Proceedings of Machine Learning Research vol 195:1–40, 2023 36th Annual Conference on Learning Theory

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

Guy Bresler GUY@MIT.EDU
77 Massachusetts Ave., Cambridge, MA, USA

Tianze Jiang TJIANG@MIT.EDU
77 Massachusetts Ave., Cambridge, MA, USA

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

Planted Dense Subgraph (PDS) problem is a prototypical problem with a computational-statistical gap. It also exhibits an intriguing additional phenomenon: different tasks, such as detection or recovery, appear to have different computational limits. A *detection-recovery gap* for PDS was substantiated in the form of a precise conjecture given by Chen and Xu (2014) (based on the parameter values for which a convexified MLE succeeds), and then shown to hold for low-degree polynomial algorithms by Schramm and Wein (2022) and for MCMC algorithms for Ben Arous et al. (2020).

In this paper we demonstrate that a slight variation of the Planted Clique Hypothesis with *secret leakage* (introduced in Brennan and Bresler (2020)), implies a detection-recovery gap for PDS. In the same vein, we also obtain a sharp lower bound for refutation, yielding a detection-refutation gap. Our methods build on the framework of Brennan and Bresler (2020) to construct average-case reductions mapping secret leakage Planted Clique to appropriate target problems.

Keywords: Average-case Complexity, Planted Clique, Algorithmic Hardness

1. Introduction

The last decade has witnessed a dramatic shift in our understanding of the fundamental limits of high-dimensional statistics problems. Rather than the *statistical limit* being the most relevant quantity governing the minimum amount of data or signal strength needed to solve a problem, it has emerged that for many problems of central importance there is a distinct and often much larger *computational limit* at which computationally efficient algorithms begin to succeed. Berthet and Rigollet (2013) showed how a *statistical-computational gap* for a binary variant of sparse PCA follows via reduction from the planted clique hardness conjecture (Conjecture 1), spurring intense research activity (see, e.g., Brennan and Bresler (2020) and references therein).

In this paper we investigate how the computational complexity of different tasks, including detection, recovery, and refutation, can vary even for the same statistical model. The phenomena of interest are exemplified by the Planted Dense Subgraph (PDS) problem, defined next.

Planted Dense Subgraph (PDS). A sample from the distribution $\mathsf{PDS}(n, k, p, q)$ is obtained by:

1. Sample $G \sim G(n, q)$ an Erdős-Renyi graph with edge density q.
2. Select a subset S of vertices uniformly among the $\binom{n}{k}$ subsets of size k.
3. Re-sample edges with both endpoints in S independently, including each with probability $p > q$.

The *detection* (or decision) problem is to decide, given a graph G, between the two hypotheses

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

The *recovery* problem is to (exactly) find the planted support S. (Weaker notions of recovery can be found in Section 1.4.)

The special case of PDS where $p = 1$ is known as the Planted Clique (PC) problem. Let $G(n, k, p) = \mathsf{PDS}(n, k, 1, p)$. We denote by $\mathsf{PC}_D(n, k, p)$ the problem of deciding between

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

Both detection and recovery have efficient (polynomial-time) algorithms whenever $k = \Omega(\sqrt{n})$ (Alon et al. (1998)), but a growing body of evidence (Barak et al. (2019); Feldman et al. (2017)) suggests that these problems become hard for clique size $k = n^\beta$ with $\beta < 1/2$.

Conjecture 1 (PC Conjecture) *Fix 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 $\limsup_{n\to\infty} \log_n k_n < \frac{1}{2}$. If G is an instance of $\mathrm{PC}_D(n, k, p)$, then*

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

In our work, we will use a (stronger) variation of this assumption proposed by Brennan and Bresler (2020) where some structural information of the planted clique is assumed (the *secret leakage*). See Conjecture 4 and the associated discussion.

