Detection-Recovery and Detection-Refutation Gaps via Reductions from Planted Clique

Guy Bresler* Tianze Jiang†

Abstract

One fundamental topic in high dimensional statistics is to perform inference on a model with hidden planted structure inside noisy data. In particular, a large number of those problems have been shown to exhibit a phenomenon referred to as the statistical-computation gap, where the statistical limit of signal strength needed for a specific task is much smaller than what is needed for efficient algorithms. While there has been progress on proving such computation lower bounds with restricted classes of algorithms such as low-degree polynomials and the Sum of Squares hierarchy, we aim to establish a more rigorous reduction-based approach for proving (computation) lower bounds on this class of problems. Specifically, we try to construct reductions based on average-case hardness assumptions on the decision of *Planted Clique* (PC).

In this work, we are interested in a fundamental variant of PC: Planted Dense Subgraph (PDS). We study the average-case complexity of inference tasks of detection, refutation, and recovery of the hidden structure. We show that there exists a gap between hardness of detection with recovery and refutation in the PDS problem via average case reductions from the PC detection hypothesis and use it to imply similar results in Gaussian Biclustering and biased Sparse PCA. Our reduction generalizes complexity lower bounds proven and conjectured in other frameworks to a broad spectrum of (almost all existing) algorithms that abides to the PC hardness assumption. We also characterize statistical lower bounds for those problems.

1 Introduction

1.1 Statistical-Computation Gaps

A central class of problems in statistical inference is stated as the following form: how much data (signal strength) is needed to carry out a specific task on random structured models, and what is the limit of efficient algorithms? A variety of phenomenal works in the past have been focused on the information-theoretical limit of inference. However, in recent decades, an emerging field of studies focuses on the computational limit of (random) inference tasks when it is statistically possible instead and shows the phenomenon of the *statistical-computation gap*, where polynomial algorithms fail when statistically easy.

Two general approaches in proving computational hardness (when the problem is information-theoretically feasible) are through either proving guaranteed failure of powerful classes of (polynomial) algorithms ([Gam21][Jon+21][SW22]), or using the conventional complexity-theory wisdom to constructed (random) *average-case reductions* from one average-case hard problem to another

*Massachusetts Institute of Technology. Department of EECS. Email: guy@mit.edu.

†Massachusetts Institute of Technology. Department of EECS. Email: tjiang@mit.edu.

([BR13][MW15][BBH18]), very similar to how NP-hardness (completeness) was proven by constructing a worst-case reduction from SAT.

There has been numerous results on failures of specific types of algorithms. In [Bar+19], a novel technique of lower bounds on the Sum-of-Squares hierarchy via pseudocalibration was invented and employed to establish the tight hardness of the Planted Clique problem. In following work [Jon+21], another lower bound was proven in the SoS framework on a variation of Planted Clique. In [SW22; Rus+22], a full characterization of the average-case power of low-degree polynomials are inspected and tight lower bounds associated were proposed. In [Dec+11], a non-rigorous heuristic from statistical physics was provided to establish the hardness of a class of problems including recovering stochastic block models up to the *Kesten-Stigum* threshold, which is (roughly) the boundary of spectral methods. And very recently, a class of lower bound results focusing on the geometrics of algorithmic landscape ([Gam21]) has been established.

Progress on the reduction end of the attempts have been much fewer, partly because it is notably different from the well-studied worst-case reductions. One of the first works in this field, [BR13] has established average-case reduction to show (partial) hardness results of Sparse Principal Component Analysis (SPCA) via the classical Planted Clique hypothesis, which states that it is hard to detect between a Erdős-Renyi graph $G(n, 1/2)$ and one with a clique of size $k \in o(\sqrt{n})$ planted inside. In [WBS16][GMZ14][BB19b], a variety of average-case reduction techniques were introduced to complete the (tight) image of Sparse PCA from the Planted Clique (PC) hypothesis. In [MW15], an extension of Planted Clique called the Planted Dense Subgraph (PDS), in which a random $G(k, p)$ is planted in an instance of $G(n, q)$ with unknown location, was considered. While seemingly very close to PC, it is highly non-trivial ([MW15][HWX15a][BBH18]) to provide reductions from PC_D to PDS_D (the decision problem) without losing much information. More recently, in [BB20] a class of average-case methods were invented to transform Gaussian signals from a slightly refined hypothesis PC_ρ, proving hardness of a class of planted Gaussian detection problems. In this work, we will mostly focus on reductions towards Planted Dense Subgraphs and their applications.

1.2 Inference Tasks Beyond Decision

Aside from the detecting the existence of a planted signal, there are other fundamental inference tasks, such as refuting, certifying, and recovering the signal. It was proven ([HWX15a; Bar+19]) and intuitively so that refutation, certification, and recovery are all harder than simple detection. Indeed, *detection lower bounds imply certification, recovery and refutation lower bounds* (see Section 6.2). However, it is also noted that refutation and recovery seem intrinsically harder [HWX16; Jon+21; SW22] than detection for Planted Dense Subgraph and a variety of similar problems.

In this work, we try to answer the following question:

Where to find polynomial-time reductions from an average-case computational detection hardness hypothesis to harder objectives than detection on structured models?

by presenting several average-case reductions from Planted Clique to refutation and recovery for PDS, with natural implications extending to other planted structured models via reduction.

We first define those tasks formally. Consider H_0 as the null hypothesis (usually an Erdős-Renyi Graph), and H_1 is a graph with planted subset on support $v \in \{0, 1\}^n$. Consider a (non-polynomial) function val such that: $val(x)|x \sim H_0 < \delta - \epsilon$ and $val(x)|x \sim H_1 > \delta + \epsilon$ both with probability $1 - o_n(1)$. In PDS, val would be the densest-k-subgraph density function for G.

Refutation A refutation algorithm with successful probability p is a (randomized) algorithm $\mathcal{A}$ supported on all graphs with size n:

- If $val(G) > \delta + \epsilon$, then $\mathcal{A}(G) = 1$.
- For $G \sim H_0|_{val(G)<\delta-\epsilon}$, output $\mathcal{A}(G) = 0$ with probability at least p.

Recovery Consider the distribution π on planted support v. A recovery blackbox $\mathcal{A} : G \to \{0,1\}^n$ such that $\|\mathcal{A}(G)\|_1 = \|v\|_1$ matching the output/truth planted size are said to achieve (the expectation is taken over both $v \sim \pi$ and the randomness of graph conditioned on v):

1. *Partial recovery:* If $\mathbb{E}_{G_\pi}[v^T\mathcal{A}(G_v)] = \Omega(||v||_1)$.
2. *Weak recovery:* If $\mathbb{E}_{G_\pi}[v^T\mathcal{A}(G_v)] = \|v\|_1 - o(||v||_1)$.
3. *Exact (precise) recovery:* If $\mathbb{P}_{G_\pi}[\mathcal{A}(G_v) = v] = \Omega(1)$.

In most of the models considered, those variants only differ in sub-polynomial factors (via reduction as presented in Section 6.3). Moreover, note that in our setting we can replace π to be concentrated on one v as $\mathcal{A}$ can simply permute the nodes first. So the prior distribution π does not matter.

1.3 Planted Clique and Secret Leakage

We introduce the formal setting of the PC conjecture and PC_ρ, an alternative hardness assumption that was adopted in [BBH19][BB20]. First, consider the $\mathrm{PC}_D(n,k,p)$ as the hypothesis testing problem where $H_0 : G(n,p)$ and $H_1 : G(n,k,p)$ with a k node clique uniformly from $\binom{V}{k}$ planted.

Conjecture 1.1 (PC Conjecture)**.** *Fix some constant $p \in (0,1)$. Suppose that $\{A_n\}$ is a sequence of randomized polynomial time algorithms $A_n : G_n \to \{0,1\}$ 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 central idea for PC_ρ is that, instead of having a uniform prior on the location of the clique, we instead have a *secret leakage* of a tiny amount of information guarantees of where the planted clique lies. When the amount of extra information given to us is small enough, we may still assume that its corresponding hardness conjecture holds.

The setting we will consider is k-PC, where the tester knows *a priori* a partition of $[n]$ to k equal subsets, and is guaranteed that the planted set is sampled such that each subset contains exactly one node. In other words, the distribution on the planted clique is the product of each one node uniformly drawn in a subset. We refer to hardness assumption as the k-PC conjecture, and use it to similarly define k-PDS, the PDS variant. Consider the k-$\mathrm{PC}_D(n,k,p)$ as the hypothesis testing problem where $H_0 : G(n,p)$ and $H_1 : G_E(n,k,p)$ where E is a given partition and the planted clique location is sampled uniformly on $\prod_{i=1}^k \binom{E_i}{1}$:

Conjecture 1.2 (k-PC Conjecture)**.** *Fix some constant $p \in (0,1)$. Suppose that $\{A_n\}$ is a sequence of randomized polynomial time algorithms $A_n : G_n \to \{0,1\}$ 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 k-$\mathrm{PC}_D(n,k,p)$, it holds that*

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

We refer to [BB20] for an extensive discussion on the family of leakage PC conjectures.

1.4 Planted Graphical Models

Planted Dense Subgraph (PDS). Consider generalizing the setting of Planted Clique where we now define the distribution $\mathrm{PDS}(n, k, p, q)$ as follows:

1. Select a (random) subset S of vertices from $G \sim G(n, q)$ uniformly from subsets of size k.

2. Resample edges within nodes of S with probability $p > q$.

The planted dense subgraph problem is thus the hypothesis testing between:

$$H_0 : G \in G(n, q), \quad H_1 : G \sim \mathrm{PDS}(n, k, p, q). \tag{1}$$

In [MW15][HWX15a][BBH18], it was proven that the rate-optimal (statistical and computational) algorithm for detection in the PDS problem is to threshold the number of edges, which does not rely on the community structures and is *easy*. However, an important note from Section 1.2 is that while detection concerns both H_0 and H_1, all other tasks deals with only one part of the hypotheses and very weakly the other. To this end, we may turn to define qualifying "quite" hypotheses such that it has hard(er) decision task and imply similar recovery (refutation) guarantees.

Mean corrected Planted Dense Subgraph (PDS*). This is essentially PDS with a different H_0 to avoid the natural first degree test. Consider edge strengths $q < p_0 < p$ and size k such that:

$$p_0 = q + \gamma = p - (\frac{n^2}{k^2} - 1)\gamma$$

and formally define PDS* as hypothesis testing between:

$$H_0 : G \sim G(n, p_0), \quad H_1 : G \sim \mathrm{PDS}(n, k, p, q). \tag{2}$$

Imbalanced Stochastic Block Model (ISBM). Consider a two-community Stochastic Block Model $\mathrm{ISBM}(n, k, P_{11}, P_{12}, P_{22})$ to be the graph model generated by sampling $S_1 \sim \binom{[n]}{k}$ and $S_2 = [n] \setminus S_1$. Connect nodes $u \in S_i, v \in S_j$ with probability $P_{ij} = P_{ji}$.

With ISBM, if H_0 is $G(n, P_0)$, we are able to force the degree constraints *on each node*:

$$n \cdot P_0 = k \cdot P_{11} + (n - k) \cdot P_{12} = k \cdot P_{12} + (n - k) \cdot P_{22}$$

and formulate the decision problem ISBM_D as:

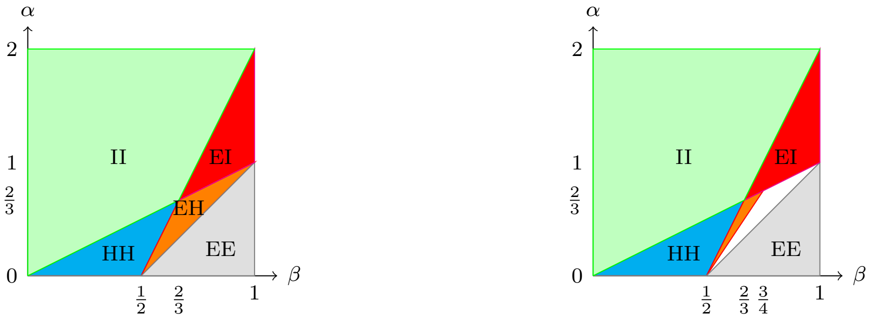
$$H_0 : G \sim G(n, P_0), \quad H_1 : G \sim \mathrm{ISBM}(n, k, P_{11}, P_{12}, P_{22}). \tag{3}$$

This model can be consider as a mean-field analogue of recovering a first community in a general k-block SBM model (maintaining one community and averaging out the rest).

1.5 Evidence and Obstacles for the Recovery Hardness

In this section we introduce the PDS recovery conjecture, which was an important open problem in various works such as [CX14][HWX15a][CLR17][BBH18][SW22]. The conjecture asserts the computational hardness of (even weakly) recovery of PDS. A matching upper bound can be achieved by various classes of algorithms such as spectral algorithms ([AWZ20]), semi-definite programming ([HWX16]), and low-degree polynomial ([SW22]). However, the lower bound in general is still left open except for specific classes of algorithms such as in [SW22; Rus+22] with low-degree polynomials and [AWZ20] for a class of MCMC algorithms. In [CLR17], a reduction-based approach on the subGaussian relaxation was stated based on detection hardness for finding a clique in a *regular* graph, which is fundamentally different from our problem[1] due to its dependency introduced.

[1] For instance, it is unclear how to sample a $n/2$-regular graph with a planted clique of any given size



(a) Detection vs Refutation.

(b) Detection vs Recovery.

Figure 1: The pictures above concerns $\text{PDS}(n, k, p, q)$ when p, q are bounded away from 0 and 1, and $k \in \widetilde{\Theta}(n^{\beta}), D_{KL}(p\|q) \in \widetilde{\Theta}(n^{-\alpha})$, where E denotes easy, H (computationally) hard, and I (statistically) intractable, hence the orange EH (computationally easy to detect but hard to refute/recover) is desired. Our statistical EI and computational EH characterization of refutation (left) in this density regime are both novel. The orange-white region in the right denotes the conjectural EH regime, where we prove hardness for orange and leave white open.

Conjecture 1.3 (PDS recovery conjecture). *Suppose $G_n \sim \text{PDS}(n, k_n, p_n, q_n)$ where $k_n > \sqrt{n}$ and:*

$$\limsup \log_n k^2 \frac{(p-q)^2}{q(1-q)} < 1,$$

then no polynomial algorithm $\mathcal{A} : G \to \binom{[n]}{k}$ can achieve exact recovery of PDS *asymptotically.*

As sharp detection boundaries can be shown from reduction, the next question is: can a sharp recovery boundary be shown via reduction as well? It turns out the answer is pessimistic. The main obstacle in proving such gap (as well as the refutation gap) is that suppose one has a reduction Φ that sends a PC_D instance to two classes of distributions where one is PDS (we call it H_1), then H_0 has to be indistinguishable from H_1 (which is why we introduce PDS^*_D). It turns out that a tightly "quiet" H_0 is extremely hard to find ([SW22]), and even for good H_0 candidates constructing Φ is likely hard if possible at all. Below we will show one reduction Φ to PDS^*_D at Section 3.

We finally remark that with some *relaxed* condition on the elevated signal, one can prove tight recovery hardness that resembles PDS in structure. See Theorem 6.2 and the discussions therein.

1.6 Main Results

In this work, our main technical contributions are:

- By introducing and characterizing detection hardness of a new variant of PDS, we present the first proof of computational *detection-recovery gap* (Corollary 3.3) for planted dense subgraph and planted Gaussian submatrix via an average-case reduction from $(\text{PC}_\rho)_D$.
- By studying the two-community Imbalanced Stochastic Block Model, we prove a log-optimal computational lower bound on refutation for densest k-subgraphs in $G(n, p)$ and densest principal submatrix in $\mathcal{N}(0, 1)^{\otimes(n \times n)}$ assuming $(\text{PC}_\rho)_D$ (*detection-refutation gap*) (Theorem 4.2), extending a couple of algorithmic results ([Bar+19; Jon+21]) via reductions.

Finally, in the last section we will discuss reductions to other problems of interests such as binomially sized planted subgraph, Gaussian biclustering, and biased sparse principle component analysis and our reductions' implication on their lower bounds. We also provide arguments on statistical boundaries for the above problems, including a sharp boundary on refuting densest k-subgraphs in Erdő's Renyi Graphs via reductions from recovery.

2 Techniques Overview

2.1 Distribution shift and Cloning

We present the fundamental lemmas of distribution shifting. In short, the idea is that performing transformations on Bernoulli random variables without loss is hard, but we can frontload the loss (in total variation) by first transforming into another distribution (such as Gaussian or Poisson random variables), and perform (almost lossless) transformation of signals to obtain the desired target with minimal loss in total variation. This idea was exploited in various literature such as [MW15][BBH18][BB19a][BB19b]. In this work, we start with the basic lemmas of rejection sampling that turns an instance of bit $x \in \{0,1\}$ from one of two known Bernoulli distributions $\mathrm{Bern}(p)$, and $\mathrm{Bern}(q)$ to two known distributions P_1 and P_2, respectively and that the KL divergence between P_1 and P_2 is almost preserved according to Data Processing Inequality. Specifically, in the scope of this work we are concerned about the cases when P_1, P_2 are (approximately in d_{TV}):

- A scalar Gaussian distribution $\mathcal{N}(\mu, 1)$ with mean μ.
- Tensor products of Bernoulli random variables $\mathrm{Bern}(\lambda)^{\otimes t}$ with specified bias.

Lemma 1 (Gaussian Rejection Kernels – Lemma 5.4 in [BBH18]). *Let R_{rk} be a parameter and suppose that $p = p(R_{rk})$ and $q = q(R_{rk})$ satisfy that $0 < q < p \le 1$, $\min(q, 1-q) = \Omega(1)$ and $p - q \ge R_{rk}^{-O(1)}$. Let $\delta = \min\left\{\log\left(\frac{p}{q}\right), \log\left(\frac{1-q}{1-p}\right)\right\}$. Suppose that $\mu = \mu(R_{rk}) \in (0,1)$ satisfies that*

$$\mu \le \frac{\delta}{2\sqrt{6\log R_{rk} + 2\log(p-q)^{-1}}}$$

Then the map RK_G with $N = \lceil 6\delta^{-1}\log R_{rk}\rceil$ iterations can be computed in poly(R_{rk}) time and satisfies

$$d_{\mathrm{TV}}\left(\mathrm{RK}_G(\mu, \mathrm{Bern}(p)), \mathcal{N}(\mu,1)\right) = O\left(R_{rk}^{-3}\right) \quad \textit{and} \quad d_{\mathrm{TV}}\left(\mathrm{RK}_G(\mu, \mathrm{Bern}(q)), \mathcal{N}(0,1)\right) = O\left(R_{rk}^{-3}\right).$$

Another Bernoulli transformation allows us to produce two independent graph instances following the same distribution from one with new density parameters, which will help us producing a bipartite graph from an undirected graph with the same (unknown) underlying latent structure[2].

Lemma 2 (Graph Cloning – Lemma 5.2 in [BBH19]). *Let $t \in \mathbb{N}$, $0 < q < p \le 1$ and $0 < Q < P \le 1$ satisfy that*

$$\frac{1-p}{1-q} \le \left(\frac{1-P}{1-Q}\right)^t \quad \textit{and} \quad \left(\frac{P}{Q}\right)^t \le \frac{p}{q}$$

Then the algorithm $\mathcal{A} = $ Graph-Clone runs in poly(t, n) *time and satisfies that for each $S \subseteq [n]$,*

$$\mathcal{A}\left(G(n,q)\right) \sim G(n,Q)^{\otimes t} \quad \textit{and} \quad \mathcal{A}\left(\mathrm{PDS}(n,S,p,q)\right) \sim \mathrm{PDS}(n,S,P,Q)^{\otimes t}$$

2.2 Bernoulli Rotations

In this section we introduce a key reduction technique Bern-Rotations first explored by [BB20], which turns a *planted bit* distribution, in which a $\mathrm{Bern}(p)$ bit has been planted on the ith place into a vector $\mathrm{Bern}(q)^{\otimes n}$, into a Gaussian vector with identity variance and specified mean.

In short, the ideas behind Bern-Rotations are with the following:

[2] While it seems natural from cloning a Bernoulli bit that the one can map from PDS to its bi-partite variant, the diagonal part on the output still needs to be carefully handled, see Lemma 17

1. After performing bit-wise Gaussianization by Lemma 1, we are looking at an instance of $\mu e_i + \mathcal{N}(0, I_n)$ where e_i is the one-hot encoding of signal location.

2. Then, we hit the entire vector by a *design matrix* $A \in \mathbb{R}^{m\times n}$, which will turn the instance into $\mu A_i + \mathcal{N}(0, AA^T)$ where A_i is the ith column of A. Denote the square of the top-singular value of A to be $\lambda = \sigma^2(A)$.

