset of d. Let v be the distribution on matrices $\theta u_S u_S^\top$ where $\theta \leq 1$ and S is chosen uniformly at random from the set of k-subsets of $[d]$. Note that $\mathcal{L}_{H_0}(X) = N(0, I_d)^{\otimes n}$ and $\mathcal{L}_{H_1}(X) = \mathbb{E}_v[N(0, I_d + \theta u_S u_S^\top)^{\otimes n}]$. By Lemma 9.27, it follows that

$$\chi^2\left(\mathcal{L}_{H_1}(X), \mathcal{L}_{H_0}(X) \right) = \mathbb{E}\left[\det\left(I_d - \theta^2 u_S u_S^\top u_T u_T^\top \right)^{-n/2} \right] - 1$$

$$= \mathbb{E}\left[\left(1 - \frac{\theta^2}{k^2} \cdot |S \cap T|^2 \right)^{-n/2} \right] - 1$$

$$\leq \mathbb{E}\left[\exp\left(\frac{n\theta^2}{k^2} \cdot |S \cap T|^2 \right) \right] - 1$$

113

where S and T are independent random k-subsets of $[d]$. The last inequality above follows from the fact that $(1-t)^{-1/2} \leq e^t$ if $t \leq 1/2$, $\theta^2 \leq 1/2$ and $|S \cap T| \leq k$. Now note that $|S \cap T| \sim$ Hypergeometric(d, k, k) and let

$$b = \frac{n\theta^2}{k^2} \cdot \left(\max \left\{ \frac{1}{k} \log \left(\frac{ed}{k} \right), \frac{d^2}{k^4} \right\} \right)^{-1}$$

The given condition on θ implies that $b \leq \beta_0$. It follows by Lemma 9.3 that $\chi^2 \left(\mathcal{L}_{H_1}(X), \mathcal{L}_{H_0}(X) \right) \leq \tau(\beta_0) - 1$ and by Cauchy-Schwarz that if $w(\beta_0) = \frac{1}{2}\sqrt{\tau(\beta_0) - 1}$ then

$$d_{\mathrm{TV}} \left(\mathcal{L}_{H_0}(X), \mathcal{L}_{H_1}(X) \right) \leq \frac{1}{2}\sqrt{\chi^2 \left(\mathcal{L}_{H_1}(X), \mathcal{L}_{H_0}(X) \right)} \leq w(\beta_0)$$

where $w(\beta_0) \to 0$ as $\beta_0 \to 0^+$, proving the theorem. $\qquad\square$

Proof of Theorem 9.34. Let $u = v_{\max}^k \left(\hat{\Sigma}(X) \right) - v$ and note that

$$\|u\|_2^2 = \|v\|_2^2 + \left\| v_{\max}^k \left(\hat{\Sigma}(X) \right) \right\|_2^2 - 2 \left\langle v, v_{\max}^k \left(\hat{\Sigma}(X) \right) \right\rangle = 2 \cdot L \left(v_{\max}^k \left(\hat{\Sigma}(X) \right), v \right)^2$$

If $i \in \mathrm{supp}(v)$ where $v \in \mathcal{V}_{d,k}$, then $|v|_i \geq \frac{1}{\sqrt{k}}$. Therefore each $i \in S(X)\Delta\mathrm{supp}(v)$ satisfies that $|u|_i \geq \frac{1}{2\sqrt{k}}$, which implies that

$$\frac{1}{k} |S(X)\Delta\mathrm{supp}(v)| \leq 4 \cdot \sum_{i \in S(X)\Delta\mathrm{supp}(v)} |u|_i^2 \leq 4\|u\|_2^2 \leq 8\sqrt{2} \cdot L \left(v_{\max}^k \left(\hat{\Sigma}(X) \right), v \right)$$

using the fact that $L(u, v) \leq \sqrt{2}$ if $\|u\|_2 = \|v\|_2 = 1$. This inequality along with the previous theorem completes the proof. $\qquad\square$