1.1. Computational feasibility of PDS

PDS Detection. Feasibility of detection in PDS is described by a *phase diagram* (see Fig. 1) indicating for each possible parameter choice whether the problem is: (1) information-theoretically impossible, (2) solvable in principle but computationally hard, or (3) solvable in polynomial time. Complete phase diagrams were shown by reduction from PC_D in the regime $q = \Theta(1)$ by Ma and Wu (2015)[1], for the sparse regime $p = cq$ for constant c and $q = 1/\mathrm{poly}(n)$ by Hajek et al. (2015a), and by Brennan et al. (2018) for a general regime interpolating between the two. Despite the similarity between PC_D and PDS_D, it is non-trivial to construct reductions that are tight against algorithms, since PDS_D exhibits a trade-off between subgraph size and signal strength.

In all of the above parameter regimes, whenever $k = \omega(\sqrt{n})$ the optimal polynomial-time test T_{sum} simply compares the total number of edges to a threshold. A second moment calculation shows that

$$T_{\mathrm{sum}} \text{ succeeds w.h.p. if } \quad \frac{k^4(p-q)^2}{n^2 q(1-q)} = \omega(1)\,.$$

By its nature, success of the sum statistic yields no information whatsoever about the location of the planted dense subgraph. What can be said about recovery?

1. In the regime $q = \Theta(1)$, PDS is easily seen to be computationally equivalent up to log factors in the parameter values to the Gaussian matrix model with corresponding means.

PDS Recovery. The best currently-known algorithms (such as spectral, semi-definite programming, and low-degree polynomials) for *recovery* turn out to require a dramatically higher signal strength (Chen and Xu (2014); Hajek et al. (2016)). The following conjecture posits that this signal strength is optimal for the recovery problem (Chen and Xu (2014); Hajek et al. (2015a)).

Conjecture 2 (PDS recovery conjecture) *Suppose* $G_n \sim \mathsf{PDS}(n, k_n, p_n, q_n)$. *If* $k = \omega(\sqrt{n})$ *and*

$$\limsup \log_n \frac{k^2(p-q)^2}{nq(1-q)} < 0,$$

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

The lower bound in this conjecture has been shown for restricted classes of algorithms: in Schramm and Wein (2022) for low-degree polynomials and Ben Arous et al. (2020) for Markov Chain Monte Carlo algorithms.

Recovery lower bound via reduction? Lower bounds have been shown for a wide variety of detection problems via reduction from PC, and for the majority of these problems recovery is algorithmically feasible in the same parameter regime (to within a constant factor) in which detection is algorithmically feasible. Yet for problems where recovery seems strictly harder than detection, demonstrating a detection-recovery gap via reduction from PC has remained elusive. Attempts in this direction include those of Cai et al. (2017) showing hardness for a matrix model with highly correlated entries (different from the independent edges in PDS^*), and Brennan and Bresler (2020) showed that the conjectured recovery lower bound follows from the PC conjecture for a *semirandom* variant of PDS where an adversary may "helpfully" remove edges outside of the dense subgraph. The main question motivating our work is:

Can a detection-recovery gap be shown for Planted Dense Subgraph via reduction?

A first conceptual challenge is that, as shown by Alon et al. (2007), detection and recovery for PC are *equivalent*. What this means is that the detection-recovery gap appearing in PDS is inherent to PDS, and in particular, we cannot simply map from PC detection and PC recovery separately.

In fact, our reductions will still map to detection problems (with implications to recovery). But we cannot simply map to the PDS detection hypotheses PDS_D: Otherwise, we would be mapping a conjecturally hard instance of PC to an easy instance of PDS_D! Our goal in this paper is considerably more modest than to refute the planted clique conjecture, so we must find another way.

1.2. Contributions

In this work, we will utilize the insight that by constructing different statistical models with similar underlying properties, tailored to corresponding inference tasks, we can go beyond the simple detection boundary to prove tighter results. Our main contributions are:

- We present the first reduction-based evidence of a computational *detection-recovery gap* (Corollary 11) for recovering the hidden community in planted dense subgraph, via an average-case reduction from Planted Clique with secret leakage (Conjecture 4).