3. On the result vector $\mathcal{N}(\mu\lambda^{-1/2}A_i, AA^T/\lambda)$, we can add a constant noise $\mathcal{N}(0, I - AA^T/\lambda)$ (note that the variance term is PSD) to get exactly $N(\frac{\mu}{\sigma(A)}A_i, I_n)$.

Formalizing the above process, we get the following lemma:

Lemma 3 (Dense Bernoulli Rotations – Lemma 8.1 in [BB20])**.** *Let m and n be positive integers and let $A \in \mathbb{R}^{m\times n}$ be a matrix with singular values all at most $\lambda > 0$. Let R_{rk}, $0 < q < p \leq 1$ and μ be as in Lemma 1. Let $\mathcal{A}$ denote* Bern-Rotations *applied with rejection kernel parameter R_{rk}, Bernoulli probability parameters $0 < q < p \leq 1$, output dimension m, matrix A with singular value upper bound λ and mean parameter μ. Then $\mathcal{A}$ runs in* $\text{poly}(n, R_{rk})$ *time and it holds that*

$$d_{\text{TV}}\left(\mathcal{A}\left(\text{PB}(n,i,p,q)\right), \mathcal{N}\left(\mu\lambda^{-1}\cdot A_i, I_m\right)\right) = O\left(n \cdot R_{rk}^{-3}\right)$$
$$d_{\text{TV}}\left(\mathcal{A}\left(\text{Bern}(q)^{\otimes n}\right), \mathcal{N}\left(0, I_m\right)\right) = O\left(n \cdot R_{rk}^{-3}\right)$$

for all $i \in [n]$, where $\text{PB}(n,i,p,q)$ *is the distribution on $\{0,1\}^{\otimes n}$ where the ith bit is sampled from* $\text{Bern}(p)$ *and all others from* $\text{Bern}(q)$ *independently.*

2.3 Random Construction of Design Matrices

In this section we establish a couple of important lemmas constructing (a set of) random matrices that are crucial to our reduction. The central idea for constructing a design matrix A for Bern-Rotations is that, we need the output mean vector to correspond to the target output (edge probability, or Gaussian mean), and we want the transformation matrix to have operator norm at most 1 (constant). It is also worth noting that we need to map precisely to the desired boundary of the target distribution obtained by an upper bound. In other words, Bern-Rotations applied with A cannot lose much information, otherwise one can map to any instance of wish and simply scale it down by $\sigma(A)$. We refer to Section 6.4.1 for a detailed discussion.

A line of works ([BS87][TY19][LLV15][BH16]) have given high probability bounds on the spectral of concentration $\|A - \mathbb{E}(A)\|_{op}$ of adjacency matrix A for a random graph G with given degree distributions (corresponding to our fraction of signal variables). Here we are interested in the case when A is the adjacency matrix of a directed d-regular graph (each node has out-degree and in-degree exactly d). In this case the operator norm of concentration can be expressed with the second largest singular value of A. In [TY19], a (tight) high probability upper bound on the said quantity have been proven when $n^\alpha < d < n/2$:

$$|s_2(A)| \leq C_{\alpha,m}\sqrt{d}$$

with probability at least $1 - n^{-m}$ for a constant C independent with n, d for the second largest singular value where α, m are constants. With this result we can establish the following:

Lemma 4 (Random Matrix with regular constraints)**.** *Given constant $\alpha > 0$, there exist constant C_α, such that for a $n \times n$ (random) matrix $R = R_{n,1/r}$ where $r < n^{1-\alpha}$ is an even divisor of n, with entries sampled from the following procedure:*

1. *Sample G uniformly from all directed n/r-regular graphs with size n.*

2. $R_{ij} = \frac{-1}{\sqrt{nr}} + 1_{e_{ij} \in E_G} \cdot \sqrt{\frac{r}{n}}$ *for $j \neq i$ off diagonal.*

3. $R_{ii} = \frac{-1}{\sqrt{nr}}$ *on the diagonal.*

Then with probability $1 - o_n(1)$ this matrix satisfies:

$$\|R\|_{op} \leq C_\alpha.$$

Note that we can sample from directed regular graphs efficiently by bounding the mixing time of a simple edge-flipping Markov process by polynomials, as proven in works such as [Gre11][Coo+17]. With a polynomial sampler, we have the following result for generating design matrices for PDS*:

Lemma 5 (Construction of (fixed size) random $K_n^{1/r}$)**.** *For given α, exist absolute constant $C_\alpha > 0$, such that for every $n > r > 2$ where $r < n^{1-\alpha}$ divides n, there exist n subsets $A_1, A_2, \ldots, A_n$ of $[n]$ such that $|A_i| = \frac{n}{r}$, and that the $n^2 \times n^2$ symmetric matrix $K_{(ij),(kl)} : i, j, k, l \in [n]$ defined as follows ($\mu = (C_\alpha + 1)^{-2} \in \Theta_n(1)$):*

$$K_n^{1/r} := K_{(ij),(kl)} = \mu \sqrt{\frac{r}{n}} \cdot (1_{k \in A_i \text{ and } l \in A_j} \cdot \frac{r}{n} - \frac{1}{nr})$$

(observe that K has sum of entries 0) has largest singular value at most 1. Specifically, K has the following form:

$$K = \mu \sqrt{\frac{r}{n}} \left[(R + \frac{1}{\sqrt{nr}} J) \otimes (R + \frac{1}{\sqrt{nr}} J) - \frac{1}{nr} J \otimes J \right].$$

where J is the all-one matrix and R satisfies the criteria from the previous lemma. Finally, with probability over $1 - o_n(1)$ we can find a satisfying assignment in polynomial time.

3 Hardness for detection in degree-1 corrected null hypothesis

We present a formal statement on detection hardness for the degree-1 corrected null hypothesis testing by constructing an average case mapping in this section. This result, combined Lemma 16, completely resolves the computational PDS$_D^*$ problem (2) in the dense regime within log factors.

Theorem 3.1 (Reduction to PDS*)**.** *Given any fixed constant $\alpha > 0$. Let N, k_0 be parameters of planted clique graph size, (n, k) be the target graph sizes where $\frac{n}{k} =: r < \left(\frac{N}{k_0}\right)^{1-\alpha}$. We present the following reduction ϕ with absolute constant $C > 1$:*

- Initial k-PDS Parameters: *vertex count N, subgraph size $k_0 \in o_N(N)$ dividing N, edge probabilities $0 < q < p \leq 1$ with $\min\{q, 1-q, p-q\} = \Omega(1)$, and a partition E of $[N]$. We further assume that $k_0 \in o(\sqrt{N})$ holds (otherwise detection for the PDS problem will be easy).*

- Target PDS* parameters: *(n, r, k) where $r \in o(\sqrt{n})$ is a specified parameter, $k = n/r$ is the target subgraph size, and n is the smallest multiple of $k_0 r$ greater than $(1 + \frac{p}{Q})N$ where*

$$Q = 1 - \sqrt{(1-p)(1-q)} + 1_{p=1}(\sqrt{q} - 1)$$

is the cloned signal strength from pre-processing.

- Target PDS* edge strength:

$$\gamma = \mu(\frac{k_0 r}{n})^{1.5}, \quad P_1 = \Phi(\frac{(r^2-1)\gamma}{r^2}), \quad P_2 = \Phi(-\frac{\gamma}{r^2}),$$

  where $\mu \in (0,1)$ satisfies that

$$\mu \leq \frac{1}{12C\sqrt{\log(N) + \log(p-Q)^{-1}}} \cdot \min\{\log(\frac{p}{Q}), \log(\frac{1-Q}{1-p})\}.$$

  where γ denotes the signal strength $\gamma = \Theta(D_{KL}(P_1 \| P_2))$ roughly the KL-divergence between two output Bernoullis.

- *Applying ϕ on the given input graph instance G yields the following:*

$$d_{\mathrm{TV}}(\phi(G(N, \frac{1}{2})), G(n, \frac{1}{2})) = o_n(1)$$

$$d_{\mathrm{TV}}(\phi(PC_E(N, k_0, \frac{1}{2})), \mathrm{PDS}(n, k, P_1, P_2)) = o_n(1)$$

Finally we conclude the boundary by applying the following procedures:

1. Generalizing the above reduction to all density $\{P_0 : \min\{P_0, 1-P_0\} \in \Omega(1)\}$ by diluting the signals by a constant. Then, with some care in bounding total variation and computing the signal ratio we retain the same lower bound for any P_0.

2. Getting rid of Gaussian densities with approximations so that the average degree in $\phi(G(N, 1/2))$ matches exactly with the average degree in $\phi(PC_\rho)$. This extra step ensures that the (random) approximate density in constructing the design matrices aligns with PDS*.

Theorem 3.2 (Lower bounds for efficient PDS* detection)**.** *Consider hypothesis testing* PDS* *for $H_0 : G(n, p_0)$ versus H_1 :* $\mathrm{PDS}(n,k,q,p)$ *where $p_0 = p - (\frac{n^2}{k^2} - 1)\gamma = q + \gamma$. Let parameters $p_0 \in (0,1)$, $\alpha \in [0,2), \beta \in (0,1)$ and $\beta < \frac{1}{2} + \frac{2}{3}\alpha$. There exist a sequence $\{(N_n, K_n, p_n, q_n)\}$ of parameters such that:*

- *The parameters are in the regime $\gamma \in \widetilde{\Theta}(N^{2\beta-\alpha-2})$, $K \in \widetilde{\Theta}(N^{\beta})$. Or formally:*

$$\lim_{n\to\infty} \frac{\log \gamma_n}{\log N_n} = 2\beta - \alpha - 2, \quad \lim_{n\to\infty} \frac{K_n}{N_n} = \beta, \quad \lim_{n\to\infty} \frac{\log(p_n - q_n)^{-1}}{\log N_n} = \alpha.$$

- *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* $\mathrm{PDS}^*(N_n, K_n, p_n, q_n)$ *is at least 1 assuming the PC conjecture holds with density $p = 1/2$.*

Furthermore, we note that there exists a matching upper bound for PDS^*_D based on the empirical variance of degrees. See Section 6.5 for a full statement.

With this theorem, we can now state one of our main results for recovery hardness in the PDS problem, which improves upon the prior known bound by a coefficient on the exponent.

Corollary 3.3 (Recovery Hardness for PDS)**.** *Let parameters $p_0 \in (0,1)$, $\alpha \in [0,2), \beta \in (0,1)$ and $\alpha < \beta < \frac{1}{2} + \frac{2}{3}\alpha$,. Then for any $p_0 \in (0,1)$ there exists a sequence $\{(N_n, K_n, p_n, q_n)\}$ of parameters such that the following holds:*

Algorithm From k-PDS to PDS*:
Inputs: Graph G of size N, subgraph parameter k_0 dividing N, edge density $q < p \in (0, 1]$ and a partition E of $[N]$ to k_0 equal parts $E_1, E_2, \dots, E_t$. Target planted ratio r.
Steps :

1. *To-bipartite and planted diagonal:* The first step transforms PDS to a bipartite variant. Let n be the smallest integer multiple of k_0 that is greater than $(1 + \frac{p}{Q})N$. Apply Graph Cloning to input G to obtain G_1 and G_2 with edge density $Q < p$ where:

$$Q = 1 - \sqrt{(1-p)(1-q)} + 1_{p=1}(\sqrt{q} - 1)$$

   then construct $F \in \mathbb{R}^{n \times n}$ equipped with a partition $S_1, S_2, \dots, S_{k_0}$ of $[n]$ such that:

   - Given that each $|S_i| = n/k_0$, (uniformly) sample a random subset T_i in S_i of size N/k_0. Construct (any) bijective map $\pi_i : T_i \to E_i$.
     Sample a subset $X_i \subset T_i$ where each element is included independently with probability p and sample $y_i \sim \max\{\text{Bin}(n/k_0, Q) - |X_i|, 0\}$. Sample subset $Y_i \subset (S_i \setminus T_i)$ (with size y_i) uniformly from $\binom{S_i \setminus T_i}{y_i}$.
   - Construct F for each F_{S_i,S_j} in the following fashion:
     - If $i \neq j$, then:

$$F_{T_i,T_j} = \begin{cases} G_1[\pi_i(T_i), \pi_j(T_j)] & i > j \\ G_2[\pi_j(T_j), \pi_i(T_i)] & i < j \end{cases}$$

$$F_{(i,j) \in S_i \times S_j \setminus T_i \times T_j} \sim \text{Bern}(Q).$$

     - For the diagonal blocks:

$$F_{T_k,T_k}(i,j) = \begin{cases} G_1[\pi_k(T_k), \pi_k(T_k)]_{ij} & i < j \\ G_2[\pi_k(T_k), \pi_k(T_k)]_{ij} & i > j \\ \mathbf{1}\{i \in X_i\} & i = j \end{cases}$$

$$F_{(i,j) \in S_k \times S_k \setminus T_k \times T_k} = \begin{cases} \sim \text{Bern}(Q) & \text{i} \neq j \\ \mathbf{1}\{i \in y_i\} & i = j \end{cases}.$$

2. *Flattened Bernoulli Rotations:* Let S be a partition of $[n]$ into k_0 equal parts $S_1, S_2, \dots, S_{k_0}$ obtained from the previous part. Construct output matrix M:

   (a) For i, j in $\{1, 2, \dots, k_0\}$, flatten matrix F_{S_i,S_j} to a $(n/k_0)^2$ size vector.

   (b) Apply Bernoulli Rotation on this vector with design matrix $(K^{1/r}_{n/k_0})^T$, Bernoulli parameter strengths $Q < p \leq 1$, output dimensions $v_{ij} \in \mathbb{R}^{(n/k_0)^2}$.

   (c) Layout vector $v_{ij} \in \mathbb{R}^{(n/k_0)^2}$ to $(n/k_0) \times (n/k_0)$ matrix in the order from part (a). Apply permutation to $[n]$.

3. *Thresholding:* Given matrix M from the previous step, construct $G' = \phi(G)$ such that: for distinct indices $i < j$, $e_{ij} \in E(G')$ if and only if $M_{ij} \geq 0$ and output.

Figure 2: Reduction from k-PDS to PDS*.

- *The parameters are in the regime* $\gamma := D_{KL}(p\|q) \in \widetilde{\Theta}(N^{-\alpha})$, $K \in \widetilde{\Theta}(N^{\beta})$.
- *For any sequence of (randomized) polynomial-time algorithm* $\phi_n : \mathcal{G}_{N_n} \to \binom{[N_n]}{K_n}$, ϕ_n *cannot achieve asymptotic weak recovery on* $\mathrm{PDS}(N_n, K_n, p_n, q_n)$.

We note that the constraint $\alpha < \beta$ comes from the fact that weak recovery is statistical impossible at $\alpha \geq \beta$ (see Theorem 6.1). For completeness, we defer to the Appendix (Theorem 6.3) for an extended discussion on the statistical boundaries associated.

4 Hardness results for refutation.

4.1 Detection hardness for ISBM

In dealing with refutation, one usually want to find some "quiet" distribution, such that it has the correct valuation but is hard to distinguish from a null instance. We thus propose the ISBM model (3) in this section as a qualifying planted distribution. The hardness result can be proven using a similar procedure as before leading to a reduction. As in the proof of reduction to PDS*, we can then generally establish the complete boundary in ISBM detection. This is an extension of Theorem 3.2 in [BB20] where their (deterministic rotation kernel) reduction only works with a number theoretic constraint restricted to a relatively small set of parameters (most notably $r = n^{1/2t} \in O(n^{1/4})$ where $t > 1$ is an integer). Here we extend their results to the full boundary line by the regular concentration lemma on random matrices.

Theorem 4.1 (Hardness of detection in ISBM)**.** *Consider hypothesis testing* ISBM_D *for* $H_0 : G(n, p_0)$ *versus* $H_1 : \mathrm{ISBM}(n, r, P_{11}, P_{12}, P_{22})$ *where*

$$p_0 = P_{11} - \gamma = P_{12} + \frac{\gamma}{r-1} = P_{22} - \frac{\gamma}{(r-1)^2}.$$

Let parameters $p_0 \in (0, 1)$, $\alpha \in [0, 2), \beta \in (0, 1)$ *and* $\beta > \frac{1}{2} - \alpha$. *There exist a sequence* $\{(N_n, R_n, P_{11}^{(n)}, P_{12}^{(n)}, P_{22}^{(n)})\}$ *of parameters such that:*

- *The parameters are in the regime* $\gamma \in \widetilde{\Theta}(N^{-\alpha})$, $R \in \widetilde{\Theta}(N^{\beta})$. *Or formally:*

$$\lim_{n\to\infty} \frac{R_n}{N_n} = \beta, \quad \lim_{n\to\infty} \frac{\log(P_{11}^{(n)} - p_0)^{-1}}{\log N_n} = \alpha.$$

- *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 decision problems* $\mathrm{ISBM}_D(N_n, R_n, P_{11}^{(n)}, P_{12}^{(n)}, P_{22}^{(n)})$ *will be at least 1 assuming the PC conjecture holds with density* $p = 1/2$.

4.2 Refutation hardness for planted dense subgraph in $G(n, p)$.

Equipped with the hardness results in ISBM, which has a nice large dense subgraph, we obtain the formal hardness results in refutation as follows:

Theorem 4.2 (Hardness in refutation of PDS in the dense regime)**.** *Consider refutation problem for* $H_0 : G(n, p_0)$ *and val function* $v(G)$ *defined as the edge density of the largest* $k-$*subgraph. Let parameters* $p_0 \in (0, 1)$, $\alpha \in [0, 2), \beta \in (0, 1)$ *and* $\beta > \frac{1}{2} - \alpha$. *Then for any sequence of parameters* $\{(N_n, K_n, p_1^{(n)}\}$ *satisfying:*

- *The parameters are in the regime $p_1 - p_0 \in \widetilde{\Theta}(N^{-\alpha})$, $K \in \widetilde{\Theta}(N^{\beta})$. Or formally:*

$$\lim_{n\to\infty} \frac{K_n}{N_n} = \beta, \quad \lim_{n\to\infty} \frac{\log(p_1^{(n)} - p_0)^{-1}}{\log N_n} = \alpha.$$

- *No sequence of (randomized) polynomial-time algorithms ϕ_n can achieve refutation with asymptotic successful probability strictly above* 0.

Finally, we note that a matching (computational) upper bound can be constructed via a semidefinite programming relaxation (see appendix). Moreover, the statistical boundary for refutation lies exactly as the statistical boundary for recovery, ignoring log factors, from looking at the distribution of densest k subgraph strength and applying a reduction to recovery.

Theorem 4.3 (Statistical bounds for refutation)**.** *Consider refutation problem for $G \sim G(n, p_0)$ and val function $v(G)$ defined as the edge density of the largest $k-$subgraph. Assuming that p_0 is bounded away from 0 and 1, and $k \in \widetilde{\Theta}(n^{\gamma})$ for some $\gamma \in (0.5, 1)$, then:*

- *When $kD_{KL}(p\|p_0) \in \widetilde{\omega}(1)$, the densest k subgraph $val(G) \leq p$ with probability $\to 1$.*
- *When $kD_{KL}(p\|p_0) \in \widetilde{o}(1)$, the densest k subgraph $val(G) \geq p$ with probability $\to 1$.*

5 Extensions and discussions

5.1 Reduction from fixed planted size to Binomial planted size

In the contexts of Planted Clique and planted dense subgraphs (Gaussian submatrices), two different settings are usually considered: fixed planted size k (most of the reductions above) as well as binomial k where each node is selected with fixed probability $\rho = k/n$. They correspond to step 1 of generating H_1 in Section 1.2. While those two settings are generally considered similar (because the binomial distribution is highly concentrated on its mean when n is large), many tricks will not work on both settings, which appears in works such as [Alo+07][HWX15a], unless with certain stronger assumptions on either the class of algorithms or on the correctedness guarantees.

In the scope of this paper, most of our works are done in the setting of fixed size planted set. However, we note that with a lemma from spectral concentration on Erdős-Renyi graphs, we would be able to establish an average-case reduction from a **single** fixed k submatrix to binomial submatrix detection without extraneous assumptions or properties on the model.

Theorem 5.1 (Fixed k to Binomial)**.** *Ignoring* log*-scale losses on the parameters, we have:*

1. *For the decision and recovery variant of the planted dense subgraph problem, there exists a reduction from fixed planted size k to binomial distributed $k \sim \mathrm{Binom}(n, k/n)$ where the edge densities $|p' - q'| \in \widetilde{\Theta}(|p - q|)$ (at most loses poly-*log *scaled signal) when $\{p, q, 1 - p\} \subset \Omega(1)$.*
2. *There exists reduction from PC conjecture with fixed community size to binomial sized community of the detection variant of* PDS* *and* ISBM *with fixed parameter such that the computational barrier remains the same.*

Finally, we note that the first part of the above statement states that the PC conjecture of fixed size is "stronger" than its binomial variant. Whether this strong-ness is strict remains open.