Proof of Theorem 9.35. First assume that H_0 holds and $X \sim N(0, I_d)^{\otimes n}$. Observe that

$$\frac{n}{d} \cdot \mathbf{1}^\top \hat{\Sigma}(X)\mathbf{1} = \frac{1}{d} \sum_{i=1}^n \langle \mathbf{1}, X_i \rangle^2$$

where the values $\frac{1}{\sqrt{d}}\langle \mathbf{1}, X_i \rangle$ are independent and distributed as $N(0, 1)$. Therefore $\frac{n}{d} \cdot \mathbf{1}^\top \hat{\Sigma}(X)\mathbf{1}$ is distributed as a χ^2 distribution with n degrees of freedom. Since $\frac{1}{d}\langle \mathbf{1}, X_i \rangle^2 - 1$ is zero-mean and sub-exponential with norm 1, Bernstein's inequality implies that for all $t \geq 0$

$$\mathbb{P}\left[\frac{n}{d} \cdot \mathbf{1}^\top \hat{\Sigma}(X)\mathbf{1} \geq n + t \right] \leq 2\exp\left(-c \cdot \min\left(\frac{t^2}{n}, t \right) \right)$$

for some constant $c > 0$. Substituting $t = \frac{2n\delta^2 k\theta}{d} \leq n$ yields that

$$\mathbb{P}\left[\mathbf{1}^\top \hat{\Sigma}(X)\mathbf{1} \geq d + 2\delta^2 k\theta \right] \leq 2\exp\left(-c \cdot \min\left(\frac{4n\delta^4 k^2\theta^2}{d^2}, \frac{2n\delta^2 k\theta}{d} \right) \right) = 2\exp\left(-\frac{4cn\delta^4 k^2\theta^2}{d^2} \right)$$

which tends to zero as $n \to \infty$. Now assume that H_1 holds and $X \sim N(0, I_d + \theta vv^\top)^{\otimes n}$ for some $v \in \mathcal{BV}_{d,k}$. Note that each X_i can be written as $X_i = \sqrt{\theta} \cdot g_i v + Z_i$ where $g_1, g_2, \ldots, g_n \sim_{\text{i.i.d.}} N(0, 1)$ and $Z_1, Z_2, \ldots, Z_n \sim_{\text{i.i.d.}} N(0, I_d)$. If $s(v) = \sum_{j=1}^d v_j$ is the sum of the entries of v, then

$$\langle \mathbf{1}, X_i \rangle = \sqrt{\theta} \cdot g_i s(v) + \sum_{j=1}^d Z_{ij} \sim N\left(0, d + \theta s(v)^2 \right)$$

Furthermore, these inner products are independent for $i = 1, 2, \ldots, n$. Therefore $\frac{n}{d+\theta s(v)^2} \cdot \mathbf{1}^\top \hat{\Sigma}(X)\mathbf{1}$ is also distributed as a χ^2 distribution with n degrees of freedom. Since $v \in \mathcal{BV}_{d,k}$, it either follows that $|\mathrm{supp}_+(v)| \geq \left(\frac{1}{2} + \delta\right) k$ or $|\mathrm{supp}_-(v)| \geq \left(\frac{1}{2} + \delta\right) k$. If $|\mathrm{supp}_+(v)| \geq \left(\frac{1}{2} + \delta\right) k$, then by Cauchy-Schwarz we have that

$$
\begin{aligned}
s(v) &= \sum_{i \in \mathrm{supp}_+(v)} v_i - \sum_{i \in \mathrm{supp}_-(v)} |v_i| \geq \left(\frac{1}{2} + \delta\right) \sqrt{k} - \left(\sum_{i \in \mathrm{supp}_-(v)} |v_i|^2 \right)^{1/2} \cdot \sqrt{|\mathrm{supp}_-(v)|} \\
&= \left(\frac{1}{2} + \delta\right) \sqrt{k} - \left(1 - \sum_{i \in \mathrm{supp}_+(v)} |v_i|^2 \right)^{1/2} \cdot \sqrt{|\mathrm{supp}_-(v)|} \\
&\geq \left(\frac{1}{2} + \delta\right) \sqrt{k} - \left(1 - \frac{|\mathrm{supp}_+(v)|}{k} \right)^{1/2} \cdot \sqrt{\left(\frac{1}{2} - \delta\right) k} \geq 2\delta\sqrt{k}
\end{aligned}
$$