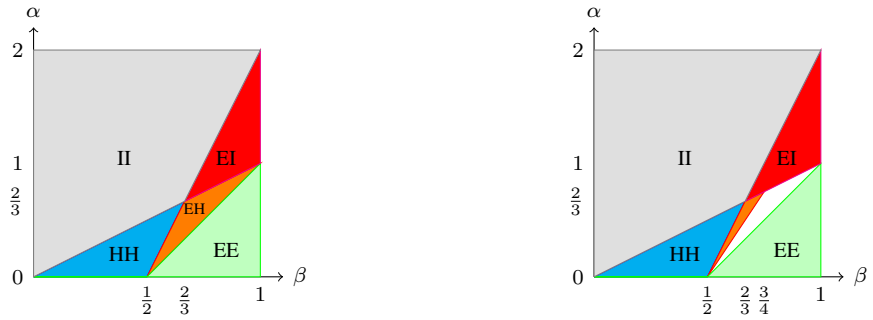

Figure 1: The pictures above (left: Detection vs Refutation; right: Detection vs Recovery) concerns $\mathsf{PDS}(n, k, p, q)$ when p, q are bounded away from 0 and 1, and $k \in \widetilde{\Theta}(n^{\beta})$, $D_{KL}(p\|q) \in \widetilde{\Theta}(n^{-\alpha})$, where E denotes easy, H (computationally) hard, and I (statistically) intractable, hence the orange EH (computationally easy to detect but hard to refute/recover) is our main results. 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 close for orange and leave white open.

- We show how detection hardness for the two-community Imbalanced Stochastic Block Model (ISBM), shown by reduction from Conjecture 4 by Brennan and Bresler (2020), can be used to obtain a log-optimal lower bound on refuting dense k-subgraphs in $G(n, p)$ and Gaussian principal submatrix with large mean. This matches the algorithm-specific results of Barak et al. (2019) and Jones et al. (2021) and shows a reduction-based *detection-refutation gap*.

- Combining our results with existing reductions yields analogous results also for other average-case planted models such as Gaussian biclustering and biased sparse PCA. This yields detection-recovery gaps for these problems and answers a question from Brennan et al. (2018).

- Finally, we also give insight into the relationships between the statistical boundaries for the above problems, including showing a nearly sharp limit on refuting densest k-subgraphs in Erdős Rényi Graphs via a novel reduction from recovery.

1.3. Reductions and Other Evidence for Hardness

Average-case Reductions. We will define an (average-case) reduction (in total variation) from two source distributions P_0, P_1 to target pair Q_0, Q_1 as a (random) polynomial-time computable map Φ such that the pushforward $d_{\mathrm{TV}}(\Phi(P_i), Q_i) = o(1)$ for $i = 1, 2$. The implication is that if P_0, P_1 are computationally hard to distinguish, then the same holds true for Q_0 versus Q_1: any poly-time algorithm $\mathcal{A}$ for the latter task would yield a poly-time algorithm $\mathcal{A} \circ \Phi$ for the former by composing with the reduction, contradicting the presumed hardness of P_0 versus P_1.

While reductions form the bread and butter of complexity theory, there is a general sentiment in the community that average-case reductions are notoriously delicate. Such reductions must not only map to a valid problem instance, they must precisely map entire probability distributions.

The upside is that reductions can give the strongest possible evidence for computational hardness, and moreover, they demonstrate a connection between two formerly disparate problems which is often of interest independent of hardness. We refer to Brennan and Bresler (2020) for a review of the reductions literature.