5.2 Biclustering and biased sparse PCA

In this section we point out a couple of other models that have a detection hardness gap as a result of PDS hardness. Those connections were first observed in [CLR17][BBH18] but under the conjectural tight hardness bound. In our framework, by performing some post-processing, we see that our reduction techniques apply directly from the PC hypothesis.

Bi-clustering This model is planting a $k \times k$ (not necessarily principal) submatrix and can be formulated as the following Gaussian detection problem:

$$H_0 : Z \sim \mathcal{N}(0,1)^{\otimes n \times n}, \qquad H_1 : Z \sim \mathcal{N}(0,1)^{\otimes n \times n} + \lambda u v^T \tag{4}$$

where $u, v \sim \mathrm{Bern}(k/n)^{\otimes n}$ (or uniform from all subsets of size k) independently. The recovery problem is to localize the latent vectors u, v given an instance $Z \sim \mathcal{N}(0,1)^{\otimes n \times n} + \lambda u v^T$, and the refutation task is to refute a $k \times k$ submatrix with large mean.

Biased SPCA Consider the *spiked covariance model*:

$$\begin{aligned} &H_0 : X_1, X_2, \ldots, X_n \sim \mathcal{N}(0, I_d)^{\otimes n} \quad \text{and} \\ &H_1 : X_1, X_2, \ldots, X_n \sim \mathcal{N}\left(0, I_d + \theta v v^\top\right)^{\otimes n} \text{ where } v \in \mathrm{Unif}[S_k] \end{aligned} \tag{5}$$

where S_k is the set of all k-sparse unit vectors with non-zero entries equal to $\pm\frac{1}{\sqrt{k}}$. The recovery task is to estimate $\mathrm{supp}(v)$ given observations $X_1, X_2, \ldots, X_n$ sampled from $N\left(0, I_d + \theta v v^\top\right)^{\otimes n}$. Here we study a biased variant of the problem. Namely, when v comes from the further restricted family (let $\|v\|_0^+$ denotes the number of positive entries of v):

$$BS_k := \{v \in S_k : \left| \|v\|_0^+ - \frac{k}{2} \right| > \delta \cdot k\}$$

where δ is a constant. Specifically for this variant where the sum test can be shown optimal for detection, our result implies a detection-recovery gap which is lacking in its general unbiased form.

5.3 Open Problems

We point out several open problems and conjectures related to our work. Specifically, the following questions of interest are unaddressed:

1. To resolve the PDS recovery conjecture with reduction from the PC conjecture.

2. To find a "quiet" null H_0 where there is no dense subgraph (satisfying the *val* condition for DkS) but hard to distinguish with PDS up to $\frac{k^2}{n} \cdot \frac{(p-q)^2}{q(1-q)} \in \widetilde{O}(1)$. This would also imply a *detection-certificate gap* as well as the recovery hardness conjecture.

3. Whether the reverse direction for Theorem 5.1 can be proven. Specifically, is there a direct reduction losing only poly-log signal strength from binomial planted set to fixed planted set.

4. Whether one can show that statistical lower bounds for PDS* and ISBM are tight. Specifically, since there are no known polynomial algorithms, lower bound $d_{\mathrm{TV}}(H_0, H_1)$ at the regime where recovery is statistically impossible ($D_{KL} \in \widetilde{\Theta}(n^2/k^4)$).

References

[Alo+07] Noga Alon et al. "Testing K-Wise and Almost k-Wise Independence". In: *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*. STOC '07. San Diego, California, USA: Association for Computing Machinery, 2007, pp. 496–505. ISBN: 9781595936318. DOI: 10.1145/1250790.1250863. URL: https://doi.org/10.1145/1250790.1250863.

[AS16] Venkat Anantharam and Justin Salez. "The densest subgraph problem in sparse random graphs". In: *The Annals of Applied Probability* 26.1 (2016), pp. 305–327. DOI: 10.1214/14-AAP1091. URL: https://doi.org/10.1214/14-AAP1091.

[AWZ20] Gérard Ben Arous, Alexander S. Wein, and Ilias Zadik. *Free Energy Wells and Overlap Gap Property in Sparse PCA*. 2020. DOI: 10.48550/ARXIV.2006.10689. URL: https://arxiv.org/abs/2006.10689.

[Bal+19] Paul Balister et al. "Dense subgraphs in random graphs". In: *Discrete Applied Mathematics* 260 (2019), pp. 66–74. ISSN: 0166-218X. DOI: https://doi.org/10.1016/j.dam.2019.01.032. URL: https://www.sciencedirect.com/science/article/pii/S0166218X19300678.

[Ban+20] Afonso S. Bandeira et al. "Spectral Planting and the Hardness of Refuting Cuts, Colorability, and Communities in Random Graphs". In: *CoRR* abs/2008.12237 (2020). arXiv: 2008.12237. URL: https://arxiv.org/abs/2008.12237.

[Bar+19] Boaz Barak et al. "A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem". In: *SIAM Journal on Computing* 48.2 (2019), pp. 687–735. DOI: 10.1137/17M1138236. eprint: https://doi.org/10.1137/17M1138236. URL: https://doi.org/10.1137/17M1138236.

[BB19a] Matthew Brennan and Guy Bresler. *Average-Case Lower Bounds for Learning Sparse Mixtures, Robust Estimation and Semirandom Adversaries*. 2019. DOI: 10.48550/ARXIV.1908.06130. URL: https://arxiv.org/abs/1908.06130.

[BB19b] Matthew Brennan and Guy Bresler. *Optimal Average-Case Reductions to Sparse PCA: From Weak Assumptions to Strong Hardness*. 2019. DOI: 10.48550/ARXIV.1902.07380. URL: https://arxiv.org/abs/1902.07380.

[BB20] Matthew Brennan and Guy Bresler. *Reducibility and Statistical-Computational Gaps from Secret Leakage*. 2020. DOI: 10.48550/arxiv.2005.08099. URL: https://arxiv.org/abs/2005.08099.

[BBH18] Matthew Brennan, Guy Bresler, and Wasim Huleihel. *Reducibility and Computational Lower Bounds for Problems with Planted Sparse Structure*. 2018. DOI: 10.48550/arxiv.1806.07508. URL: https://arxiv.org/abs/1806.07508.

[BBH19] Matthew Brennan, Guy Bresler, and Wasim Huleihel. *Universality of Computational Lower Bounds for Submatrix Detection*. 2019. DOI: 10.48550/ARXIV.1902.06916. URL: https://arxiv.org/abs/1902.06916.

[BH16] Afonso S. Bandeira and Ramon van Handel. "Sharp nonasymptotic bounds on the norm of random matrices with independent entries". In: *The Annals of Probability* 44.4 (July 2016). DOI: 10.1214/15-aop1025. URL: https://doi.org/10.1214%2F15-aop1025.

[BI13] Cristina Butucea and Yuri I. Ingster. "Detection of a sparse submatrix of a high-dimensional noisy matrix". In: *Bernoulli* 19.5B (2013), pp. 2652–2688. DOI: 10.3150/12-BEJ470. URL: https://doi.org/10.3150/12-BEJ470.

[BR13] Quentin Berthet and Philippe Rigollet. "Complexity Theoretic Lower Bounds for Sparse Principal Component Detection". In: *Proceedings of the 26th Annual Conference on Learning Theory*. Ed. by Shai Shalev-Shwartz and Ingo Steinwart. Vol. 30. Proceedings of Machine Learning Research. Princeton, NJ, USA: PMLR, Dec. 2013, pp. 1046–1066. URL: https://proceedings.mlr.press/v30/Berthet13.html.

[BS87] Andrei Broder and Eli Shamir. "On the second eigenvalue of random regular graphs". In: *28th Annual Symposium on Foundations of Computer Science (sfcs 1987)*. 1987, pp. 286–294. DOI: 10.1109/SFCS.1987.45.

[CLR17] T. Tony Cai, Tengyuan Liang, and Alexander Rakhlin. "Computational and statistical boundaries for submatrix localization in a large noisy matrix". In: *The Annals of Statistics* 45.4 (Aug. 2017). DOI: 10.1214/16-aos1488. URL: https://doi.org/10.1214%2F16-aos1488.

[Coo+17] Colin Cooper et al. *The flip Markov chain for connected regular graphs*. 2017. DOI: 10.48550/ARXIV.1701.03856. URL: https://arxiv.org/abs/1701.03856.

[CX14] Yudong Chen and Jiaming Xu. *Statistical-Computational Tradeoffs in Planted Problems and Submatrix Localization with a Growing Number of Clusters and Submatrices*. 2014. DOI: 10.48550/ARXIV.1402.1267. URL: https://arxiv.org/abs/1402.1267.

[Dec+11] Aurelien Decelle et al. "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications". In: *Physical Review E* 84.6 (Dec. 2011). DOI: 10.1103/physreve.84.066106. URL: https://doi.org/10.1103%2Fphysreve.84.066106.

[Gam21] David Gamarnik. "The Overlap Gap Property: a Geometric Barrier to Optimizing over Random Structures". In: *CoRR* abs/2109.14409 (2021). arXiv: 2109.14409. URL: https://arxiv.org/abs/2109.14409.

[GMZ14] Chao Gao, Zongming Ma, and Harrison H. Zhou. *Sparse CCA: Adaptive Estimation and Computational Barriers*. 2014. DOI: 10.48550/ARXIV.1409.8565. URL: https://arxiv.org/abs/1409.8565.

[Gre11] Catherine Greenhill. *A polynomial bound on the mixing time of a Markov chain for sampling regular directed graphs*. 2011. DOI: 10.48550/ARXIV.1105.0457. URL: https://arxiv.org/abs/1105.0457.

[HWX15a] Bruce Hajek, Yihong Wu, and Jiaming Xu. "Computational Lower Bounds for Community Detection on Random Graphs". In: *Proceedings of The 28th Conference on Learning Theory*. Ed. by Peter Grünwald, Elad Hazan, and Satyen Kale. Vol. 40. Proceedings of Machine Learning Research. Paris, France: PMLR, Mar. 2015, pp. 899–928. URL: https://proceedings.mlr.press/v40/Hajek15.html.

[HWX15b] Bruce Hajek, Yihong Wu, and Jiaming Xu. *Information Limits for Recovering a Hidden Community*. 2015. DOI: 10.48550/ARXIV.1509.07859. URL: https://arxiv.org/abs/1509.07859.

[HWX16] Bruce Hajek, Yihong Wu, and Jiaming Xu. *Semidefinite Programs for Exact Recovery of a Hidden Community*. 2016. DOI: 10.48550/ARXIV.1602.06410. URL: https://arxiv.org/abs/1602.06410.

[Jon+21] Chris Jones et al. “Sum-of-Squares Lower Bounds for Sparse Independent Set”. In: *CoRR* abs/2111.09250 (2021). arXiv: 2111.09250. URL: https://arxiv.org/abs/2111.09250.

[LLV15] Can M. Le, Elizaveta Levina, and Roman Vershynin. *Concentration and regularization of random graphs*. 2015. DOI: 10.48550/ARXIV.1506.00669. URL: https://arxiv.org/abs/1506.00669.

[MW15] Zongming Ma and Yihong Wu. “Computational barriers in minimax submatrix detection”. In: *The Annals of Statistics* 43.3 (2015), pp. 1089–1116. ISSN: 00905364. URL: http://www.jstor.org/stable/43556548 (visited on 11/14/2022).

[Rus+22] Cynthia Rush et al. *Is it easier to count communities than find them?* 2022. DOI: 10.48550/ARXIV.2212.10872. URL: https://arxiv.org/abs/2212.10872.

[SW22] Tselil Schramm and Alexander S. Wein. “Computational barriers to estimation from low-degree polynomials”. In: *The Annals of Statistics* 50.3 (2022), pp. 1833–1858. DOI: 10.1214/22-AOS2179. URL: https://doi.org/10.1214/22-AOS2179.

[TY19] Konstantin Tikhomirov and Pierre Youssef. “The spectral gap of dense random regular graphs”. In: *The Annals of Probability* 47.1 (2019), pp. 362–419. DOI: 10.1214/18-AOP1263. URL: https://doi.org/10.1214/18-AOP1263.

[Ver+12] Alexander Veremyev et al. “Dense Percolation in Large-Scale Mean-Field Random Networks Is Provably “Explosive””. In: *PloS one* 7 (Dec. 2012), e51883. DOI: 10.1371/journal.pone.0051883.

[WBS16] Tengyao Wang, Quentin Berthet, and Richard J. Samworth. “Statistical and computational trade-offs in estimation of sparse principal components”. In: *The Annals of Statistics* 44.5 (2016), pp. 1896–1930. DOI: 10.1214/15-AOS1369. URL: https://doi.org/10.1214/15-AOS1369.

6 Appendix:

6.1 Preliminaries and notations

We briefly introduce the notations. We use $\mathcal{L}(X)$ to denote the law of a random variation X, $d_{\mathrm{TV}}, D_{KL}, \chi^2$ to denote the total variation distance, KL-divergence, and χ^2 divergence. Specifically we shorthand $d(\mathrm{Bern}(p), \mathrm{Bern}(q)) := d(p, q)$ for Bernoullis with bias p, q. We use the $\widetilde{O}(\cdot)$ notation to denote big-O ignoring log-factors. For instance, $r \in \widetilde{\omega}(n)$ means $r \in \omega(n \log^k n)$ for any constant k, $r \in \widetilde{\Omega}(n)$ means $r \in \Omega(n \log^k n)$ for some k, and $\widetilde{O}$, $\widetilde{o}$ likewise. Specifically, $\widetilde{\Theta}(n) = \widetilde{O}(n) \bigcap \widetilde{\Omega}(n)$. We use $\prod_i P_i$ to denote the tensor product of distributions, specifically $P^{\otimes k} = \prod_{i=1}^k P$.

For a given partition F of $[n]$ to k sets, we use $\mathcal{U}_n(F)$ to denote the uniform distribution of k-subsets of $[n]$ with each element in one of F_i. We let $Unif_n(k)$ to denote the uniform distribution over all k-subsets of $[n]$. For a planted structure distribution, we use $\mathcal{M}_{A\times B}(S \times T, P, Q)$ to denote planted structure on community $A \times B$ with a planted submatrix $S \times T$ where the in-community entries sampled from P and otherwise from Q. Specifically, if $A = B$ and $S = T$ are unknown sampled from $\mathcal{P}$, denote $\mathcal{M}_{A\times B}(\mathcal{P}, P, Q) := \mathbb{E}_{S\sim\mathcal{P}} \left(\mathcal{M}_{A\times B}(S \times S, P, Q)\right)$ the symmetric planting.

We use $A \otimes B \in \mathbb{R}^{n^2\times n^2}$ for matrices $(A, B) \in (\mathbb{R}^{n\times n}, \mathbb{R}^{n\times n})$ to denote the Kronecker product between A, B. We usually parameterize indices of $A \otimes B$ by a pair $(ij) : i, j \in [n]$ such that $(A \otimes B)_{(ij),(kl)} = A_{ik}B_{jl}$. Fixing i, j and laying out the row of $A \otimes B$ as a $n \times n$ matrix, it is exactly $A^T_{i,\cdot} B_{j,\cdot}$ the product of two row-vectors.

We then introduce the following (common) lemmas as preliminaries. Let f be a Markov transition kernel and P be any distribution, we denote the law of $\mathcal{L}(f(A))_{A\sim P} =: f(P)$. We also denote sets in $V \in 2^{[n]}$ and vectors $v \in \{0, 1\}^n$ interchangeably, and $\mathrm{PDS}(n, S, p, q)$ to be the planted dense subgraph instance conditioned on planted location at set S.

Lemma 6 (Data Processing Inequality)**.** *Let $\mathcal{A}$ be a Markov transition kernel and A, B be two distributions, then:*

$$d_{\mathrm{TV}}(f(A), f(B)) \leq d_{\mathrm{TV}}(A, B).$$

Lemma 7 (Basic geometry on TV)**.** *The following holds:*

1. *For any three distributions P, Q, R we have*

$$d_{\mathrm{TV}}(P, R) \leq d_{\mathrm{TV}}(P, Q) + d_{\mathrm{TV}}(Q, R).$$

2. *Assuming a latent variable $\theta \in \Theta$ and a class of distributions parameterized by θ satisfies $d_{\mathrm{TV}}(P_\theta, P) > \epsilon$ for any $\theta \in \Theta$, then the mixture:*

$$d_{\mathrm{TV}}(\mathbb{E}_\theta P_\theta, P) > \epsilon.$$

Lemma 8 (Tensorization of TV)**.** *Let P_i, Q_i be distributions for $i = 1, 2, \ldots, n$. Then:*

$$d_{\mathrm{TV}}(\prod P_i, \prod Q_i) \leq \sum_i d_{\mathrm{TV}}(P_i, Q_i).$$

Lemma 9 (Accumulation of TV distance)**.** *Consider a finite set of sequential functions on distributions $\mathcal{A}_i : i = 1, 2, \ldots, k$. Assuming one has distributions $P_0, P_1, P_2, \ldots, P_k$ such that:*

$$d_{\mathrm{TV}}(\mathcal{A}_i(P_{i-1}), P_i) \leq \epsilon_i$$

for all $i = 1, 2, \dots, k$, then we have:

$$d_{\mathrm{TV}}(\mathcal{A}_k(\dots \mathcal{A}_1(\mathcal{A}_1(P_0))\dots), P_k) \le \sum_i \epsilon_i.$$

The last lemma comes directly from data processing and induction. Next, we present a couple of lemmas on Bernoulli distributions.

Lemma 10 (KL divergence between Bernoullis)**.** *Assume a sequence of $\{p_n\}$ and $\{q_n\}$ such that $q_n < p_n < cq_n$, $1 - q_n < c(1 - p_n)$ for some constant c, then:*

$$D_{KL}(p\|q) := D_{KL}(\mathrm{Bern}(p)\|\mathrm{Bern}(q)) = \Theta(\frac{(p-q)^2}{q(1-q)}).$$

Proof. Note that the quantity $\frac{(p-q)^2}{q(1-q)}$ is the χ^2 divergence between two Bernoulli, which dominates the KL divergence. For the other side, note that on the support of these two distributions ($\{0, 1\}$) their ratio of density is bounded. Thus by a reverse Pinsker's inequality the result follows. □

Lemma 11 (TV divergence between Binomials)**.** *Consider two parameters $p, q \in (0, 1)$, then:*

$$d_{\mathrm{TV}}(\mathrm{Bern}(q)^{\otimes n}, \mathrm{Bern}(q)^{\otimes n}) \le \sqrt{\frac{n(p-q)^2}{2q(1-q)}}.$$

Proof. This comes directly from the Pinsker's inequality on TV, KL, and χ^2 divergences:

$$\begin{aligned} d_{\mathrm{TV}}(\mathrm{Bern}(q)^{\otimes n}, \mathrm{Bern}(q)^{\otimes n}) &\le \sqrt{\frac{D_{KL}(\mathrm{Bern}(q)^{\otimes n}, \mathrm{Bern}(q)^{\otimes n})}{2}} \\ &= \sqrt{\frac{nD_{KL}(\mathrm{Bern}(p), \mathrm{Bern}(q))}{2}} \\ &\le \sqrt{\frac{n\chi^2(\mathrm{Bern}(p), \mathrm{Bern}(q))}{2}} = \sqrt{\frac{n(p-q)^2}{2q(1-q)}} \end{aligned}$$

due to $2d_{\mathrm{TV}}^2 \le D_{KL} \le \chi^2$ and the factorization of D_{KL} for independent distributions. □

Finally, a result on the Kronecker product for the operator norm:

Lemma 12 (Kronecker product on operator norm)**.** *For any two matrices A, B, we have the following equality on operator norm (largest singular value):*

$$\|A \otimes B\|_{op} = \|A\|_{op} \cdot \|B\|_{op}$$

where $\otimes$ is the Kronecker product.

6.2 Reductions to detection

In this section we point out that all of the inference variants considered, detection is (almost) the weakest version of all. This can be viewed from the perspective of reductions where a blackbox for a different task implies a blackbox for detection. Such reductions were discussed in [HWX15a][Bar+19][Ban+20]. Here we re-formulate the necessary proofs:

Lemma 13 (Refutation implies detection)**.** *For two hypothesis H_0, H_1 and valuation function val with separation thresholds ϵ and gap δ. If there is an efficient refutation blackbox A with asymptotic success probability $p = \lim_{n\to\infty} p^{(n)} > 0$, then (weak) detection is computationally possible for H_0 and H_1.*

Proof. Consider the canonical form of refutation as described in Section 1.2 with a polynomial-timed refutation blackbox A having success probability $p > 0$. We show that A satisfies the detection criteria.