Bernstein's inequality with $t = \frac{2n\delta^2 k\theta}{d+\theta s(v)^2} \leq n$ now implies that

$$
\mathbb{P}\left[\mathbf{1}^\top \hat{\Sigma}(X)\mathbf{1} \leq d + 2\delta^2 k\theta \right] \leq 2\exp\left(-\frac{4cn\delta^4 k^2 \theta^2}{(d + \theta s(v)^2)^2} \right)
$$

which tends to zero as $n \to \infty$ since $\theta s(v)^2 \leq \theta k \leq \frac{d}{2\delta^2}$ by Cauchy-Schwarz. This completes the proof of the theorem. $\square$

Proof of Theorem 9.39. First observe that $\hat{\Sigma}(X) = \frac{1}{n} X X^\top$ and thus $\lambda_1(\hat{\Sigma}(X)) = \frac{1}{n}\sigma_1(X)^2$. Under H_0, it follows that $X \sim N(0,1)^{\otimes d \times n}$. By Corollary 5.35 in [Ver10], it follows that

$$
\mathbb{P}\left[\lambda_1(\hat{\Sigma}(X)) > 1 + 2\sqrt{c} \right] \leq \mathbb{P}\left[\sigma_1(X) > \sqrt{n} + 2\sqrt{d} \right] \leq 2e^{-d/2}
$$

Under H_1, suppose that $X \sim N(0, I_d + \theta v v^\top)^{\otimes n}$ where $v \in \mathcal{V}_{d,k}$. As in the proof of Theorem 9.35, write $X_i = \sqrt{\theta} \cdot g_i v + Z_i$ where $g_1, g_2, \ldots, g_n \sim_{\text{i.i.d.}} N(0,1)$ and $Z_1, Z_2, \ldots, Z_n \sim_{\text{i.i.d.}} N(0, I_d)$. Now observe that

$$
v^\top \hat{\Sigma}(X)v = \frac{1}{n}\sum_{i=1}^n \langle v, X_i \rangle^2 = \frac{1}{n}\sum_{i=1}^n \left(\sqrt{\theta} \cdot g_i + \langle v, Z_i \rangle \right)^2
$$

Note that since $\|v\|_2 = 1$, it holds that $\sqrt{\theta} \cdot g_i + \langle v, Z_i \rangle \sim N(0, 1 + \theta)$ and are independent for $i = 1, 2, \ldots, n$. Now note that $(1 + \theta)^{-1}\langle v, X_i \rangle^2 - 1$ is zero-mean and sub-exponential with norm 1. Therefore Bernstein's inequality implies that

$$
\begin{aligned}
\mathbb{P}\left[\lambda_1(\hat{\Sigma}(X)) \leq 1 + 2\sqrt{c} \right] &\leq \mathbb{P}\left[v^\top \hat{\Sigma}(X)v \leq 1 + 2\sqrt{c} \right] \\
&\leq \mathbb{P}\left[\sum_{i=1}^n \left[(1 + \theta)^{-1}\langle v, X_i \rangle^2 - 1 \right] \leq -\frac{2n\sqrt{c}}{1 + \theta} \right] \\
&\leq 2\exp\left(-c_1 \cdot \frac{4cn}{(1 + \theta)^2} \right) \to 0 \text{ as } n \to \infty
\end{aligned}
$$

for some constant $c_1 > 0$ and since $\frac{2n\sqrt{c}}{1+\theta} < n$. This completes the proof of the theorem. $\square$

Proof of Theorem 9.40. Let $\hat{v} = g \cdot v + \sqrt{1 - g^2} \cdot u$ where u is the projection of $\hat{v}$ onto the space orthogonal to v. Also assume that $g \in [0, 1]$, negating v if necessary. By Theorem 4 in [Pau07], it follows that

$$g \to \sqrt{\frac{\theta^2 - c}{\theta^2 + \theta c}} \quad \text{with probability } 1 - o(1) \text{ as } n \to \infty$$