Algorithm-specific hardness There have been numerous results showing lower bounds for classes of algorithms and we mention a few of the results that relate to PDS. In Barak et al. (2019), a lower bound for refutation of large cliques in $G(n, \frac{1}{2})$ was shown for the Sum-of-Squares hierarchy. Schramm and Wein (2022); Rush et al. (2022) sharply characterize the power of low-degree polynomials for recovery in PDS. The overlap gap framework, introduced in Gamarnik and Sudan (2014), connects algorithmic infeasibility with properties of the solution space geometry (see also Gamarnik (2021); Gamarnik and Zadik (2019)). Other relevant results include that of Feldman et al. (2017) on the statistical query model analyzed in the case of a bipartite "samples" version of PC and Brennan et al. (2021) relating the power of low-degree polynomials and the statistical query model.

1.4. Inference tasks beyond decision

Denote by H_0 the null hypothesis (usually an Erdős-Rényi Graph), and H_1 is a graph with a planted structure with support $v \in \{0,1\}^n$. Consider a valuation function val on graphs such that: $\mathbb{P}_{H_0}(\mathsf{val}(G) < \delta - \epsilon)$ and $\mathbb{P}_{H_1}(\mathsf{val}(G) > \delta + \epsilon)$ are both $1 - o_n(1)$. In the case of PDS, val is the densest-k-subgraph density. Consider the following:

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

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

Recovery Let π be a distribution over size k planted supports $v \in \{0,1\}^n$, and for each v let P_v be a distribution over planted graphs. Let $G \sim P = \mathbb{E}_{v\sim\pi} P_v$. A recovery blackbox $\mathcal{A} : G \to \{0,1\}^n$ is said to achieve

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

In most models we consider, these variants of recovery only differ in sub-polynomial factors (via reduction in Appendix C). We further remark that a refutation algorithm is only evaluated on the input distribution H_0, whereas a recovery algorithm is only evaluated on the distribution H_1. The latter fact was leveraged by Schramm and Wein (2022) and both will be crucial to our proofs.

Lemma 3 (Informal, see Lemma 22) *For any $\widetilde{H}_0$ that does not have a k-subgraph with density above $\frac{p+q}{2}$ with high probability, weak recovery oracles nontrivially distinguish $\widetilde{H}_0, H_1 = \mathsf{PDS}$.*

1.5. Planted Clique and Secret Leakage

We require a slight modification of the planted clique conjecture, proposed by Brennan and Bresler (2020): Instead of a uniformly located clique, the clique is sampled according to some distribution ρ over the $\binom{n}{k}$ possible clique positions. One may interpret this as a form of *secret leakage*, whereby some information about the clique position has been revealed to the algorithm.

The form of secret leakage we will use in our reductions is k-PC$_D(n, k, p)$, where there is some fixed (known) partition E of $[n]$ into k equally-sized subsets, and under H_1 the planted set

is obtained by selecting exactly one node uniformly from each part. We refer to the corresponding hardness assumption as the k-PC conjecture.

Conjecture 4 (k-PC Conjecture) *Fix 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-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 Brennan and Bresler (2020) for a general leakage PC conjecture and supporting evidence. When the amount of leaked information is small enough, both low-degree polynomials and statistical query algorithms succeed only above the same $\sqrt{n}$ clique size as in ordinary PC.

Remark 5 (Binomial planted set) *In the literature it is sometimes assumed that the planted set is of fixed size k, and other times it is of binomial size (where each node is planted with probability k/n independently). We use a fixed size k and note that all of our (hardness) results extend to corresponding binomial versions by virtue of closeness of the hypergeometric and binomial distributions in appropriate parameter regimes (which can be understood as an instance of a finite de Finetti type theorem Diaconis and Freedman (1980)). In particular, one may carry out a reduction by keeping a random $o(n)$ sized fraction of the nodes and discarding the rest.*

2. Reduction Techniques Overview

2.1. Selecting hypotheses

As discussed in Section 1.1, we cannot map to the standard two PDS hypotheses. A key insight from Section 1.4 is that while detection concerns both H_0 and H_1, all other tasks deal with only one of the two hypotheses. Specifically, for any pair of hypotheses with distributions satisfying the val criteria, recovery algorithms are only evaluated on an input distributed according to H_1 and not H_0. To this end, we are free to select qualifying "quiet" hypothesis $\widetilde{H}_0$ that is not Erdős-Renyi such that it has a harder decision task and imply stronger recovery lower bounds. Similarly, for refutation we may map to $\widetilde{H}_1$ that is different from the standard H_1.