- When $G \sim H_0$, $A(G) = 0$ with probability at least $p \cdot P(val(G) < \epsilon - \delta | G \sim H_0)$, thus the Type I error is at most $1 - p \cdot P(val(G) < \epsilon - \delta | G \sim H_0)$.
- When $val(G) > \epsilon + \delta$ the output is always 1, thus the Type II error is at most the probability of a low valuation $P(val(G) > \epsilon + \delta | G \sim H_1)$.

Therefore, the sum of errors is bounded above by:

$$1 - p \cdot P(val(G) < \epsilon - \delta | G \sim H_0) + P(val(G) > \epsilon + \delta | G \sim H_1)$$

Note that as $n \to \infty$,

$$P(val(G) < \epsilon - \delta | G \sim H_0) \to 1, \quad P(val(G) > \epsilon + \delta | G \sim H_1) \to 0.$$

Therefore, the detection error of this blackbox is bounded above by $1 - p^{(n)} < 1$ in the limit. This implies that it returns a better-than-random detection asymptotically. □

Lemma 14 (Recovery implies detection)**.** *For two hypothesis H_0, H_1 with valuation val. Suppose there is a secret key k_G such that an approximation $val_k(G) \leq val(G)$ can be computed in polynomial time and the alternate distribution $G \sim H_1$ satisfies $val_k(G) > \epsilon$ with probability $1 - o_n(1)$, then detection is possible in polynomial time equipped with a polynomial oracle that generates k_G for a given graph (assuming $val(G \sim H_0) \leq \epsilon - \delta$ with high probability).*

Proof. Consider the alternate valuation function $val' = val_k$, which can be computed in polynomial time by first computing k. From the previous lemma and the separation conditions we know that:

1. For $G \sim H_0$, $val'(G) \leq val(G) \leq \epsilon - \delta$ with high probability.
2. For $G \sim H_1$, $val'(G) > \epsilon$ with high probability.

Therefore, val' is a polynomial time valuation, which obviously implies that refutation on this blackbox can be done in polynomial time. By Lemma 13, we have the desired conclusion. □

6.3 Different varieties of recovery

Consider the notion of *minimal recovery* of strength $\alpha > 0$, which is outputting a guess $\widehat{P}$ for the planted location such that

$$\lim_{n\to\infty} \frac{(\log k)^{\alpha} \mathbb{E}[|\widehat{P} \bigcap P|]}{k} \geq 1.$$

Specifically weak recovery is just partial recovery of strength 1 and partial recovery implies minimal recovery of strength $\alpha \to 0$. We can go on to prove that with partial recovery one can achieve precise recovery with only sub-polynomial signal boost. This means that the PDS recovery conjecture can be weakened to only assume hardness for *minimal* recovery. The following lemma applies to both the statistical and computational boundary.

Lemma 15 (Minimal recovery implies exact recovery)**.** *For the* PDS(n,k,p,q) *recovery problem when p,q are bounded by away by zero and one. If one can achieve minimal recovery with strength $\alpha > 1$ on a sequence of parameters (N_n, K_n, P_n, Q_n) in polynomial time where $K_n \in \widetilde{\omega}(\sqrt{N_n})\bigcap o(N_n^{\gamma})$ for some exponent $\gamma \in (\frac{1}{2}, 1)$, then one can achieve exact recovery on a modified sequence of parameters (N_n, K_n, P_n, Q_n') where Q_n' satisfies $D_{KL}(P_n \| Q_n) = \Theta((\log k)^{2\alpha} D_{KL}(P_n \| Q_n'))$.*

Proof. The critical component here lies in the subroutine Graph-Clone (Lemma 2) in which we generate independent graph instances conditioned on planted instance locations. The lemma can be read off as the following form:

- Suppose we have a hidden planted location η, and a one-time sampler from the planted distribution $\mathcal{M}_{[n]\times[n]}(\eta \times \eta, \mathrm{Bern}(p), \mathrm{Bern}(q))$, then we have a one-time sampler from the tensor product $\mathcal{M}^{\otimes 2}_{[n]\times[n]}(\eta\times\eta, \mathrm{Bern}(p), \mathrm{Bern}(Q))$ where $Q = 1 - \sqrt{(1-p)(1-q)}$. Specifically, the divergence measure $\chi^2(p, Q) > \frac{1}{2}\chi^2(p,q)$.

Corollary: Suppose we have a hidden planted location η and as above a one-time sampler, then we can have a sample generated from $\mathcal{M}^{\otimes 2}_{[n]\times[n]}(\eta\times\eta, \mathrm{Bern}(p), \mathrm{Bern}(Q))$ where $\chi^2(p,Q) > \frac{1}{2t}\chi^2(p,q)$.

Note that in the case when p,q are bounded away from zero and one, $D_{KL}(p\|q) \sim \chi^2(p,q) \sim (p-q)^2$ are of the same order. Therefore, with the cost of reducing $(\log k)^{2\alpha}$ in the distance we can generate one instance from $\mathcal{M}^{\otimes(\log k)^{2\alpha}}_{[N]\times[N]}(\eta\times\eta, \mathrm{Bern}(P), \mathrm{Bern}(Q))$ from a single instance of the original $\mathcal{M}_{[N]\times[N]}(\eta\times\eta, \mathrm{Bern}(P), \mathrm{Bern}(Q'))$.

Note that our assumption states that we have a black box to perform minimal recovery of strength α on $\mathcal{M}_{[N]\times[N]}(\eta\times\eta, \mathrm{Bern}(P), \mathrm{Bern}(Q))$, and we arrive at $\log^{2\alpha} k$ estimates of the planted η, denoted by $\widehat{\eta}_1, \ldots, \widehat{\eta}_{\log^{2\alpha} k} := \widehat{\eta}_r$. Moreover, given that the cloned copies are *independently generated* conditioned on η and hence so are those $\widehat{\eta}_i$'s, we wish to reconstruct η through those independent estimate $\widehat{\eta}_i$'s.

Note that it is safe to assume that any black-box takes in the input unlabeled, because we can apply a hidden permutation π to the graph and feed it to the black-box instead, we know that for each index $i \in \eta$, the probabilty that it lies in a $\widehat{\eta}$ is exactly $E(\widehat{\eta}\cdot\eta)/k \sim \log^{\alpha} k$ where expectation ranges over $\mathcal{M}_{[N]\times[N]}(\eta\times\eta, \mathrm{Bern}(P), \mathrm{Bern}(Q))$. For each node not in η, the probability of $\widehat{\eta}$ hitting it is at most $\frac{k}{n-k} \le \frac{2k}{n}$. Therefore, if we compile a histogram of $\widehat{\eta}_i$ hits, we have k copies of $\mathrm{Binom}(r, \sqrt{r^{-1}})$ and $n-k$ copies of (at most) $\mathrm{Binom}(r, \frac{2k}{n})$.

We now only need to show that when $r \in \Theta((\log k)^{2\alpha})$, the two distributions (smallest from η and largest from $\overline{\eta}$) separates with probability $\to 1$. Note that the probability that $\mathrm{Binom}(r, 2k/n)$ is at least constant C is:

$$\mathbb{P}(\mathrm{Binom}(r, 2k/n) > C) \le r \cdot r^C \cdot (\frac{2k}{n})^C \lesssim (\log n)^{2(C+1)\alpha} n^{-C(1-\gamma)}$$

Pick any $C(1-\gamma) > 1$, then the above probability goes to $\widetilde{o}(n^{-1})$, and a union bound over all vertices in $\overline{\eta}$ says that with probability $1 - o_n(1)$ all the counts in that group is bounded by constant C.

Now we consider the group of nodes in η. The probability of one being bouneded by C is at most

$$\mathbb{P}(\mathrm{Binom}(r, \sqrt{r^{-1}}) < C) \le C\binom{r}{C}(1-\sqrt{r^{-1}})^{r-C} \le Cr^C \exp\left(-(r-C)\sqrt{r^{-1}}\right)$$

because $1 - x < e^{-x}$, and the union bound says that the minimum for counts in η is at most:

$$k\mathbb{P}(\mathrm{Binom}(r, \sqrt{r^{-1}}) < C) \lesssim kr^C \exp\left(-\frac{1}{2}\sqrt{r}\right) = \exp\left(\log k - \frac{(\log k)^{\alpha}}{2} + O(\log r)\right)$$

assuming that $\alpha > 1$, the above goes to 0 as $k \to \infty$.

Therefore, if we sum over statistics for $\widehat{\eta}_i$'s we get that the entries in η goes above any constant with probability $\to 1$ whereas with high probability the other entries are bounded by a constant, and hence precise recovery is achievable via the most popular nodes. Moreover, this entire procedure applies in polynomial time. □

We further comment that the case of when p, q are not bounded away by one ("dense") are similar, but require some further bounds on the exponents. Here we only present proof of this dense regime. Moreover, the condition that $k < n^\gamma$ instead of $k \in o(n)$ is also only needed for minimal recovery (not required for a reduction from partial to exact), but for our purposes this is a fine assumption to make. The condition $\alpha > 1$ is to some extent unnecessary either because recovery of α implies recovery of α^+ for any $\alpha^+ > \alpha$ ($\alpha > 1$ is only needed for convenience in expressing signal decay).

Moreover, note that the non-homogenity of planted instance in Theorem 6.2 means that the direct reduction in Lemma 15 does not apply. In fact, it will be easy to see that (hardness in) weak recovery for our instance in Theorem 6.2 is can be implied by PDS recovery.

For completeness of arguments (which will be useful in statistical bound for refutation), we also present a statistical condition for (exact) PDS recovery. Below is a sufficient result from [HWX15b]:

Theorem 6.1 (PDS recovery – Theorem 2 in [HWX15b])**.** *Again consider the settings (and parameters correspondence) as above, then exact recovery is statistically possible if:*

$$\frac{kD_{KL}(p\|q)}{\log n} > C$$

and impossible if

$$\frac{kD_{KL}(p\|q)}{\log n} < c$$

for some absolute constants c, C.

6.3.1 Evidence of recovery hardness

We present a statement on tight recovery hardness but with a relaxed community structure.

Theorem 6.2. *For any $k_n \in \omega(\sqrt{n}), p_n \in (0,1)$, there exists a symmetric edge density matrix on subgraphs $D_n \in \mathbb{R}^{k_n \times k_n}$, such that the row (and column) sums of D are uniformly $k_n \cdot \lambda_n$ and the graph constructed by:*

1. *On $G \sim G(n, p_n)$, randomly select a subset S of vertices of size k_n. Choose a random bijection π from S to $[k_n]$.*
2. *For the nodes $u, v \in S$, resample uv with probability $D_{\pi(u)\pi(v)} + p$.*

And if $\limsup_{n\to\infty} \frac{k_n^2}{n} \frac{\lambda_n^2}{p_n(1-p_n)} \in \widetilde{o}(1)$, no (randomized) polynomial algorithm can achieve exact recovery on the planted instance, even given the knowledge of D, assuming the planted clique conjecture with $p = 1/2$.

It is not hard to check that for this general model (where recovery is at least as hard as PDS), one can still find a log-optimal algorithmic matching upper bound. Specifically, consider simply taking the k most popular nodes (those with the highest degrees), then as long as the ratio $\frac{k^2\lambda^2}{np(1-p)\log n} \to \infty$, the output satisfies *exact recovery* criteria.

The proof comes from a very basic intuition: the recovery conjecture is essentially the bound where the *average* elevated degree is $\widetilde{\Theta}(\sqrt{n})$ for each planted node, which happens to be the same thing for the original PC conjecture. Therefore, we can imagine a very simple reduction where we *plant* an extra dense subgraph in addition to PC, so that we have a (slightly) bigger planted dense subgraph, but the average elevated degree remains the same. However, strong recovery here will not be possible because even if we have the knowledge of the extra planted dense subgraph, we still cannot tell the original location of PC from Conjecture 1.1.

Proof. We start from the fact that, *recovery for* PC *is hard at the regime when detection is hard*, which is a direct implication of the PC conjecture and Lemma 14. Consider the following procedure applied on a graph $G \sim \mathrm{PDS}(n, k, 1, 1/2)$ to obtain G':

1. Add $(t-1)k$ vertices to G that will be part of the (new) planted structure where $t > 1$ is a specified parameter such that the total planted size is tk.

2. For the $(t-1)k$ extra vertices, connect each pair with probability $\frac{t}{2(t-1)}$. Connect each edge between the original n vertices to the new $(t-1)k$ vertices with probability $1/2$.

3. Permute the nodes in G' randomly.

Under this reduction, consider the new planted density matrix in $\mathbb{R}^{tk \times tk}$ where a $\mathbb{R}^{(t-1)k \times (t-1)k}$ principal submatrix is $\frac{1}{2(t-1)} J_{(t-1)k}$ and the other $k \times k$ principal submatrix has all entries $1/2$. The recovery hardness comes from the fact that even if the blackbox knows the exact location where our planted $(t-1)k$ nodes are, it can still not precisely recover the original k vertices in the planted clique instance with high probability, thus strong recovery is impossible.

Note that this satisfies the row-column sum constraint where the expected degree of each planted node is exactly $\frac{n+(t+1)k}{2}$ and the planted structure lifted each node's degree by $k/2$. Consider the distribution of the degrees for a node not in planted structure (which is $d \sim \mathrm{Binom}(n+(t-1)k, 1/2)$) and the node inside planted structure which is either $d \sim k + \mathrm{Binom}(n + (t-2)k, 1/2)$ or $d \sim \mathrm{Binom}(n, 1/2) + \mathrm{Binom}((t-1)k, 2t/(t-1))$ depending on which part of the planted set. The separation of the first distribution (null) with the later two (latent) follows immediately by a very simple Chernoff Bound when $k \in \widetilde{\omega}(\sqrt{n})$. □

6.4 Proofs in section 3

6.4.1 Criteria of a "Good" Design Matrix

In this section we discuss how to go from a design matrix to the desired quiet distribution. Note that to use thresholding from Gaussian in our last step of reduction, our target Gaussian mean should be approximately the bias of desired planted Bernoulli (subtract $1/2$). Therefore we should expect the matrix itself to have some "planted property", although it is by no mean random itself during Bern-rotation process.

Specifically, for the purpose of design matrices in the problems here, we want to construct matrices K that have the following properties ($r = n/k$):

1. It should have top singular value (operator norm) at most 1.

2. It should have a *planted submatrix* structure e.g. where a $1/r \times 1/r$ submatrix of entries having an elevated mean corresponding to the planted dense subgraph.

3. Finally, we want the design matrix to be map to the objective in the regime that the computation barrier is still (roughly) preserved. It cannot be too lossy in transforming the signals.

Note that the row vectors have norm 1 in the design matrix construction in Lemma 4, hence it is equivalent to proving that it is (roughly) an isometry. Therefore, a (tight) spectral bound on the matrix is needed when our target only demands row-regularity, which is why we consider directed regular graphs (on which we can prove spectral boundaries).

It is worth noting that a degree-2 test is (roughly) equivalent to our spectral rotation, suggesting some "tightness" on Bernoulli Rotations. For instance, the degree-2 test on ISBM_D would be to sum up the square of all (Gaussian) entries, and on PDS^*_D would be to sum up the degrees. Informally, for a target mean $\lambda\mu_i$ with signal λ, if for any matrix A mapping planted bits to (a permutation) of $\lambda\mu_i$, there exist a fixed vector v such that $\|Av\| > \|v\|$, then outputting the norm of $v \cdot \lambda\mu_i$ itself is a valid degree-2 detection algorithm.

However, as raised in [SW22], there may very well be Gaussian planted models where the degree-2 testing power is sub-optimal. It remains open to understand whether Bern-Rotation or any other method can be used to provide a reduction to such lower bounds.

6.4.2 Proof of Lemma 4

Proof. Note that the adjacency matrix of the sampled directed graph A is not symmetric. However, we do know that the operator norm equals to the largest singular value of

$$(A - \frac{d}{n}\mathbb{1}\mathbb{1}^T)(A^T - \frac{d}{n}\mathbb{1}\mathbb{1}^T) = AA^T - \frac{d^2}{n}\mathbb{1}\mathbb{1}^T$$

where $d = n/r$ and $\mathbb{1} \in \mathbb{R}^{n\times 1}$ is the all-one vector.

We know that the largest eigenvalue of AA^T is d^2/n corresponding to the all one vector because it is a scaled doubly stochastic matrix. Therefore, from the Courant-Fischer Theorem we can show that the second largest singular value of A which is the second largest eigenvalue of AA^T is the largest eigenvalue of $AA^T - \frac{d^2}{n}\mathbb{1}\mathbb{1}^T$, which is the largest singular value of $A - \frac{d}{n}\mathbb{1}\mathbb{1}^T$.

Note that Theorem.B of [TY19] asserts that under the conditions in the lemma, the said quantity is bounded by $C\sqrt{d}$ with probability $1 - o_n(1)$. Therefore, our constructed R, which is exactly $\sqrt{\frac{r}{n}}(A - \mathbb{A}) = \sqrt{d^{-1}}(A - \frac{d}{n}\mathbb{1}\mathbb{1}^T)$ has max singular value at most C with probability $1 - o_n(1)$. □

6.4.3 Proof of Lemma 5

Proof. Take R from the previous lemma. Note that the top singular value of a matrix is in fact sub-additive, and the Kronecker product is a linear operator that preserves the product of the operator norm of a matrix. We have:

$$\sigma(\mu^{-1}\sqrt{\frac{n}{r}}K) \le \sigma(R \otimes R) + \frac{2}{\sqrt{nr}}\sigma(R \otimes J) \le C^2 + \frac{2C}{\sqrt{nr}}\sigma(J) = C^2 + 2C\sqrt{\frac{n}{r}}$$

because the top eigenvalue of $J = \mathbb{1}\mathbb{1}^T$ is exactly n. Therefore $\sigma(K) \le 1$ for any R satisfying the criteria of Lemma 4.

Finally, note Theorem 1 in [Gre11] states that the switch Markov Chain, on which the unique stationary distribution is the uniform distribution over all directed $d-$regular graphs, is fast-mixing, and Lemma 4 still holds if the sampling condition is approximate in $L1$. Therefore, in polynomial time we can find one candidate K satisfying our lemma above. Moreover, note that if we sample $O(n)$ times indepdently, the probability of failure becomes exponentially small. □

6.5 Proofs in section 4

As a starter to detection problems in PDS*, we present a simple PDS$^*_D$ upper bound by a degree 2 polynomial test extending Proposition B.4 in [SW22].

Lemma 16 (Upper bound on PDS$^*_D$)**.** *Consider the degree corrected* PDS* *model with planted subgraph size k and average edge probability $0 < q < p < \frac{2}{3}$, then as long as the product ratio $\frac{k^3}{n^{1.5}} \cdot \frac{(p-q)^2}{q(1-q)} \in \omega_n(1)$, one can computationally efficiently resolve hypothesis testing for* PDS**.*

The proof goes by considering the statistics $f = \sum d_i^2$ where d_i are the degrees of G and computing the mean different over variance. The upper bound gives a boundary strictly between the sum-test level for PDS and the spectral recovery level (Kesten-Stigum threshold), suggesting some consideration into the "community" structures compare to a vanilla sum test.

In the following proof we consider PDS* in the form of

$$H_0 : G \sim \text{Bern}(p_0)^{\otimes n \times n} \quad \text{and} \quad H_1 : G \sim \mathcal{M}_{[n]\times[n]}(S \times S, \text{Bern}(p_1), \text{Bern}(q_1))$$

where $k^2 p_1 + (n^2 - k^2) q_1 = n^2 p_0$, $S \sim Unif_n(k)$. Moreover, assume that $k \in o(n) \bigcap \omega(\sqrt{n})$, $p_1 - p_0 = \gamma > 0$, $p_1(1-p_1), q_1(1-q_1) \in \Theta\left(p_0(1-p_0)\right)$. The case of a symmetric graph (instead of bi-partite) and binomially-planted graph follows (almost) exactly the same. Let v denote the planted set.

Proof of Lemma 16. Consider the test statistics $f(G) = \sum d_i^2$ where d_i are the (independent) degrees. We show that there exist τ such that $\mathbb{P}_{H_0}(f(G) > \tau) + \mathbb{P}_{H_1}(f(G) < \tau) \to 0$ as $n \to \infty$.