By Theorem 6 in [Pau07], u is distributed uniformly on the $(d-1)$-dimensional unit sphere of vectors in $\mathbb{R}^d$ orthogonal to v. By Lemma 4.1 in [KNV15], it holds that $|u_i| \le h\sqrt{\frac{\log d}{d}}$ for all $i \in [d]$ with probability tending to one. Condition on this event. Since $\frac{k \log d}{d} \to 0$, it follows that $\frac{\log d}{d} = o\left(\sqrt{\frac{\log d}{kd}}\right)$. Therefore with probability tending to one, each $i \notin \text{supp}(v)$ satisfies that $|\hat{v}_i|^4 < \frac{\log d}{kd}$ for sufficiently large n, k and d. Now note that each $i \in \text{supp}(v)$ satisfies that

$$|v_i| \ge \frac{g}{\sqrt{k}} - \sqrt{1 - g^2} \cdot |u_i| \ge \frac{g}{\sqrt{k}} - h\sqrt{\frac{\log d}{d}} \ge \sqrt[4]{\frac{\log d}{kd}}$$

for sufficiently large n, k and d with probability tending to one. This is because $g = \Omega(1)$ with probability $1 - o(1)$ and $\frac{1}{\sqrt{k}} = \omega\left(\sqrt{\frac{\log d}{kd}}\right)$. Therefore it follows that $S = \text{supp}(v)$ with probability tending to one as $n \to \infty$. $\qquad\square$

G Deferred Proofs from Section 10

Proof of Lemma 10.1. We first show that the distributions in Step 2 of PDS-CLONING are well-defined. First note that both are normalized and thus it suffices to verify that they are nonnegative. First suppose that $p > q$, then the distributions are well-defined if

$$\frac{1-p}{1-q} \le \left(\frac{P}{Q}\right)^{|v|_1} \left(\frac{1-P}{1-Q}\right)^{2-|v|_1} \le \frac{p}{q}$$

for all $v \in \{0, 1\}^2$, which follows from the assumption on P and Q. Now observe that if $\mathbf{1}_{\{i,j\} \in E(G)} \sim \text{Bern}(q)$ then x^{ij} has distribution

$$\mathbb{P}[x^{ij} = v] = q \cdot \mathbb{P}[x^{ij} = v | \{i, j\} \in E(G)] + (1 - q) \cdot \mathbb{P}[x^{ij} = v | \{i, j\} \notin E(G)] = Q^{|v|_1}(1 - Q)^{2 - |v|_1}$$

and hence $x^{ij} \sim \text{Bern}(Q)^{\otimes 2}$. Similarly, if $\mathbf{1}_{\{i,j\} \in E(G)} \sim \text{Bern}(p)$ then $x^{ij} \sim \text{Bern}(P)^{\otimes 2}$. It follows that if $G \sim G(n, q)$ then $(G^1, G^2) \sim G(n, Q)^{\otimes 2}$ and if $G \sim G(n, S, p, q)$ then $(G^1, G^2) \sim G(n, S, P, Q)^{\otimes 2}$, proving the lemma. $\qquad\square$

Proof of Lemma 10.2. Since the entries of M and G are independent, the σ-algebras $\sigma\{M_{ij}^1, M_{ij}^2\}$ for $i, j \in [n]$ are independent. Now note that M_{ij}^1 and M_{ij}^2 are jointly Gaussian and $\mathbb{E}[M_{ij}^1 M_{ij}^2] = \frac{1}{2} \cdot \mathbb{E}[M^2 - G^2] = 0$, which implies that they are independent. It follows that M_{ij}^k are independent for $k = 1, 2$ and $i, j \in [n]$. The lemma follows from the fact that each of M^1 and M^2 is identically distributed to $\frac{1}{\sqrt{2}}A + N(0, 1)^{\otimes n \times n}$. $\qquad\square$