Now, suppose that we want to map from the two hypotheses in PC to $\widetilde{H}_0, H_1$ in a target graph such that H_1 is PDS (so that a recovery blackbox enables us to test between $\widetilde{H}_0$ and H_1). We have the following naturally competing constraints:

1. For a recovery blackbox to achieve detection, $\mathsf{val}(G)|_{H_0}$ has to be small with high probability, suggesting the fact that $\widetilde{H}_0$ has to be *far* from H_1, with respect to some metric.

2. We need to construct a reduction. From a data-processing inequality perspective, this means that $\widetilde{H}_0$ has to be *closer* to H_1 than the distance between source hypotheses.

It turns out that for recovery, the correct $\widetilde{H}_0$ is extremely hard to find (Appendix B in Schramm and Wein (2022)), and even for good $\widetilde{H}_0$ candidates, constructing a tight reduction seems challenging. However, we will show that by changing H_0 to simply match the first-moment in H_1, one can achieve a $\widetilde{H}_0$ realizing a non-trivial gap between detection and recovery in PDS, while still being feasible for us to map to from PC. Changing H_0 as we do here was also analyzed for the case of low-degree polynomials by Schramm and Wein (2022). Note that Brennan and Bresler (2020) in their result on semirandom PDS modified H_1, rather than H_0, and we will use this same reduction to demonstrate a detection-refutation gap. We define the following models:

Mean-corrected Planted Dense Subgraph (PDS^*)**.** Consider PDS with H_0 modified to prevent success of the obvious first moment test. Consider edge strengths $q < p_0 < p$ and size k such that

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

and define $\mathsf{PDS}^*(n, k, p, q)$ as hypothesis testing between

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

Imbalanced Stochastic Block Model (ISBM). Consider a two-community Stochastic Block Model $\text{ISBM}(n, k, P_{11}, P_{12}, P_{22})$ 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 $P_{ij} = P_{ji}$. Moreover, we 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 (let $r = n/k$):

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

This model can be considered as a mean-field analogue of recovering a first community in a general balanced r-block SBM model (keeping one block while averaging out the rest). [2]

2.2. Signal transformation

We start our reduction by viewing our problem as a *planted bits* problem, which is simply a vector $v \sim \text{Bern}(q)^{\otimes n}$ with planted bits $v_I \sim \text{Bern}(p)$ at the index set $I \subseteq [n]$ with a different bias. Concretely, because of the one clique vertex per partition assumption of k-PC, each $\frac{n}{k} \times \frac{n}{k}$ block of the adjacency matrix has *a single* planted 1 entry. All of the reductions we consider can be viewed as mapping a set of planted bits to another desired target set of planted bits with a larger planted size and specific biases.

The difficulty at the core is thus the following: *how to transform the planted bits distribution with unknown location to a desired target distribution while not losing signal-to-noise ratio* (measured by the KL-divergence) between planted and null bits and the size of planted location I (Brennan et al. (2019)), so that the target instance remains at the threshold of algorithmic feasibility.

As in Brennan and Bresler (2020), we will use Gaussian distributions as intermediate steps in transforming from k-PC. While Bernoulli data are challenging to non-trivially transform without signal loss, we will leverage the nice behavior of Gaussians under linear maps, enabling us to carefully control the added noise within the transformation (as discussed in the next subsection).

To see the approximate equivalence between Gaussians and Bernoulli variables, we note that a Gaussian $\mathcal{N}(\mu, 1)$ can be readily mapped to $\text{Bern}(\Phi(\mu))$, where Φ is the Gaussian CDF, by thresholding at 0. If $\mu \ll 1$, the KL-divergence decreases only by a numerical constant factor independent of μ. In the other direction, a rejection sampling procedure can map a pair of Bernoulli variables to a pair of Gaussians with little information loss[3]:

2. Note that both models contain a dense subgraph (high val), and PDS^* is just a translated PDS.
3. This process introduces a log-factor, which is the (only) reason in later sections we ignore poly-log factors in rates.