Firstly, consider what happens to the degrees under H_0: they are n independent samples from $\text{Binom}(n, p_0)$ with expectation given by

$$\mathbb{E}(f) = n \cdot \mathbb{E}_{x \sim \text{Binom}(n,p_0)} x^2 = n^2 p_0(1-p_0) + n^3 p_0^2 = n^2 p_0 + (n^3 - n^2) p_0^2$$

Similarly in H_1, there are $n - k$ nodes that are not in the planted set and their corresponding second moment of degree is:

$$\mathbb{E}(\sum_{i \notin v} d_i^2) = (n-k)\mathbb{E}_{x \sim \text{Binom}(n,q_1)} x^2 = n(n-k)q_1 + \left((n-k)n^2 - n(n-k)\right) q_1^2$$

and the k nodes planted has:

$$\begin{aligned}
\mathbb{E}(\sum_{i \in v} d_i^2) &= k\mathbb{E}_{x \sim \text{Binom}(n-k,\, q_1), y \sim \text{Binom}(k,\, p_1)} (x+y)^2 \\
&= k\left(\left((n-k)q_1(1-q_1) + (n-k)^2 q_1^2\right) + kp_1(1-p_1) + k^2 p_1^2 + 2q_1 p_1 k(n-k)\right) \\
&= k(n-k)q_1 + k^2 p_1 + k((n-k)q_1 + kp_1)^2 - k(n-k)q_1^2 - k^2 p_1^2
\end{aligned}$$

The difference between expectations in H_0 and H_1 is thus:

$$k((n-k)q_1 + kp_1)^2 + (n-k)(n^2 - n - k)q_1^2 - k^2 p_1^2 - n^2(n-1)p_0^2 \in \Theta(k^3(p_1 - q_1)^2).$$

Now we turn to estimating the variance of f. First consider the variance of $f \sim H_0$, which can be computed via the moments of binomial distribution

$$\begin{aligned}
\mathsf{var}(f) &= n \cdot \mathsf{var}(d_i^2) \\
&= \mathbb{E}_{x \sim \text{Binom}(n,p_0)}(x^4) - \left(\mathbb{E}_{x \sim \text{Binom}(n,p_0)}(x^2)\right)^2 \\
&= n^2 p_0(1-p_0)(1 + (2n-6)p_0(1-p_0)) \in \Theta(n^3 p_0^2 (1-p_0)^2)
\end{aligned}$$

due to a simple computation $\mathbb{E}_{x\sim \text{Binom}(n,p_0)}((x-np)^4) = np(1-p)(1+(3n-6)p(1-p))$.

For the variance of $f \sim H_1$, consider

$$\begin{aligned}
\mathsf{var}(f) &= (n-k)\cdot \mathsf{var}_{i\notin v}(d_i^2) + k\cdot \mathsf{var}_{j\in v}(d_j^2)\\
&\le \Theta(n^3\,(q_1(1-q_1))^2) + k\cdot \mathbb{E}_{j\in v}\left((d_j-\bar{d})^4\right)\\
&\le \Theta(n^3\,(q_1(1-q_1))^2) + O(kn^2\,(q_1(1-q_1))^2)\\
&= \Theta(n^3 p_1^2(1-p_1)^2)
\end{aligned}$$

because of local inequality $(x+y)^4 \le 16(x^4+y^4)$. Therefore, we know that

$$\frac{\mathbb{E}_{H_1}(f) - \mathbb{E}_{H_0}(f)}{\sqrt{\mathsf{var}_{H_0}(f) + \mathsf{var}_{H_1}(f)}} \in O(\frac{k^3(p_1-q_1)^2}{n^{1.5}p_0(1-p_0)}) = O((\frac{k^2}{n})^{1.5} D_{KL}(p_1\|q_1))$$

and when this value is in $\omega(1)$, the two hypothesis can be separated (by, for instance, thresholding at $(\mathbb{E}_{H_1}(f) + \mathbb{E}_{H_0}(f))/2$). □

6.5.1 Proof of Theorem 3.1

Proof sketch: Step by step, our proof proceeds from establishing the following lemmas for each step of our reduction in figure 2, a formal proof for the lemmas will be presented later.

Lemma 17 (To-k-Partite-Submatrix – Lemma 7.5 in [BB20])**.** *With the given assumptions, step 1 (denote as $\mathcal{A}_1$) of the reduction runs in poly(N) time and it follows that:*

H_0: $d_{\mathrm{TV}}(\mathcal{A}_1(G(N,q)), \mathrm{Bern}(Q)^{\otimes n\times n}) \le 4k_0 \exp\left(\frac{-Q^2N^2}{48pkn}\right)$.

H_1: $d_{\mathrm{TV}}(\mathcal{A}_1(G(N,\mathcal{U}_N(E),p,q)), \mathcal{M}_{[n]\times[n]}(\mathcal{U}_n(S),p,Q) \le 4k_0 \exp\left(\frac{-Q^2N^2}{48pkn}\right) + \sqrt{\frac{C_Q k_0^2}{2n}}$.

where E is the partition of $[N]$ and S is the partition of $[n]$.

Lemma 18 (Bernoulli Rotations for PDS*)**.** *Let $\mathcal{A}_2$ denote the output matrix M from the second step of our reduction (before permutation). Suppose S is a partition of $[n]$ to k_0 equal parts and planted set $|T\cap S_i| = 1$ for all i. Let $M_i : S_i \to [n/k_0]$ be any fixed bijection. Let $K^{1/r}_{n/k_0}$ be the design matrix obtained from Lemma 5 with embedded sets $A_1, A_2, \ldots, A_{n/k_0} \subset [n/k_0]$, then the following holds:*

$$d_{\mathrm{TV}}(\mathcal{A}_2(\mathrm{Bern}(Q)^{\otimes n\times n}),\ \mathcal{N}(0,1)^{\otimes n\times n}) = O(n^{-1})$$

$$d_{\mathrm{TV}}(\mathcal{A}_2(\mathcal{M}_{[n]\times[n]}(\mathcal{U}_n(S),p,Q)),\ \mathcal{L}(\gamma\cdot X + \mathcal{N}(-\frac{\gamma}{r^2},1)^{\otimes n\times n})) = O(n^{-1})$$

where $X\in\mathbb{R}^{n\times n}$ is the random variable defined in each block $S_i\times S_j$ as a function of T:

$$X_{S_i,S_j} = \left(\mathbb{1}(M_i^{-1}(A_{f(i)}) \times M_j^{-1}(A_{f(j)}))\ \textit{where}\ f(i) = M_i(T\cap S_i)\right)$$

Lemma 19 (Thresholding from Gaussians)**.** *Let $\mathcal{A}_3$ be the final step from the above reduction, with the same notations as the previous lemma, then:*

$$\mathcal{A}_3(\mathcal{N}(0,1)^{\otimes n\times n}) \sim G(n,1/2)$$

$$\mathcal{A}_3(\mathcal{L}(\gamma\cdot X + \mathcal{N}(-\frac{\gamma}{r^2},1)^{\otimes n\times n})) \sim \mathrm{PDS}(n,k,P_1,P_2).$$

```
Algorithm To-$k$-Partite-Submatrix
Inputs: $k$-PDS instance $G$ with clique size $k$ that divides $N$ and partition $E$ of $[N]$, edge probabilities $0 < q < p \le 1$ with $q = N^{-O(1)}$ and target dimension $n \ge \left(\frac{p}{Q} + 1\right) N$ where $Q = 1 - \sqrt{(1-p)(1-q)} + \mathbf{1}_{\{p=1\}}\left(\sqrt{q} - 1\right)$ and $k$ divides $n$

  1. Apply Graph-Clone to $G$ with edge probabilities $P = p$ and $Q = 1 - \sqrt{(1-p)(1-q)} + \mathbf{1}_{\{p=1\}}\left(\sqrt{q} - 1\right)$ and $t = 2$ clones to obtain $(G_1, G_2)$.

  2. Let $F$ be a partition of $[n]$ with $[n] = F_1 \cup F_2 \cup \cdots \cup F_k$ and $|F_i| = n/k$. Form the matrix $M_{\text{PD}} \in \{0,1\}^{n \times n}$ as follows:

     (1) For each $t \in [k]$, sample $s_1^t \sim \text{Bin}(N/k, p)$ and $s_2^t \sim \text{Bin}(n/k, Q)$ and let $S_t$ be a subset of $F_t$ with $|S_t| = N/k$ selected uniformly at random. Sample $T_1^t \subseteq S_t$ and $T_2^t \subseteq F_t \backslash S_t$ with $|T_1^t| = s_1^t$ and $|T_2^t| = \max\{s_2^t - s_1^t, 0\}$ uniformly at random.

     (2) Now form the matrix $M_{\text{PD}}$ such that its $(i,j)$th entry is

         $(M_{\text{PD}})_{ij} = \begin{cases} \mathbf{1}_{\{\pi_t(i),\pi_t(j)\} \in E(G_1)} & \text{if } i < j \text{ and } i,j \in S_t \\ \mathbf{1}_{\{\pi_t(i),\pi_t(j)\} \in E(G_2)} & \text{if } i > j \text{ and } i,j \in S_t \\ \mathbf{1}_{\{i \in T_1^t\}} & \text{if } i = j \text{ and } i,j \in S_t \\ \mathbf{1}_{\{i \in T_2^t\}} & \text{if } i = j \text{ and } i,j \in F_t \backslash S_t \\ \sim_{\text{i.i.d.}} \text{Bern}(Q) & \text{if } i \neq j \text{ and } (i,j) \notin S_t^2 \text{ for a } t \in [k] \end{cases}$

         where $\pi_t : S_t \to E_t$ is a bijection chosen uniformly at random.

  3. Output the matrix $M_{\text{PD}}$ and the partition $F$.
```

Figure 3: Subroutine To-k-Partite-Submatrix for mapping from an instance of k-partite planted dense subgraph to a k-partite Bernoulli submatrix problem in [BB20]. See Lemma 7.5 therein.

As an important pre-processing step, we single out the steps in Lemma 17 first in Figure 3, the proof in its the exact form is deferred to [BB20]. The general idea is that, to construct a bi-partite variant, after applying Graph-Clone to the instance and occupying the lower half of the adjacency matrix, we still need to figure out what happens in the diagonal. However, when we plant around $\sqrt{k}$ entries in a diagonal it's *almost* the same as not planting anything in total variation, which means that we only need to blow up the size by a little bit to *hide* the diagonal. After Lemma 17, we arrive at a bi-partite k-PDS instance with slightly different parameters, and we prove the following lemmas to complete the reduction.

Proof of Lemma 18. We take a close look at what Bernoulli rotation produces for H_1. In the flattening step, we first define k_0 bijections π_i from $S_i \to [n/k_0]$ (the order to be flattened). Looking at each submatrix block with a planted bit at $(T \bigcap S_i, T \bigcap S_j)$ is equivalently an instance of

$$F_{S_i \times S_j} \sim \text{PB}((n/k_0)^2, t, p, Q)$$

where the location indices are defined with $(i,j) : i,j \in [n/k_0]$ and planted bit

$$t = (\pi_i(T \bigcap S_i), \pi_j(T \bigcap S_j)) := (t_i, t_j).$$

Therefore, the output row of $K^{1/r}_{n/k_0}$ is precisely (indexed by r, s):

$$K_{(t_i,t_j),(rs)} = \mu \left(\mathbb{1}\{r \in A_{t_i} \text{ and } s \in A_{t_j}\} \cdot \sqrt{\frac{r^3 k_0^3}{n^3}} - \sqrt{\frac{k_0^3}{rn^3}} \right).$$

After sending $\mathcal{A}(\cdot)_{(rs)} \to M_{\pi_i^{-1}(r), \pi_j^{-1}(s)}$, we know that $M \in \mathbb{R}^{(n/k_0)\times(n/k_0)}$ is a bi-partite matrix with $A_{t_i} \times A_{t_j}$ submatrix being elevated and (approximately) distributed as

$$\mathcal{M}_{S_i \times S_j}(\pi_i^{-1}(A_{t_i}) \times \pi_j^{-1}(A_{t_j}), \mathcal{N}((r^2-1)\gamma/r^2, 1), \mathcal{N}(-\gamma/r^2, 1)).$$

with total variation loss at most $O((\frac{n}{k_0})^2 R_{rk}^{-3})$ by Lemma 3.

For the other hypothesis H_0, simply note that the matrix gets sent to $\mathcal{N}(0,1)$ independently for each entry and gets send to independent standard normal Gaussians. Therefore the rotation matches.

Finally, note that in each block we differs from the target by at most $O(n^2 R_{rk}^{-3})$ in d_{TV}, which results in at most $O(n^4 R_{rk}^{-3})$ difference in d_{TV} by the tensorization property. However, note that we can choose R_{rk} to be any polynomial of n, and hence the Lemma holds. □

Proof of Lemma 19. Note that if we threshold at zero, then:

1. $\mathcal{N}(0,1) \to \mathrm{Bern}(1/2)$.
2. $\mathcal{N}(\mu,1) \to \mathrm{Bern}(\Phi(\mu))$ for any μ.

Therefore we know that $\mathcal{N}(\frac{-\gamma}{r^2}, 1) \to \mathrm{Bern}(\Phi(P_1))$, and $\mathcal{N}(\frac{(r^2-1)\gamma}{r^2}, 1) \to \mathrm{Bern}(\Phi(P_2))$, so the strength matches (and hence the case for H_0 is proven).

Note that in our previous step in H_1, in each block $S_i \times S_j$, a sub-block $A_{t_i} \times A_{t_j}$ is elevated to $\mathrm{Bern}(P_2)$ whereas the rest are $\mathrm{Bern}(P_1)$. This means that in the overall graph, the sets:

$$\bigcup_i \pi_i^{-1}(A_{t_i}) \times \bigcup_i \pi_i^{-1}(A_{t_i})$$

have elevated density $\mathrm{Bern}(P_2)$ where the rest has density $\mathrm{Bern}(P_1)$.

Finally, note that the total size of $\bigcup_i \pi_i^{-1}(A_{t_i})$ is exactly $\sum_{i=1}^{k_0} \frac{n}{k_0 r} = \frac{n}{r}$. Therefore, after permuting the nodes we get exactly $\mathrm{PDS}(n, n/r, P_2, P_1)$ as the output. □

Proof of Theorem 3.1. Define the steps of $\mathcal{A}$ to map inputs to outputs as follows

$$(G,E) \xrightarrow{\mathcal{A}_1, \epsilon_1} (F,S) \xrightarrow{\mathcal{A}_2, \epsilon_2} M \xrightarrow{\mathcal{A}_3, \epsilon_3 = 0} G'$$

where the following ϵ_i denotes the total variation difference in each step (from output of $\mathcal{A}$ to the next target). Under H_1, consider the following sequence of distributions:

$$\mathcal{P}_0 = G_E(N,k,p,q)$$
$$\mathcal{P}_1 = \mathcal{M}_{[n]\times[n]}(S \times S, \mathrm{Bern}(p), \mathrm{Bern}(Q)) \quad \text{where } S \sim \mathcal{U}_n(F)$$

$$\mathcal{P}_2 = \gamma \cdot \mathbb{1}_S \otimes \mathbb{1}_S + \mathcal{N}(-\frac{\gamma}{r^2}, 1)^{\otimes n \times n} \quad \text{where } S \sim \mathit{Unif}_n(k)$$
$$\mathcal{P}_4 = \mathrm{PDS}(n, k, P_1, P_2)$$

Applying Lemma 17 before, we can take

$$\epsilon_1 = 4k_0 \cdot \exp\left(-\frac{Q^2N^2}{48pk_0n}\right) + \sqrt{\frac{C_Q k_0^2}{2n}}$$

where $C_Q = \max\left\{\frac{Q}{1-Q}, \frac{1-Q}{Q}\right\}$. For ϵ_2, Lemma 18 guarantees that $\epsilon_2 = O(n^{-1})$ suffices. The final step $\mathcal{A}_3$ is exact and we can take $\epsilon_3 = 0$. Finally, note that from the data processing inequality applied to d_{TV} that $d_{\mathrm{TV}}(\mathcal{A}_i(\cdot), \mathcal{A}_i(\cdot')) \leq d_{\mathrm{TV}}(\cdot, \cdot')$ so each step the total variation loss at most accumulates (Lemma 9), thus by the triangle inequality on TV we get

$$d_{\mathrm{TV}}(\mathcal{A}(G_E(N, k, p, q)), \mathrm{PDS}(n, k, P_1, P_2)) \leq \epsilon_1 + \epsilon_2 = o(1).$$

Under H_0, consider the distributions

$$\mathcal{P}_0 = G(N, q)$$
$$\mathcal{P}_1 = \mathrm{Bern}(Q)^{\otimes n \times n}$$
$$\mathcal{P}_3 = \mathcal{N}(0, 1)^{\otimes n \times n}$$
$$\mathcal{P}_4 = G(n, 1/2)$$

As above, Lemmas Lemma 17, Lemma 18 and Lemma 19 imply that we can take

$$\epsilon_1 = 4k_0 \cdot \exp\left(-\frac{Q^2N^2}{48pk_0n}\right), \quad \epsilon_2 = O(n^{-1}), \quad \text{and} \quad \epsilon_3 = 0$$

Again by the data processing inequality (Lemma 9), we therefore have that

$$d_{\mathrm{TV}}\left(\mathcal{A}\left(G(N, q)\right), G(n, 1/2)\right) = O(\epsilon_1 + \epsilon_2) = o(1)$$

which completes the proof of the theorem. □

6.5.2 Proof of Theorem 3.2

Suppose we now have a direct reduction to $\mathrm{PDS}(n, k, P_1, P_2)$ following the previous notations, the final step of reduction have to do with the making the uniform degree condition exact, and applying it to general (dense) P_0. Consider the following post-reduction process with given target (P_0, p_1, p_2) such that $P_0 = p_1 - (\frac{n^2}{k^2} - 1)\delta = p_2 + \delta$ for some δ:

1. Apply k-PDS-to-PDS* on given instance $G_E(n, k, p, q)$ and output G_1 with specified μ, γ such that the exact condition $\Phi(\frac{(r^2-1)\gamma}{r^2}) = \frac{P_1}{2P_0}$ holds. As before, denote output density P_1, P_2 (then $p_1 = 2P_0P_1$, and δ can be expressed by P_0, γ, r).

2. If $P_0 > 1/2$, then output G_2 by including all edges in G_1 and independently including all non-edge in G_1 with probability $2P_0 - 1$, else include all edges in G_1 with probability $2P_0$.

We show that the output of the above second post-processing step $\mathcal{A}_4$ satisfies:

$$\mathcal{A}_4(G(n, 1/2)) = G(n, P_0)$$

$$d_{\text{TV}}(\mathcal{A}_4(\text{PDS}(n,k,P_1,P_2)),\, \text{PDS}(n,k,p_1,p_2)) = o(1)$$

These two equations completely settles the reduction from PC to PDS* to the general density.

Note that the first equation concerning H_0 is trivial, because a Bern(s) instance get transferred (independently) directly to Bern$(2P_0 s)$ by $\mathcal{A}_4$. Thus we only need to deal with the second equation. The general insights is that, when ν is small, $\Phi(\nu)$ (Gaussian CMF) is almost a linear function of ν where $\Phi(\nu) \sim \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\nu$ and the error term (when $\nu < 0.1$) is:

$$\left|\Phi(\nu) - \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\nu\right| = \left|\frac{1}{\sqrt{2\pi}}\int_0^\nu (e^{-x^2/2} - 1)dx\right| \le \frac{1}{\sqrt{2\pi}}\left|\int_0^\nu x^2 dx\right| = \frac{1}{3\sqrt{2\pi}}|\nu|^3$$

since $|e^x - 1| < 2|x|$ when $|x| < 0.01$. Therefore the average degree condition *approximately* but not exactly holds with P_1 and P_2 already.

Formally, note that $\mathcal{A}_4(\text{PDS}(n,k,P_1,P_2)) = \text{PDS}(n,k,p_1,2P_2P_0)$, and we only need to show that $d_{\text{TV}}(\text{PDS}(n,k,p_1,2P_2P_0), \text{PDS}(n,k,p_1,p_2)) = o(1)$. The trick here is to use the data processing inequality again: because the distribution $\text{PDS}(n,k,p,q)$ is obtained by applying the (random) planted dense subgraph over $G(n,q)$, thus the total variation:

$$d_{\text{TV}}(\text{PDS}(n,k,p_1,2P_2P_0), \text{PDS}(n,k,p_1,p_2)) \le d_{\text{TV}}(G(n,2P_2P_0), G(n,p_2))$$

$$= d_{\text{TV}}(\text{Bern}(2P_0P_2)^{\otimes\binom{n}{2}}, \text{Bern}(p_2)^{\otimes\binom{n}{2}}).$$

Moreover, by Lemma 11 we know that the above is bounded by $|2P_0P_2 - p_2| \cdot O(n)$ because the denominator $P_0(1-P_0) \in \Theta(1)$. Now we only need to prove that $|2P_0P_2 - p_2| \in o(n^{-1})$. Note that this can be computed as exactly:

$$\begin{aligned}
|2P_0P_2 - p_2| =& 2P_0\left|\Phi(\frac{-\gamma^2}{r^2}) - \frac{1}{2} + \frac{1}{r^2-1}\left(\frac{p_1}{2P_0} - \frac{1}{2}\right)\right| \\
=& 2P_0\left|\Phi(\frac{-\gamma^2}{r^2}) - \frac{1}{2} + \frac{1}{r^2-1}(\Phi(\frac{(r^2-1)\gamma}{r^2}) - \frac{1}{2})\right| \\
\le& 4\frac{\gamma^3}{r^2} = \frac{\mu}{n}\left(\frac{k_0^2}{n}\right)\left(\frac{k_0 r}{n}\right)^{2.5} = o(n^{-1})
\end{aligned}$$

because $\frac{k_0^2}{n} < \frac{k_0 r}{N}$, $\frac{k_0 r}{n} < \frac{k_0^2}{N}$ are all assumed to be smaller than one, and $\mu \to_n 0$.