Lemma 6 (Gaussian Rejection Kernels – Ma and Wu (2015); Brennan et al. (2018)) *Let R be a parameter and suppose that $0 < q < p \leq 1$, $\min(q, 1-q, p-q) = \Omega(1)$. Suppose that $\mu < \left(1 \wedge \frac{\delta}{2\sqrt{6\log R + 2\log(p-q)^{-1}}}\right)$ where $\delta = \min\left\{\log\left(\frac{p}{q}\right), \log\left(\frac{1-q}{1-p}\right)\right\}$, then there exist map $\text{RK}(\cdot)$ can be computed in poly(R) time such that the push-forward maps satisfy*

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

Now that we have a Gaussian signal with planted mean, we apply a *rotation* (treating the entire matrix as a vector). Specifically, in Brennan and Bresler (2020) the following process BERN-ROTATIONS was introduced to transform an instance of $\mathcal{N}(v, I_\ell)$ where $v \in \mathbb{R}^\ell$ contains signal.

1. We right-multiply the Gaussian vector by a *design matrix* $A \in \mathbb{R}^{\ell \times m}$, which yields $vA + \mathcal{N}(0, AA^T)$. Denote the square of the top-singular value of A to be $\lambda = \sigma^2(A)$.

2. On the re-scaled result vector $\mathcal{N}(\lambda^{-1/2}vA, AA^T/\lambda)$, we can add a Gaussian noise $\mathcal{N}(0, I - AA^T/\lambda)$ independent of μ to get exactly $N(\frac{vA}{\sigma(A)}, I_n)$, which has unit variance.

In short, *we transform signals as mean vectors of isotropic Gaussian distributions by rotating the space and paying an extra whitening noise to produce an isotropic distribution again.*

Lemma 7 (Dense Bernoulli Rotations – Lemma 8.1 in Brennan and Bresler (2020)) *Let m and ℓ be positive integers and let $A \in \mathbb{R}^{\ell \times m}$ be a matrix with singular values all at most $\lambda > 0$. Let R, $0 < q < p \leq 1$ and μ be as in Lemma 6. Let $\mathcal{A}$ denote* BERN-ROTATIONS *applied with rejection kernel parameter R, Bernoulli biases $0 < q < p \leq 1$, output dimension m, matrix A with singular value upper bound λ and mean parameter μ. Then $\mathcal{A}$ runs in* poly(ℓ, R) *time and*

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

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

As noted earlier, with the k-PC constraint we have r^2 different blocks, given by the partition, where each block has exactly one planted bit. This allows us to view the entire k-PC matrix as a collection of PB problems and apply BERN-ROTATIONS on each $(n/k) \times (n/k)$ matrix ($\ell = n^2/k^2$).

There are two remaining things to consider. Firstly, how to get from Gaussians back to Bernoullis and the final output, and secondly, what criteria does our design matrix $A \in \mathbb{R}^{k^2 \times k^2}$ have to follow. For the first step, as noted above, transforming $\mathcal{N}(0,1), \mathcal{N}(\nu, 1)$ to two Bernoullis by thresholding at 0 will not lose too much information measured by d_{TV} when μ is small, and the transformed signal will be approximately $\text{Bern}(0.5)$ and $\text{Bern}(0.5 + \frac{\mu}{\sqrt{2\pi}})$ since the Normal CDF is continuous.

To deal with the other part, we need each row of A to *map directly to the edge density parameter* of output. Specifically, for any (unknown) input PB instance, it gets mapped to an unknown row of A, which then becomes the output PDS mean. Our *design* in A is thus formulated as: how to find a suitable A such that each row of A corresponds to a possible mean adjacency matrix in target PDS.