We now turn to the formal lower bound from the reduction. Consider the following parametrized model $\text{PDS}(n,k,p_1,p_2)$ versus $G(n,p_0)$ such that:

$$p_0 = p_1 - \frac{r^2-1}{r^2}\gamma = p_2 + \frac{1}{r^2}\gamma$$

We prove that there is a computational threshold for all signal levels below $\gamma^2 \in \widetilde{o}((\frac{r^2}{n})^{1.5})$ by filling out all possible growth rates below.

Note that fix $P_0 \in (0,1)$ throughout (we only need it to be bounded away from 0 and 1), for the reduction to work with a given sequence (N, k_0, p, q) to $n = kr$ where $r \in \widetilde{o}(k)$, $k_0 \in \widetilde{o}(n^{1/2})$ and $k \in \widetilde{\omega}(k_0)$ are (implicit) functions of N, we only need to characterize the range of viable signal strength γ that can be reduced to:

$$\gamma = \mu(\frac{k_0 r}{n})^{1.5} > \frac{1}{w(n)\sqrt{\log n}}(\frac{k_0 r}{n})^{1.5}$$

asymptotically where w can be any (slowly) increasing unbounded function (such as $n^{o(1)}$). Note that this range do indeed cover the entirety of $\widetilde{o}((\frac{r^2}{n})^{1.5})$ assuming $k_0 \in \widetilde{\Theta}(N^{0.5})$.

Therefore, we know that by the PC conjecture and the given reduction the computational lower bound for PDS* holds up to the upper bound level in Lemma 16.

6.5.3 Proof of Corollary 3.3

Proof of Corollary 3.3. By Lemma 14, we only need to show that a weak recovery blackbox output is a qualifying secret key $k(G)$ for refutation.

Consider a set R that overlaps with the real PDS planted set with size $\rho > 1/2$ ($\rho \to_n 1$ holds for weak recovery, but for the sake here we only need it at least $1/2$ for convenience). Consider PDS density parameters $p > q > n^{-1}\log n$, $p \in O(q)$, and consider the sequence of $r_n = \frac{p_n + q_n}{2}$ such that $p = O(q) = O(r)$ and $D_{KL}(p\|r) = \Theta(D_{KL}(p\|q)) \subset \widetilde{\omega}(k^{-1})$. However, by flipping the graph for Theorem 6.6 in the dense case $\lim p = \lim q = p_0$, we know that the smallest ρk-subgraph in $G(n,p)$ has density at least r with high probability. Thus the density of R is at least $\frac{1}{4}(r + 3q) = \frac{7q+p}{8} := s$ with high probability.

However, by Theorem 6.6 again we note that the densest k subgraph in $G(n,p_0)$ will not be of density at least $\frac{s+p_0}{2}$ because $(s - p_0) = \Theta(p-q)$ so $D_{KL}((s+p_0)/2\|p_0) = \Theta(D_{KL}(p\|q)) \subset \widetilde{\omega}(k^{-1})$. Therefore with high probability the densest k subgraph of $G(n,p_0)$ has a gap with the density of $1/2$ portion recovered densest k subgraph in PDS*. By Lemma 14 and Theorem 3.2 we are done with the proof. □

6.5.4 Statistical Boundary for PDS*

It is well known that the success of a statistical hypothesis testing between two distributions P, Q from one sample depends on $d_{\mathrm{TV}}(P,Q)$. However, because our alternate hypothesis is composite (mixture over latent θ), it can be challenging to compute the total variation distance between mixture $\mathbb{E}_\theta P_\theta$ and null P_0 beyond trivial geometric bounds. Thus alternative methods are needed.

In this section, for the completeness of our results on PDS*, we also present a statement of the statistical boundaries drawing comparisons with a line of statistical lower bounds in the canonical PDS such as [BI13][HWX15a][MW15] where their upper bound construction with mean comparison is now invalid in PDS*. While we can derive asymptotically similar lower bounds, there is provably no polynomial test matching this boundary in PDS* *and* recovery is impossible. Instead, we derive a boundary necessary for the χ^2 divergence between H_0 and H_1 to be large via the Ingster's trick to handle mixture in the latent structure.

Theorem 6.3 (Statistical lower bounds for PDS*). *Consider* PDS* *when* $0 < q < p_0 < q < 1/2$. *Consider the setting with a sequence of edge densities* $p^{(n)}, p_0^{(n)}, q^{(n)}, k^{(n)}$ *with graph size* $n \to \infty$*:*

- *If* $\limsup \frac{k^{(n)4}}{n^2} \cdot \frac{(p^{(n)} - q^{(n)})^2}{q^{(n)}(1-q^{(n)})} \to 0$ *and* $\limsup k^{(n)} \frac{(p^{(n)} - q^{(n)})^2}{q^{(n)}(1-q^{(n)})} \to 0$*, then no (statistical) test on* PDS* *on those parameters can achieve type I + type II error strictly less than* 1 *asymptotically.*

Proof. Consider the χ^2 trick applied on the mixture: $P_0 = G(n,p_0)$ and $P_\theta = \mathrm{PDS}(n,\theta,p,q)$ with planted set at θ and $\theta \sim \binom{n}{k}$ be uniformly distributed.

$$\chi^2(\mathbb{E}_\theta(P_\theta)\|P_0) = \int_G \frac{\mathbb{E}_\theta(P_\theta(G))\mathbb{E}_{\theta'}(P_{\theta'}(G))}{P_0(G)} - 1$$
$$= \mathbb{E}_{\theta \perp\!\!\!\perp \theta'} \frac{P_\theta(G)\mathbb{E}_{\theta'}(G)}{P_0(G)} - 1$$

If we expand the above expression and denote the $\lambda = \chi^2(p,q) = \frac{(p-q)^2}{q(1-q)}$ then we end up with the above (tightly) upper bounded by:

$$E(\exp(\lambda(H^2 - E(H)^2))$$

where $H \sim \theta \bigcap \theta'$ is distributed according to Hypergeometric(n,k,k) (where $k > \sqrt{n}$). This evaluation goes to zero from a local inequality on Hypergeometric inequalities (Lemma 6 in Appendix C of [HWX15a]), which concludes our proof.

A better way to view it (from reductions) is as follows: we know that when the inequality condition holds, PDS(n,k,p,q) is in-distinguishable from $G(n,k,q)$ by the PDS boundary in [BBH18]. However, in this case we have:

$$d_{\text{TV}}(G(n,q), G(n,p_0)) \le d_{\text{TV}}(\text{Bern}(q)^{\otimes n^2}, \text{Bern}(p_0)^{\otimes n^2})$$
$$\le n\sqrt{\frac{(q-p_0)^2}{q(1-q)}} = \frac{n}{r^2}\sqrt{\lambda}$$

by Lemma 11. Therefore, $d_{\text{TV}}(G(n,q), G(n,p_0)) \to 0$ below the statistical PDS$_D$ testing threshold, meaning that PDS$^*_D$ can also not be performed. This reduction proves the intuition that PDS$^*_D$ is "harder" than PDS. □

Remark: We also remark a counter-result for the above lower bound. Let H be distributed according to Hypergeo(n,k,k) (where $k > \sqrt{n}$), then $E(H) = k^2/n$. Assuming that $\lambda E(H)^2 \in \widetilde{\omega}(1)$, if one can prove that:

$$E(\exp(\lambda(H^2 - E(H)^2)) \to \infty$$

as well, which is stronger than Lemma 6 in [HWX15a], then the χ^2 between two hypothesis of PDS$^*_D$ diverges, which is still only a necessary but insufficient condition for the upper bound.

6.6 Proofs in section 5

6.6.1 Proof of Theorem 4.1

Theorem 6.4 (Reduction from $k-$PDS to ISBM)**.** *Let N, k_0 be parameters of planted clique graph size, $r < N/k_0$ be a target output for ratio of planted set. Let α be a constant and assume $r < \left(\frac{N}{k_0}\right)^{1-\alpha}$. We present the following reduction ϕ with absolute constant $C_\alpha > 1$:*

- Initial k-PDS Parameters: *vertex count N, subgraph size $k_0 \in o_N(N)$ dividing N, edge probabilities $0 < q < p \le 1$ with $\min\{q, 1-q, p-q\} = \Omega(1)$, and a partition E of $[N]$. We further assume that $k_0 \in o(\sqrt{N})$ holds.*

- Target PDS parameters: *(n,r,k) where $r \in o_n(\sqrt{n})$ is the specified parameter and k is the expected subgraph size $k = n/r$ and n is the smallest multiple of $k_0 r$ that is greater than $(1+\frac{p}{Q})N$ where*

$$Q = 1 - \sqrt{(1-p)(1-q)} + 1_{p=1}(\sqrt{q} - 1).$$

- Target Binomial PDS edge strength:

$$\gamma = \mu(\frac{k_0 r}{n}), \quad P_{11} = \Phi(\frac{(r-1)^2\gamma}{r^2}), \quad P_{12} = \Phi(-\frac{(r-1)\gamma}{r^2}), \quad P_{22} = \Phi(\frac{\gamma}{r^2})$$

where $\mu \in (0,1)$ satisfies that

$$\mu \leq \frac{1}{12C\sqrt{\log(N) + \log(p-Q)^{-1}}} \cdot \min\{\log(\frac{p}{Q}), \log(\frac{1-Q}{1-p})\}.$$

- *Applying ϕ on the given input graph instance G yields the following (when $k_0 \in o(\sqrt{N})$):*

$$d_{\mathrm{TV}}(\phi(G(N, \frac{1}{2})), G(n, \frac{1}{2})) = o_n(1)$$

$$d_{\mathrm{TV}}(\phi(PC_\rho(N, k_0, \frac{1}{2})), \mathrm{ISBM}(n, k, P_{11}, P_{12}, P_{22})) = o_n(1)$$

Algorithm From k-PDS to ISBM:

Inputs: Graph G of size N, subgraph parameter k_0 dividing N, edge density $q < p \in (0,1]$ and a partition E of $[N]$ to k_0 equal parts $E_1, E_2, \ldots, E_t$. Target planted ratio $r = \epsilon^{-1}$.

Steps :

1. *To-bipartite and planted diagonal:* Compute $M_1 \in \{0,1\}^{m \times m}$ with partition F of $[n]$ as To-k-Partite-Submatrix (Lemma 17) applied with initial dimension N, partition E, edge probabilities p and q and target dimension m.

2. *Flattened Bernoulli Rotations:* Let S be a partition of $[n]$ into k_0 equal parts $S_1, S_2, \ldots, S_{k_0}$ obtained from the previous part. Construct output matrix M:

   (a) For i, j in $\{1, 2, \ldots, k_0\}$, flatten matrix F_{S_i,S_j} to a $(n/k_0)^2$ size vector.

   (b) Apply Bernoulli Rotation on this vector with design matrix $\frac{1}{C'^2}\left(R_{n/k_0,\epsilon} \otimes R_{n/k_0,\epsilon}\right)$, Bernoulli parameter strengths $Q < p \leq 1$, output dimensions $v_{ij} \in \mathbb{R}^{(n/k_0)^2}$.

   (c) Layout vector $v_{ij} \in \mathbb{R}^{(n/k_0)^2}$ to $(n/k_0) \times (n/k_0)$ matrix in the order from part (a).

   (d) Assign submatrix M_{S_i,S_j} to be the values of flattened v.

3. *Thresholding:* Given matrix M from the previous step, construct $G' = \phi(G)$ such that: for distinct indices $i < j$, $e_{ij} \in E(G')$ if and only if $M_{ij} \geq 0$. Finally, randomly permute vertex orders in G' and output.

Figure 4: Reduction from k-PDS to ISBM.

The key distinction here from the reduction in PDS* lies almost solely in the design matrix, which is simply the Kronecker (tensor) product of two matrices given by Lemma 4. We hence by establish the following lemma:

Lemma 20 (Bernoulli Rotation for ISBM$_D$)**.** *Consider the second step $\mathcal{A}_2$ applied on the output of Lemma 17 and assuming notations through Lemma 18, we have:*

$$\mathcal{A}_2(\mathrm{Bern}(Q)^{\otimes n \times n}, \mathcal{N}(0,1)^{\otimes n \times n}) = o(n^{-1})$$

$$d_{\mathrm{TV}}(\mathcal{A}(\mathcal{M}_{[n]\times[n]}(\mathcal{U}_n(S), p, Q)),\ \mathcal{L}(\frac{\gamma}{r^2}\cdot(rv-1)^T(rv-1)+\mathcal{N}(0,1)^{\otimes n\times n})) = O(n^{-1})$$

where $v \sim Unif_n(k)$.

Proof. Similar to Lemma 18, the only different part is the output to Bernoulli Rotation in H_1 after performing distribution shifts to $\mathcal{N}(0,1)$ and $\mathcal{N}(\mu,1)$.

We analyze the output for the specific design matrix. For the sub-block $S_i \times S_j$, it gets mapped to exactly the flattened product of the t_ith row of $\frac{1}{C}R_{n,r^{-1}}$ (transposed) times the t_jth row of $\frac{1}{C}R_{n,r^{-1}}$. Denote the set of positive terms in the tth row to be P_t, then the output distribution (conditioning on T, the source planted set) is exactly:

$$\mathcal{L}(\frac{\gamma}{r^2}\cdot(rv-1)^T(rv-1)+\mathcal{N}(0,1)^{\otimes n\times n}))$$

for $v = \prod_{i=1}^{k_0} P_{t_i}$ (this can also be viewed in the light *Tensor Bernoulli Rotation*, see Corollary 8.2 in [BB20]). Afterwards we can simply permute the nodes (note that v has size $k = n/r$) and the Bernoulli rotation target follows.

To finish off, we get the exact same bound on total variation at most $o(n^4 R_{rk}^{-3})$, at which point the total variation bound from applying this algorithm step by $o(n^{-1})$. □

After the Bernoulli rotation, we proceed in a similar fashion with Theorem 3.1. We defer a formal full proof in thresholding Gaussians and aligning the precise density to Corollary 14.5 and Theorem 3.2 in [BB20], here we only remove the condition (T) they imposed during Bernoulli Rotation to imply the lower bound results in a general regime. This gives us the desired lower bound.

6.6.2 Proof of Theorem 4.2

Proof of refutation hardness. Consider applying Lemma 13, we know that as long as we can find a satisfying "quiet" adversarial distribution H_1, such that

- H_1 is computationally indistinguishable from H_0.
- H_0, H_1 satisfies the refutation valuation function criteria.

then we can claim that refutation is hard for H_0. For the case of the Erdős-Renyi graph null hypothesis $H_0 : G(n, P_0)$ and $val = \mathrm{DkS}$, we may simply consider the alternate hypothesis as $H_1 : \mathrm{ISBM}(n, k, P_{11}, P_{12}, P_{22})$ with specific pairs of parameters.

In fact, when $k(P_{11} - P_0) \in \widetilde{\omega}(1)$ we know that the densest subgraph in H_1 has density at least $\frac{2P_{11}+P_0}{3}$ with probability $1 - o_n(1)$ from the Markov inequality. And by Theorem 6.6 we know that as long as $k(P_{11} - P_0)^2 \in \widetilde{\omega}(1)$ the (statistical) densest k-subgraph in H_0 is smaller than $(P_0 + P_{11})/2$ with high probability. These two constraints of parameter growth will be satisfied because the (optimal) output regime for Theorem 4.1 actually reads $(P_{11} - P_0)^2 \in \widetilde{\Theta}(n/k^2)$.

Therefore, from Theorem 4.1 we know that as long as H_0 is indistinguishable with H_1, one cannot refute in polynomial time the densest k-subgraph in $G(n, P_0)$ to have value larger than $\frac{P_0+2P_{11}}{3} := q$. Plugging in the boundary for ISBM_D we know that refutation (of PDS) is computationally impossible under the regime

$$\frac{k^2 D_{KL}(p\|q)}{n} \in \widetilde{o}(1)$$

which contrasts the detection threshold $\frac{k^4 D_{KL}(p\|q)}{n^2} \in O(1)$ above which one can perform the optimal sum-test. This fact, combined with semi-definite programming, completely resolves the refutation problem of DkS in Erdős-Renyi graphs. □

6.6.3 Computational upper bound for refutation

To prove an upper bound for refutation, we first need to introduce the semi-definite programming relaxation, which is a common method to computational approach problems such as densest-k-subgraph, considered in many works such as [HWX16][CX14].

Consider the following formulation for the densest-k-subgraph problem:

$$\begin{aligned} \widehat{\xi} = \arg\max_{\xi} \ & \sum_{i,j} e_{ij}\xi_i\xi_j \\ \text{s.t.} \ & \xi \in \{0,1\}^n \\ & \xi^\top \mathbb{1} = k, \end{aligned} \tag{6}$$

Now it is well-known that integer linear programming is NP-complete, which motivates us to consider the following variant:

$$\begin{aligned} \widehat{Z} = \arg\max_{Z} \ & \langle E, Z\rangle \\ \text{s.t.} \ & Z \succeq 0 \\ & Z_{ii} \le 1, \quad \forall i \in [n] \\ & Z \ge 0 \\ & \langle I, Z\rangle = k \\ & \langle J, Z\rangle = k^2. \end{aligned} \tag{7}$$

It is not hard to see that:

- A feasible solution of (6) is also feasible for (7), thus the latter will always return objective at least the former.
- (7) is a semi-definite programming problem, and can be numerically solved to any precision in polynomial time.

With the sufficiency results given in [HWX16] (specifically, combine their results in Lemma 14, Lemma 15, and Theorem 5), we can show that under the separation conditions of $\frac{k^2}{n}\frac{(p-q)^2}{q(1-q)} \to \infty$ and $k\frac{(p-q)^2}{p(1-p)} \to \infty$, the above formulation of convexified programming for planted dense subgraph will have the optimal solution converging to the true planted instance of our graph $P(\widehat{Z}_{SDP} = Z) \to 1$ assuming the null density satisfies $0.9 > q \in \Omega(\frac{\log n}{n})$. Here we use their results for the objective function instead (that is, the objective $\langle E, Z\rangle \le k^2p + ck\sqrt{p(1-p)}$ with probability $\to 1$ as constant $c \in \Omega_{n,k}(1)$ by Markov Inequality).

Theorem 6.5. *Consider the semi-definite programming relaxation for $G \sim G(n,q)$. Then when $\frac{k^2}{n}\frac{(p-q)^2}{q(1-q)} \to \infty$ and $k\frac{(p-q)^2}{p(1-p)} \to \infty$, the probability that the objective function is at least $k^2\frac{q+2p}{3}$ goes to zero as $n,k \to \infty$. Moreover, in $H_1 = \mathrm{PDS}(n,k,p,q)$, the objective will be at least $\frac{q+4p}{5}$ with probability $\to 1$. This means that (7) will successfully refute the densest k-subgraph valuation problem in $G(n,q)$ vs $\mathrm{PDS}(n,k,p,q)$.*

Proof. We start with the following lemma from stochastic domination:

Lemma 21. *Let $F(P)$ be the distribution of objective (7) under the graph distribution P. If edges in $G \sim P$ are sampled independently with probability matrix E_P for two distributions P, Q, such that $E_P - E_Q \ge 0$ (entry-wise), then for any $x > 0$, $\mathbb{P}(F(P) > x) \ge \mathbb{P}(F(Q) > x)$. In other words, the convex program is monotone with respect to the underlying density.*

Proof. Consider the following process:

1. On $G \sim Q$, find optimal $\widehat{Z}$ for (7).
2. Then, update G in the following way: for any $e_G = 0$, flip $e_G = 1$ with probability $\frac{E_p(e)-E_q(e)}{1-E_q(e)}$.
3. Find optimal $\widehat{Z}'$ for (7).

Note that the objective never decreases in the third step because we only add edges in the second step, whereas the unconditional distribution of graph generated from 2 is exactly P. Hence we find a coupling between two distribution of graphs such that $F(P|G)$ is bounded below by $F(\{G\})$ for any G, and the result follows. □

Moreover, note that the above lemma applies to the mixture problem too. Since (7) is symmetric, the objective will not change if we condition the planted dense subgraph to a specific location, then we can use the above lemma and conclude that $F(\mathrm{PDS}(n, k, p, q))$ dominates $F(G(n, q))$ for any density $p > q$.

Consider the alternative $\mathrm{PDS}(n, k, (p + q)/2, q)$ for (7), which would also hold the asymptotic conditions for successful recovery by (7). Note that in this case, we know that the objective:

$$\frac{1}{k^2}\langle E, \widehat{Z}\rangle \leq \frac{p+q}{2} + \frac{c}{k}\sqrt{p(1-p)}$$

with probability → 1 if $c \to \infty$. Consider plugging in p to the RHS we get $c = \frac{k}{6} \cdot \frac{(p-q)}{\sqrt{p(1-p)}} \to \infty$ by the asymptotic conditions. Thus we know that the above objective is bounded above by $k^2 \cdot \frac{q+2p}{3}$ with probability going to 1.

On the other hand, clearly for $X \sim \mathrm{Binom}(k^2, p)$, we have $X \leq k^2 \cdot \frac{4p+q}{5}$ with probability at most

$$\mathbb{P}(X \leq k^2 \cdot \frac{4p+q}{5} | \mathrm{Binom}(k^2, p)) \leq (\frac{k(p-q)}{5\sqrt{p(1-p)}})^{-2} \to 0$$

by Markov inequality. So the valuation condition for H_1 is met. □

6.6.4 Statistical bounds for refutation

We now turn to show that the statistical limit of DkS problem lies upon recovery boundary for $G(n, q)$ (ignoring log factors). This has also been studied under the name quasi-cliques (k-subgraphs with edge count at least $\gamma\binom{k}{2}$) in random graphs by a line of works such as [Ver+12][AS16][Bal+19]. In our version of the problem, a simple moment inequality would suffice.

Theorem 6.6. *Consider $d = D_{KL}(\mathrm{Bern}(p)\|\mathrm{Bern}(q)) = \Theta(\frac{(p-q)^2}{q(1-q)})$ when the densities $p/q \to \Theta(1)$ and $np > nq > \log n$. Then the densest k subgraph density of $G(n, q)$ will be smaller than $\frac{p+q}{2}$ with probability → 1 if $\frac{kd}{\log n} \to \infty$ and $k \to \infty$. Moreover, in this case the densest k subgraph density of* $\mathrm{PDS}(n, k, p, q)$ *will be at least $\frac{q+2p}{3}$. Thus statistical refutation is possible.*

Proof. The latter half of the statement is proven in the previous section. Here we only deal with the first half (upper bound with $G(n, q)$). Firstly we need a tail bound on the Binomial distribution (for $r := \lceil pN \rceil$):

$$\begin{aligned}\mathbb{P}(\text{Binom}(N,q) \geq pN) &\leq N \cdot \mathbb{P}(\text{Binom}(N,q) = r)\\ &= N\binom{N}{r} q^r (1-q)^{N-r}\\ &\leq N^2 \frac{N!}{r!(N-r)!} e^{N(p\log q + (1-p)\log(1-q))}\\ &< 2N^2 \frac{1}{\sqrt{2\pi p(1-p)N}} e^{ND_{KL}(p\|q)} = e^{-ND_{KL}(p\|q) + O(\log N)}\end{aligned}$$

from Stirling's formula and $N, r \to \infty$.

Therefore, we can go on to look at each block, which has k^2 independent Bernoullis and thus satisfies the density tail with probability at most $e^{-k^2 d + O(\log k)}$. However, there are at most $\binom{n}{k}$ such blocks, so if assign random variables $X = \sum X_i$ to those we have:

$$\mathbb{P}(X > 0) \leq \mathbb{E}(X) = \sum \mathbb{E}(X_i) \leq n^k e^{-k^2 d + O(\log k)} = e^{-k^2 d + k\log n + O(\log k)}$$

when $\frac{kd}{\log n} \to \infty$, we know that the above objective goes to zero. Replacing p with $\frac{p+q}{2}$ for the above arguments works the same, and thus we are done.

As an extension, when $\frac{kd_k}{\log n} \to_k \infty$ with parameters p_k, we can show that (via a union bound) the densest k subgraph density does not exceed $\frac{q+p_k}{2}$ for all $k > \log n$ simultaneously because the objective is bounded by $\exp\left(-k(kd - \log n) + O(\log k)\right) < \exp(-k\log n) = n^{-k}$). □

Next, we deal with the lower bound on refutation, which states that in $G(n,q)$ there is a dense subgraph with density p and size k with high probability if $kd \in \widetilde{o}(1)$. The following theorem is sufficient to close the boundary for statistical impossibility. Assuming the same set of parameters, we have the following lower bound:

Theorem 6.7 (Lower bound on refutation)**.** *Assuming that $k \in o(n^\alpha) \bigcap \widetilde{\omega}(\sqrt{n})$ for some fixed constant $\alpha < 1$ and p, q are all bounded away from 0 and 1*[3]*. Moreover, for any $\alpha > 0$, assume that $k(p-q)^2 \in \Theta(\log^{-1.01} n)$. There exist a k-subgraph in $G(n,q)$ with density at least $(p+q)/2$ with probability $1 - o_n(1)$.*

Proof. First of all, the conditions assert that $p - q \in \Omega(\frac{\log^2 n}{n})$. We examine what happens in this parameter regime for the planted distribution $\text{PDS}(n,k,p,q)$. In fact, from the statistical boundary on exact recovery given in [HWX15b], we know that recovery is impossible in this regime, even in the minimal variant (from the reduction given in Lemma 15).

Consider the densest k subgraph estimator $\widehat{E}$, which happens to be the MLE estimator on the planted instance (though we do not need this fact), we know that under this regime it correlates with the true planted mean with expected density $< \epsilon$ for any constant ϵ asymptotically (partial recovery), we try to bound the density in $\widehat{E}$ before planted dense subgraph.

Formally, assume that $\widehat{E} \cap E = T$ where E is the true planted set. Consider the original G_1 the instance from $G(n,q)$ and G_1' be the graph after planting on E. The total edges in $G_1(\widehat{E})$ is at least (since it is the densest subgraph in G_1'):

$$E_{\widehat{E}}^{G_1} = E_{\widehat{E}}^{G_1'} - E_T^{G_1'} + E_T^{G_1} \geq E_E^{G_1'} - E_T^{G_1'} + E_T^{G_1}$$

[3] Observe that here the KL divergence reduces to $\Theta((p-q)^2)$ and $\log n/k = \Theta(\log n) = \Theta(\log k)$

and we bound those terms one by one. To start, note that $|T| > \log n$, else the total edges offset in T is at most $\binom{|T|}{2} < (\log n)^2$, and $\binom{|T|}{2}/\binom{k}{2} = O((\log n)^2/k^2) \subset o(p-q)$. Now we consider the case when $|T| > \log n \to \infty$ and apply the densest $|T|$ subgraph in $(G_1')_E$:

1. $E_E^{G_1'}$ is just the edge count of the planted instance that is distributed according to $G(k,p)$. We know that the total number of edges is at least
$$E_E^{G_1'} \geq \frac{2p+q}{3}\binom{k}{2}$$
from a simple Markov inequality (as in the previous theorem).

2. $E_T^{G_1'}$ is equivalent to $|T|$-subgraph sampled from $G(n,q)$. From the previous theorem, we know that if $\frac{|T|(p-r_{|T|})^2}{\log k} \in \omega(1)$, then with probability $1-o(1)$ the densest $|T|$ subgraph in planted set has density at most $r_{|T|}$, and $E_T^{G_1'} < r_{|T|}\binom{|T|}{2}$.

3. Similar as the previous part, we know that when $\frac{|T|(p-s_{|T|})^2}{\log n} \in \omega(1)$ the probability that G_1 has such a *sparse* subgraph is at most $1-o(1)$ (note that here we use the reverse side of the tail bound, which is a trivial implication when p, q are bounded away by one) and $E_T^{G_1} > s_{|T|}\binom{|T|}{2}$.

Combining the above, we only need to show that:
$$(r_{|T|} - s_{|T|})\binom{|T|}{2} \leq \frac{1}{6}(p-q)\binom{k}{2}.$$

Let $d_{|T|} = r_{|T|} - p > 0$ and $f_{|T|} = q - s_{|T|} > 0$ then $|T|(d_T^2 + f_{|T|}^2) \in O(\log n)$, thus the sum bound over all edges $|T|^2(d_T + f_{|T|}) \in O(\sqrt{\log n}|T|^{3/2})$. Moreover, recall the condition on p, q we have $k^2(p-q) \in \Theta(k^{3/2}\sqrt{\log^{-1.01} n})$.[4]

Now note that $r - s = p - q + (d+f)$, thus what remains to show is the local inequality
$$|T| \in o(\frac{k}{\log^{2/3} k}), \qquad \text{when } k(p-q) \in \Theta(\log^{-1} n)$$
after which apply the fact that $\binom{|T|}{2}(d+f) \in o(k^2(p-q))$ we are done.

Finally, note that the information theoretical limit for precise recovery is $k(p-q)^2 \in \Theta(\log n)$ (Theorem 6.1), below which it is impossible to perform (even weak) recovery, so the above bound on $|T|$ follows immediately from Lemma 15 with strength $\alpha = 1.004$ (so the expected size of $|T|$ cannot be greater than $\frac{k}{(\log k)^{1.004}} < \frac{k}{(\log k)^{2/3}}$ by minimal recovery). □

6.7 Proofs in section 6

Lemma 22 (Design matrix with independent entries). *For a fixed constant d and a $n \times n$ (random) matrix $B = B_{n,1/r}$ with entries sampled from the following distribution:*

- $B_{ij} = B_{ji} = \frac{r-1}{\sqrt{nr}}$ *w.p.* $\frac{1}{r}$, *and* $\frac{-1}{\sqrt{nr}}$ *w.p.* $\frac{r-1}{r}$ *for* $j \neq i$ *off diagonal.*
- $B_{ii} = \frac{-1}{\sqrt{nr}}$ *on the diagonal.*

[4] Note that here the key is that (somewhat counterintuitively) we want p, q to be far enough so that we can utilize the fact that small error terms cannot dominate the total density of at least p in $\widehat{E}$.

If $r \in o_n(n/(\log n)^4)$, then there exist an absolute constant $c' > 0$ s.t. with probability $1 - n^{-d}$ this matrix satisfies the following constraints:

- *The operator norm $\|B\| \le C := c'd$.*
- *Every column of B has between $n/r - C \log n\sqrt{n/r}$ to $n/r + C \log n\sqrt{n/r}$ positive entries.*

For $r^{-1} = \epsilon \in (0, 1/2]$, let $B_{n,\epsilon} \in \mathbb{R}^{n\times n}$ denote the random symmetric matrix with independent entries sampled as follows

$$(B_{n,\epsilon})_{ij} = (B_{n,\epsilon})_{ji} \sim \begin{cases} \sqrt{\frac{1-\epsilon}{\epsilon n}} & \text{with prob. } \epsilon \\ -\sqrt{\frac{\epsilon}{(1-\epsilon)n}} & \text{with prob. } 1-\epsilon \end{cases}$$

for all $1 \le i < j \le n$, and $(B_{n,\epsilon})_{ii} = -\sqrt{\frac{\epsilon}{(1-\epsilon)n}}$ for each $1 \le i \le n$.

The following lemma strengthens Lemma 8.13 in [BB20] where we also upper bound the probability of failure by a polynomial of n by incurring a stronger lemma (Theorem 2.1 in [LLV15]) in random graph theory. This ensures that the total variation loss between a satisfying sample and a random binomial sample is small enough, which is crucial to the reduction.

Lemma 23 (Key Properties of $B_{n,\epsilon}$)**.** *If $\epsilon \in (0, 1/2]$, $\alpha > 0$ satisfies that $\epsilon n = \omega_n(\log n)$, there is a constant $C_\alpha > 0$ such that the random matrix $B_{n,\epsilon}$ satisfies the following two conditions with probability over $1 - n^{-\alpha}$:*

1. *the largest singular value of $B_{n,\epsilon}$ is at most C_α; and*
2. *every column of $B_{n,\epsilon}$ contains between $\epsilon n - C_\alpha\sqrt{\epsilon n \log n}$ and $\epsilon n + C_\alpha\sqrt{\epsilon n \log n}$ negative entries.*

Proof. We first note that the second part of the lemma is a direct consequence of the Chernoff Bound applied to Binomial distributions $X \sim \text{Binom}(n, \epsilon)$:

$$\mathbb{P}(|X - n\epsilon| > \delta n\epsilon) \le 2\exp\{-\delta^2 n\epsilon/3\}$$

where we let $\delta = C_\alpha\sqrt{(\epsilon n)^{-1}\log n}$ so the above holds with probability at most $2n^{-C^2/3}$. Now we turn to the first condition on spectral norm.

As states in Theorem 2.1 in [LLV15] without modifying the graph, we know that with probability $1 - n^{-\alpha}$, the spectral concentration of the adjacency matrix of graph $G(n, \epsilon)$ is at most $O_\alpha(\sqrt{n\epsilon})$, which results in exactly $\sigma_{op}(B_{n,\epsilon}) \in O_\alpha(1)$, as desired. □

The above result gives us a good starting point at reductions in the binomial setting. To complete the picture for reduction to binomial PDS, we still need the following lemma that transfers a planted bit tensor in $\mathbb{R}^{n\times n}$ to an elevated $\epsilon n \times \epsilon n$ planted tensor:

Lemma 24 (Design matrix for PDS$_B$)**.** *Consider the following matrix $L \in \mathbb{R}^{n^2\times n^2}$ defined as:*

$$L^{(1,2)}_{n,\epsilon} = (B^{(1)}_{n,\epsilon} + \sqrt{\frac{\epsilon}{(1-\epsilon)n}}J_n) \otimes (B^{(2)}_{n,\epsilon} + \sqrt{\frac{\epsilon}{(1-\epsilon)n}}J_n)$$

satisfies for some constant C:

1. *$B^{(t)}_{n,\epsilon}$ are satisfying matrices from the previous lemma for $t = 1, 2$.*

2. *$L_{(ij),(kl)} = (B^{(1)}_{ik} + \sqrt{\frac{\epsilon}{(1-\epsilon)n}})(B^{(2)}_{jl} + \sqrt{\frac{\epsilon}{(1-\epsilon)n}})$ is equivalent to (fix (i,j)):*

$$L_{(ij),(kl)} = \begin{cases} 0 & \textit{if } (l \notin N_G(j)) \vee (k \notin M_G(i)) \\ \frac{1}{\epsilon(1-\epsilon)n} & \textit{if } (l \in N_G(j)) \wedge (k \in M_G(i)) \end{cases}$$

where $G \sim G^{(t)}(n, \epsilon)$ is the graph $R^{(t)}$ is sampled from and N_G, M_G are their neighborhood functions.

3. *The operator norm $\sigma(L^{(1,2)}_{n,\epsilon}) \le Cn\epsilon$.*

Proof. The first property follows directly from the definition. The second property is established similar to Lemma 5. In fact, we have:

$$\sigma(L) \le \sigma(B^{(1)} \otimes B^{(2)}) + \sqrt{\frac{\epsilon}{(1-\epsilon)n}}(\sigma(B^{(1)} \times J) + \sigma(B^{(2)} \times J)) + \frac{\epsilon}{(1-\epsilon)n}\sigma(J \times J)$$

and since $\sigma(B) \le C$, $\sigma(J) = n$, the above is bounded by $C^2 + 2C\sqrt{\epsilon n} + n\epsilon$ (since $1 - \epsilon < 1$) and is clearly bounded by $C'n\epsilon$ for some constant C' as long as $n\epsilon \in \omega(1)$. □

6.7.1 Proof of Theorem 5.1

Firstly, define $\mathrm{PDS}_B(n, k, p, q)$ to be the upper half (so that it remains a symmetric graph) of planted distribution $\mathcal{M}_{[n]\times[n]}(\mathrm{Bern}(\epsilon)^{\otimes n}, \mathrm{Bern}(p), \mathrm{Bern}(q))$ where $\epsilon := k/n$. As before, we also denote $r := n/k = \epsilon^{-1}$.

We show the proof of an average case reduction from fixed size PDS (PC) to a binomially planted set PDS_B preserving the boundary tightness of computation detection (for fixed community size) as established in [MW15][BBH18]. The rest of the theorem (ISBM_B and PDS^*_B) follows similar in spirit in that we only need to slightly alter our design matrix.

We start with the first lemma for approximating the output graph distribution, which states that as long as the distribution for the latent structure are similar, the total distribution on the graphs will be similar as well.

Lemma 25. *When the distributions $\mathcal{P}_1, \mathcal{P}_2$ on the (planted) mean μ have total variation difference $d_{\mathrm{TV}}(\mathcal{P}_1, \mathcal{P}_2) \le o_n(1)$, then the Gaussian planted distribution satisfies*

$$d_{\mathrm{TV}}(\mathcal{N}(\mu \sim \mathcal{P}_1, I_n), \mathcal{N}(\mu \sim \mathcal{P}_2, I_n))) \le o_n(1).$$

Proof. This comes from a standard bound on convolution density. Let $f(x) = \frac{1}{(2\pi)^{n/2}} \exp(-\frac{1}{2}\|x\|^2)$ the Gaussian PDF, then we have:

$$\begin{aligned} 2d_{\mathrm{TV}}(\mathcal{N}(\mu \sim \mathcal{P}_1, I_n), \mathcal{N}(\mu \sim \mathcal{P}_2, I_n))) &= \left| \int_X \int_\mu |p_1(\mu) - p_2(\mu)| f(X - \mu) d\mu dX \right| \\ &\le \left| \int_\mu |p_1(\mu) - p_2(\mu)| d\mu \right| \cdot \left| \int_X |f(X)| dX \right| \\ &= 2\|p_1 - p_2\|_1 \end{aligned}$$

by Young's inequality where p_1, p_2 are the densities of $\mathcal{P}_1, \mathcal{P}_2$ and $\|p_1 - p_2\|_1 = d_{\mathrm{TV}}(\mathcal{P}_1, \mathcal{P}_2)$. Thus the above inequality holds true. □

Note that during a reduction we can disregard the latent distribution on the target output because a permutation on vertices always returns the correct uniform over all latent communities. Therefore, we only care about the size of the output community, which we want to show is distributed approximately binomial. Formally, the reduction is as below:

Theorem 6.8 (Reduction from $k-$ PDS to binomial PDS)**.** *Let N, k_0 be parameters of planted clique graph size, $r < N/k_0$ be a target output for ratio of planted set. We present the following reduction ϕ with absolute constant $C > 1$:*

- Initial k-PDS Parameters: *vertex count N, subgraph size $k_0 \in o_N(N)$ dividing N, edge probabilities $0 < q < p \leq 1$ with $\min\{q, 1-q, p-q\} = \Omega(1)$, and a partition E of $[N]$. We further assume that $k_0 \in o(\sqrt{N})$ holds (otherwise detection for the PDS problem will be easy by total edge thresholding).*
- Target PDS parameters: *(n, r, k) where $r \in o_n(\sqrt{n})$ is the specified parameter and k is the expected subgraph size $k = n/r$ and n is the smallest multiple of $k_0 r$ that is greater than $(1 + \frac{p}{Q})N$ where*
$$Q = 1 - \sqrt{(1-p)(1-q)} + 1_{p=1}(\sqrt{q} - 1)$$
is the cloned signal strength from pre-processing.
- Target Binomial PDS edge strength:
$$\gamma = \mu \cdot \frac{k_0^2 r^3}{n^2(r-1)}, \qquad P_1 = \Phi(\gamma), \qquad P_2 = \frac{1}{2},$$
where $\mu \in (0,1)$ satisfies that
$$\mu \leq \frac{1}{12C\sqrt{\log(N) + \log(p-Q)^{-1}}} \cdot \min\{\log(\frac{p}{Q}), \log(\frac{1-Q}{1-p})\}.$$
Where γ here denotes the signal strength which is $\gamma = \Theta(\frac{(p-q)^2}{q(1-q)})$ and roughly the KL-divergence between two Bernoullis.
- *Applying ϕ on the given input graph instance G yields the following (when $k_0 \in o(\sqrt{N})$):*
$$d_{\text{TV}}(\phi(G(N, \frac{1}{2})), G(n, \frac{1}{2})) = o_n(1)$$
$$d_{\text{TV}}(\phi(PC_\rho(N, k_0, \frac{1}{2})), \text{PDS}_B(n, k, P_1, P_2)) = o_n(1)$$

We note that the bulk of the proofs will be similar to those in Theorem 3.1, specifically, Lemma 17 and Lemma 19 follows (almost) exactly. Therefore, we only need two extra results on matrix sampling and Bernoulli rotation.

Lemma 26 (Uniform sample of design matrix)**.** *Let $S \in (\mathbb{R}^{(n/k_0)\times(n/k_0)})^{\otimes k_0}$ be the k_0 design matrices sampled in step 2, Let $\mathcal{R}$ be the distribution on $(\mathbb{R}^{(n/k_0)\times(n/k_0)})^{\otimes k_0}$ where the k_0 matrices are sampled as S without rejection (exact binomial), then:*
$$d_{\text{TV}}(\mathcal{L}(S), \mathcal{R}) = 1 - o(n^{-2})$$
for some constant α.

Algorithm From k-PDS to Binomial PDS:

Inputs: Graph G of size N, subgraph parameter k_0 dividing N, edge density $q < p \in (0, 1]$ and a partition E of $[N]$ to k_0 equal parts $E_1, E_2, \dots, E_t$. Target planted ratio $r = \epsilon^{-1}$.

Steps :

1. *To-bipartite and planted diagonal:* Compute $F \in \{0, 1\}^{n \times n}$ with partition S of $[n]$ as To-k-Partite-Submatrix (Lemma 17) applied with initial dimension N, partition E, edge probabilities p and q and target dimension n.
2. *Design Matrix Sampling:* Sample independently k_0 random graphs and generate the corresponding matrices $R^{(i)}$ in the manner of Lemma 24. Reject if one of the samples does not satisfy the norm bound condition, repeat until succeeded.
3. *Flattened Bernoulli Rotations:* Let S be a partition of $[n]$ into k_0 equal parts $S_1, S_2, \dots, S_{k_0}$ obtained from the previous part. Construct output matrix M:
   (a) For i, j in $\{1, 2, \dots, k_0\}$, flatten matrix F_{S_i,S_j} to a $(n/k_0)^2$ size vector.
   (b) Apply Bernoulli Rotation on this vector with design matrix $\frac{k_0}{Cn\epsilon}(L^{(i,j)}_{n/k_0,1/r})$, Bernoulli parameter strengths $Q < p \le 1$, output dimensions $v_{ij} \in \mathbb{R}^{(n/k_0)^2}$.
   (c) Layout vector $v_{ij} \in \mathbb{R}^{(n/k_0)^2}$ to $(n/k_0) \times (n/k_0)$ matrix in the order from part (a).
   (d) Assign submatrix M_{S_i,S_j} to be the values of flattened v. Permute the output.
4. *Thresholding:* Given matrix M from the previous step, construct $G' = \phi(G)$ such that: for distinct indices $i < j$, $e_{ij} \in E(G')$ if and only if $M_{ij} \ge 0$.

Figure 5: Reduction from k-PDS to Binomial PDS. Note that one key distinction from previous reductions is that here we require the strong fact that sampling design matrices is almost always constraint-satisfying to ensure the output distribution is indeed Binomial and it has to be performed per-reduction, whereas in the previous reductions only the existence of one satisfying matrix is needed and can be done *a priori*.

Proof. Consider sampling k_0 matrices according to $\mathcal{R}$, then Lemma 23 and the union bound over the samples imply that at least one of them is rejected happens with probability at most $k_0(\frac{n}{k_0})^{-\alpha}$. Note that in the regime $k_0 \le n^{1/2}$, any $\alpha > 5$ suffices to make this probability $o(n^{-2})$.

Therefore, the probability that a sample from $\mathcal{R}$ is rejected is at most $o(n^{-2})$. This means that the accepted samples (with bounded column sum and operator norm) distribute according to a conditioning on $\mathcal{R}$, and thus the d_{TV} between them at $\mathcal{R}$ is at most $(1-p)^{-1} - 1 \in o(n^{-2})$ (where p is the probability of the condition not satisfied for one of the k_0 samples). □

As a corollary, we know that fix index i, the distribution of the degree in the sample graph of the ith vertex is distributed like $\text{Binom}(n, r^{-1})$ with total variation difference at most $o(n^{-2})$ from the Data Processing Inequality. Therefore, for any mixture of planted location in the PC instance, the distribution of output is close to binomial (Lemma 7). Therefore, we can now apply Thresholding and derive the formal argument for Bern Rotation:

Lemma 27 (Bernoulli Rotation for PDS_B)**.** *Consider the third step $\mathcal{A}_2$ in the above algorithm, whose input follows the previous step given in Lemma 17: suppose S is a partition of $[n]$ to k_0 equal parts and planted set $|T \cap S_i| = 1$ for all i. Denote $t_i := S_i \bigcap T$ and M_i a bijection between S_i and $[n/k_0]$. Then the following holds:*

$$d_{\mathrm{TV}}(\mathcal{A}_2(\mathrm{Bern}(Q)^{\otimes n\times n}),\ \mathcal{N}(0,1)^{\otimes n\times n}) = O(n^{-1})$$
$$d_{\mathrm{TV}}(\mathcal{A}_2(\mathcal{M}_{[n]\times[n]}(\mathcal{U}_n(S), p, Q)),\ \mathcal{L}(\gamma \cdot \mathbb{1}_{A\times A} + \mathcal{N}(0,1)^{\otimes n\times n})) = O(n^{-1})$$

where $A \sim \mathrm{Bern}(r^{-1})^{\otimes n}$ is the binomially planted set.

Proof. As in Lemma 18, we first shift all Bernoullis to $\mathcal{N}(0,1)$ and $\mathcal{N}(0,\mu)$ and apply Lemma 3.

The first equation is trivial by the canonical Bernoulli rotation, thus we are only concerned about the second equation (what H_1 gets mapped to). For the second half, the distribution shifting remains the same as Lemma 18, so we only need to analyze the output with our sampled matrix.

Conditioning on any fixed (set of) Bernoulli rotations in submatrix $F_{S_i\times S_j}$ with design matrices $\frac{k_0}{Cnr}L_{ij} = (\sqrt{\frac{k_0}{Cnr}}B_i + \frac{\sqrt{\gamma}}{r}) \otimes (\sqrt{\frac{k_0}{Cnr}}B_j + \frac{\sqrt{\gamma}}{r})$, the output inside submatrix $M_{S_i\times S_k}$ is exactly the rank-1 product of the $M_i(t_i)$th row of $\sqrt{\frac{k_0}{Cnr}}B_i + \frac{\sqrt{\gamma}}{r}$ (transposed) times $M_j(t_j)$th row of $\sqrt{\frac{k_0}{Cnr}}B_j + \frac{\sqrt{\gamma}}{r}$. Assume that in $\sqrt{\frac{k_0}{Cnr}}B_i + \frac{\sqrt{\gamma}}{r}$ the non-zero entries of $M_i(t)$th row come from a set index set $N_t^{(i)}$ parameterized by t, i, then the output is exactly:

$$M_{S_i\times S_j} \sim \gamma \cdot \mathbb{1}_{N_{t_i}^{(i)}\times N_{t_j}^{(j)}} + \mathcal{N}(0,1)^{\otimes (n/k_0)\times(n/k_0)}.$$

Now we turn to the marginal distribution of output M without conditioning on $L_{i,j}$. We see that as argued in Lemma 26, for any prior on T, the marginal distribution of $N_{t_j}^{(j)}$ is at most $o(n^{-2})$ close in total variation to $\mathrm{Bern}(r^{-1})^{\otimes n/k_0}$. By Lemma 8, the total variation distance between the product of $N_t^{(\cdot)}$ and $\mathrm{Bern}(r^{-1})^{\otimes n}$ is at most $o(n^{-2}k_0) = o(n^{-1})$.

Finally, note that while our output from Bern-rotation has law

$$\mathcal{L}(M) = \mathcal{L}(\gamma \cdot \mathbb{1}_{N\times N} + \mathcal{N}(0,1)^{\otimes n\times n})$$

for the unknown prior of T and the cropped set of design matrices L where $N = \bigcup_i N_{t_i}^{(i)}$. Since we have that $d_{\mathrm{TV}}(N, \mathrm{Bern}(r^{-1})^{\otimes n}) = o(n^{-1})$, by data processing inequality we have

$$d_{\mathrm{TV}}(\mathcal{M}, \mathcal{L}_{A\sim\mathrm{Bern}(r^{-1})^{\otimes n}}(\gamma \cdot \mathbb{1}_{A\times A} + \mathcal{N}(0,1)^{\otimes n\times n}))$$

as well. Finally, the lemma can be proven via the bounding the loss incurred by distribution shifting to be $o(n^{-1})$ (for instance see Lemma 18) so that $d_{\mathrm{TV}}(\mathcal{A}_2(\cdot), \mathcal{L}(M)) = o(n^{-1})$ as well. □

Proof of Theorem 6.8. Define the steps of $\mathcal{A}$ to map inputs to outputs as follows

$$(G, E) \xrightarrow{\mathcal{A}_1, \epsilon_1} (F, S) \xrightarrow{\mathcal{A}_2, \epsilon_2} M \xrightarrow{\mathcal{A}_3, \epsilon_3} G'$$

where the following ϵ_i denotes the total variation difference in each step (from output of $\mathcal{A}$ to the next target). Under H_1, consider the following sequence of distributions:

$$\mathcal{P}_0 = G_E(N, k, p, q)$$
$$\mathcal{P}_1 = \mathcal{M}_{[n]\times[n]}(S \times S, \mathrm{Bern}(p), \mathrm{Bern}(Q)) \quad \text{where } S \sim \mathcal{U}_n(F)$$

$$\begin{aligned}\mathcal{P}_2 &= \gamma \cdot \mathbb{1}_S \otimes \mathbb{1}_S + \mathcal{N}(0,1)^{\otimes n \times n} \quad \text{where } S \sim \text{Bern}(r^{-1})^{\otimes n} \\ \mathcal{P}_4 &= \text{PDS}_B(n, k, P_1, 1/2)\end{aligned}$$

Applying Lemma 17 before, we can take

$$\epsilon_1 = 4k_0 \cdot \exp\left(-\frac{Q^2 N^2}{48pk_0 n}\right) + \sqrt{\frac{C_Q k_0^2}{2n}}$$

where $C_Q = \max\left\{\frac{Q}{1-Q}, \frac{1-Q}{Q}\right\}$. For ϵ_2, Lemma 27 guarantees that $\epsilon_2 = O(n^{-1})$ suffices. The final step $\mathcal{A}_3$ is exact and we can take $\epsilon_3 = 0$. Finally, note that from the data processing inequality applied to d_{TV} that $d_{\text{TV}}(\mathcal{A}_i(\cdot), \mathcal{A}_i(\cdot')) \leq d_{\text{TV}}(\cdot, \cdot')$ so each step the total variation loss at most accumulates (Lemma 9), thus by the triangle inequality on TV we get

$$d_{\text{TV}}(\mathcal{A}(G_E(N, k, p, q)), \text{PDS}_B(n, k, P_1, 1/2)) \leq \epsilon_1 + \epsilon_2 = o(1).$$

Under H_0, consider the distributions

$$\begin{aligned}\mathcal{P}_0 &= G(N, q) \\ \mathcal{P}_1 &= \text{Bern}(Q)^{\otimes n \times n} \\ \mathcal{P}_3 &= \mathcal{N}(0,1)^{\otimes n \times n} \\ \mathcal{P}_4 &= G(n, 1/2)\end{aligned}$$

As above, Lemmas Lemma 17 and Lemma 27 imply that we can take

$$\epsilon_1 = 4k_0 \cdot \exp\left(-\frac{Q^2 N^2}{48pk_0 n}\right), \quad \epsilon_2 = O(n^{-1}), \quad \text{and} \quad \epsilon_3 = 0$$

Again by the data processing inequality (Lemma 9), we therefore have that

$$d_{\text{TV}}\left(\mathcal{A}\left(G(N, q)\right), G(n, 1/2)\right) = O(\epsilon_1 + \epsilon_2) = o(1)$$

which completes the proof of the theorem. □

We omit the corollary with the lower bound of deciding PDS_B, as it has been studied extensively in various literature and well-established that those PDS and PDS_B incur the same computation boundary. Specifically, this simply require an extra post-processing step as in Theorem 3.2.

6.8 Detection-Recovery gaps in other problems

In this section we finish our discussions on two other problems that observe a detection-recovery gap from reduction to a detection-recovery gap in PDS. In [BBH18], such relations were considered assuming the PDS recovery conjecture, here we do so at a lower rate of signal from only assuming PC conjecture and Theorem 3.3. Denote the H_1 hypothesis distributions in Section 5.2 as $\text{BC}(n, k, \mu)$ and $\text{BSPCA}(m = n, k, d, \theta)$, respectively.

6.8.1 Detection-Recovery gap in Biclustering

This follows from a canonical process of simply performing To-k-Partite-Submatrix and Gaussianizing. This gives us a symmetric planted Gaussian principal submatrix with elevated mean. Lastly, we can permute the columns if needed.

Lemma 28 (Reduction to Bi-clustering – Lemma 6.7 in [BBH18])**.** *Suppose that n, μ and $\rho \geq n^{-1}$ are such that*

$$\mu = \frac{\log(1+2\rho)}{2\sqrt{6\log n + 2\log 2}} > \frac{\rho}{4\sqrt{6\log n + 2\log 2}}$$

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

$$d_{\mathrm{TV}}\left(\phi\left(\mathrm{PDS}(n, S, 1/2+\rho, 1/2)\right), \mathbb{E}_{T\sim \mathit{Unif}_n(k)}\mathcal{L}\left(\mu\cdot \mathbb{1}_S\mathbb{1}_T^\top + \mathcal{N}(0,1)^{\otimes n\times n}\right)\right) = O\left(\frac{1}{\sqrt{\log n}}\right).$$

With this, we can now state the lower bound for recovery and refutation in BC by Lemma 14:

Corollary 6.9 (Recovery Hardness for Bi-Clustering)**.** *Let $\alpha > 0$ and $\beta \in (0,1)$, then there exists such parameters (N_n, K_n, μ_n) such that: (assuming the PC_ρ detection hypothesis)*

1. *The parameters are in the regime:*

$$\lim_{n\to\infty}\frac{\log K_n}{\log N_n} \leq \beta, \quad \lim_{n\to\infty}\frac{\log \mu_n}{\log N_n} \leq -\alpha.$$

2. *If $\beta < \frac{1}{2}+\frac{2}{3}\alpha$, then there is no (randomized) polynomial-time recovery blackbox $\mathcal{A}_n : \mathbb{R}^{N_n\times N_n} \to \binom{N_n}{K_n}^2 =: (\widehat{S}, \widehat{T})$ such that $|\widehat{S}\bigcap S| + |\widehat{T}\bigcap T| - 2K_n \in o(K_n)$ with probability greater than 0 asymptotically with $\mathcal{A}$ applied over the distribution on the Bi-clustering instance conditioning on S, T and the uniform prior distribution $S \perp\!\!\!\perp T \sim \mathit{Unif}_n(k)$.*

3. *If $\beta < \frac{1}{2}+\alpha$, then there is no polynomial-time refutation blackbox $\mathcal{A}_n : \mathbb{R}^{N_n\times N_n} \to \{0,1\}$ such that $\mathcal{A}$ returns 0 with asymptotically positive probability applied on $\mathcal{N}(0,1)^{\otimes n\times n}$ and returns $\mathcal{A}(M) = 1$ if there is a $k\times k$ submatrix S in M with mean at least $k^2\mu$.*

We finally comment that the detection boundary is $\mu \in \widetilde{\omega}(n/k^2)$ from the same reduction and hence the detection problem is computationally easy when $\beta > \frac{1}{2} + \frac{1}{2}\alpha$.

6.8.2 Detection-Recovery gap in BSPCA

Theorem 6.10 (Recovery Hardness in BSPCA)**.** *Let $\alpha \in \mathbb{R}$ and $\beta \in (0,1)$. There exists a sequence $\{(N_n, K_n, D_n, \theta_n)\}_{n\in\mathbb{N}}$ of parameters such that: (assuming the PC_ρ detection hypothesis)*

1. *The parameters are in the regime*

$$\lim_{n\to\infty}\frac{\log \theta_n}{\log N_n} \leq -\alpha, \quad \lim_{n\to\infty}\frac{\log K_n}{\log N_n} \leq \beta$$

2. *If $\alpha > \beta - \frac{1}{2} > 0$, then there is no randomized polynomial-time recovery blackbox $\phi_n : \mathbb{R}^{D_n\times N_n} \to \binom{[N_n]}{k}^2$ such that the probability that ϕ_n recovers exactly the pair of latent row and column supports of an instance from $\mathrm{BSPCA}(N_n, K_n, D_n, \theta_n)$ is greater than 0 asymptotically, where the supports are independently distributed from the uniform prior.*

Proof. The proof follows from the following lemma:

Lemma 29 (Random Rotation – Lemma 8.7 in [BBH18])**.** *Let $\tau : \mathbb{N} \to \mathbb{N}$ be an arbitrary function with $\tau(n) \to \infty$ as $n \to \infty$. There exists map $\phi : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ that sends $\phi(\mathcal{N}(0,1)^{\otimes m\times n}) \sim \mathcal{N}(0,1)^{\otimes m\times n}$ and for any unit vectors $u \in \mathbb{R}^m, v \in \mathbb{R}^n$ we have that*

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

We defer the proof to [BBH18] and focus on the reduction forward. Note that the left-hand side can be viewed as the asymmetric biclustering distribution, thus combining Lemma 28 and Lemma 29 with Lemma 9 we get a polynomial-time map $\mathcal{A}$ such that:

$$d_{\text{TV}}(\mathcal{A}(\text{PDS}(n, u, \frac{1}{2}+\rho, \frac{1}{2})), \mathcal{N}(I_n + \frac{\mu^2}{\tau n}uu^T)) = o(1).$$

Now the only thing left is to define precise parameter correspondence to apply Theorem 3.3. Consider the following set of parameters (let $\gamma := \beta - \frac{1-\alpha}{2}$):

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

Observe that because $\rho \to 0$, $\log(1+2\rho) \in \Theta(\rho)$ and thus $\mu_n \in \widetilde{\Theta}(\rho_n)$, one can easily verify that the conditions are equivalent to:

$$\lim_{n\to\infty} \frac{\log(K_n^3\rho_n^2)}{\log(N_n)} = 1 - \alpha + \beta < 1.5$$

thus we can apply Theorem 3.3, which concludes that no polynomial black-box can successfully recover the planted instance u here. □

6.9 Limitations of Gaussians

An important missing case of the above work is when the edge densities goes to zero as $n \to \infty$, or, in other words, the edge density $p = cq = n^{-\alpha}$ or $p = n^{-\beta} + n^{-\alpha}, q = n^{-\alpha}$. Specifically, this requires distribution shifting in the sparse regime (so thresholding Gaussians no longer works). In [BBH18], reduction technique via Poisson variables are considered (since they resemble the mean and variance of sparse Bernoulli bits), but it is hard to consider manipulating Poisson random variables in a rotation. Specifically, we have the pessimistic lemma below:

Proposition 6.11 (Signal loss from dense to sparse)**.** *There does not exist a map M that sends $X_1 = \text{Bern}(1/2)$ and $X_2 = \text{Bern}((1+\mu_n)/2)$ to $Y_1 = \text{Bern}(a_n)$ and $Y_2 = \text{Bern}(b_n)$ where $\mu_n, a_n, b_n \to \widetilde{o}_n(1), a_n/b_n \in \Theta(1)$, and the signal loss $D_{KL}(X_1\|X_2) = \widetilde{\Theta}(D_{KL}(Y_1\|Y_2))$.*

Proof. Because our image is in $\{0,1\}$, we may assume that a map $M : 0 \to \text{Bern}(p)$ and $1 \to \text{Bern}(q)$, then we have:

$$M(\text{Bern}(\frac{1}{2})) \to \text{Bern}(\frac{p+q}{2})$$
$$M(\text{Bern}(\frac{1+\mu}{2})) \to \text{Bern}(\frac{p+q}{2} + \mu\frac{q-p}{2}).$$

Note the desired target regimes, one would need $p, q \in \widetilde{o}(1)$, $p/q \in \widetilde{\Theta}(1)$.

Hence the KL divergence (by Lemma 10) is approximately (ignoring log factors):

$$\mu^2 \sim D_{KL}(\frac{1}{2} \| \frac{1+\mu}{2}) \sim D_{KL}(M(X_1) \| M(X_2)) \sim \frac{\mu^2(p-q)^2}{q(1-q)}$$

implying that $D_{KL}(p\|q) \sim \frac{(p-q)^2}{q(1-q)} \in \widetilde{\Omega}(1)$. However, this cannot happen as $p - q \in \widetilde{\Theta}(q)$ and $q(1-q) \in \widetilde{\Theta}(q)$, leading to a contradiction! □

The above statement has the following implications. Firstly, it means that one cannot use Gaussian shifts because the mappings between Gaussian and dense Bernoullis are bi-directional (one can turn around with rejection sampling and thresholding without losing signal). Secondly, it implies that one cannot handle the problem in its dense case and hope that a final post-processing procedure can reach the sparse target.

One alternate to get around the above problem is through producing *multiple bits* $\text{Bern}(q)^{\otimes T}$ from a single Bernoulli or Gaussian. However, the data processing inequality asserts that the KL divergence of the product bit scales inversely with T, meaning that if we are to multiple n, k (and thus k^2/n) by T, the KL divergence would scale down by T^2, which leads to a wrong *slope* of parameters for our bounds in this paper. The sparse case regime remains of interest and open for future work.