2.3. Design matrices

We first remark that the key factor in BERN-ROTATIONS is the added noise $\mathcal{N}(0, I - AA^T/\lambda)$, which will in fact be the only part of our reduction process that may introduce irreversible signal loss. Consequently, we want to construct matrix A such that $I - AA^T/\lambda$ is as small as possible: A *has to be close to an isometry.* Let us first assume that $\sigma(A) = 1$ for simplicity.

As an example, suppose one wants to map from k-PC to the Gaussian version of PDS (i.e., $\mathcal{N}(\gamma, 1)^{\otimes k\times k}$ planted in $\mathcal{N}(0, 1)^{\otimes n\times n}$) with tight recovery boundary such that $k^2\gamma^2 \sim n$. As k-PC contains at most n planted bits yet the squared ℓ_2 norm of the target mean matrix is exactly $k^2\gamma^2 = \Omega(n)$, the sum of squared ℓ_2 norms of the n column vectors A_i being mapped to should be at least $\Omega(n)$, which (informally) implies that the design matrix A has to be an almost perfect isometry given $\sigma(A) = 1$.

Having independently generated random columns would allow to apply random matrix spectral bounds. For example, a matrix with i.i.d entries from some fixed distribution was used by Brennan and Bresler (2020). They proved that this methods achieves the desired spectral bound, but each column has a random number of planted bits resulting in binomial planted size rather than the desired fixed size (c.f., Remark 5).

Viewing the design matrix structure as the adjacency matrix of some graph, where i.i.d. matrices corresponds to Erdős-Rényi graphs, a natural alternative is regular graphs. These satisfy our fixed size constraint. Moreover, considering the tight recovery reduction again, one also needs all rows to have squared norms of $O(1)$ while summing up to $\Omega(n)$, making it an implicit regularization in our construction that all row norms have $\Theta(1)$ norm. This fact provides a crucial motivation into directed *regular graph* models for generating matrices such that the row norms and column norms align, and the columns are roughly independent (i.e. perpendicular).

2.4. Singular value from recentering

We will now focus on what happens in each sub-block with size $m = n/k$ given by partitioning k-PC, and treat it as our main target.[4] A line of works (Tikhomirov and Youssef (2019); Le et al. (2015)) have given high probability bounds on the spectral norm $\|A - \mathbb{E}(A)\|_{op}$ of adjacency matrix A for a random graph G with given degree distributions (planted signal). Here we consider 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 Tikhomirov and Youssef (2019), a (tight) high probability upper bound on the said quantity has been proven when $m^\alpha < d < m/2$ we have $|s_2(A)| \leq C_{\alpha,m}\sqrt{d}$ with high probability. With this result, we can establish the following lemma that will lead to the ultimate design matrix by taking the (translated) Kronecker product to make it $m^2 \times m^2$:

Lemma 8 (Random matrix with regular constraints) *Given constant $\alpha > 0$, there exists a constant C_α, such that for a $m \times m$ (random) matrix $R = R_{m,1/r}$ where $r < m^{1-\alpha}$ is an even divisor of m, with entries sampled from the following procedure:*

1. *Sample G uniformly from all directed m/r-regular graphs with size m.*
2. *$R_{ij} = \frac{-1}{\sqrt{mr}} + 1_{e_{ij}\in E_G} \cdot \sqrt{\frac{r}{m}}$ for $j \neq i$ off diagonal, $R_{ii} = \frac{-1}{\sqrt{mr}}$ on the diagonal.*

4. With a slight abuse of notations, we note this is different from the target planted size in PDS.

Then with probability $1 - o_m(1)$ *this matrix satisfies* $\|R\|_{op} \le C_\alpha$.

This (centered) matrix has a nice property in that it is an approximate isometry, where each row has $\frac{r-1}{r}$ fraction of $-\gamma = -1/\sqrt{mr}$ and $1/r$ fraction of $(r-1)\gamma$ with norm 1. However, it is not yet in the form we target (recall that we want to each column to map to the mean of the $m \times m$ adjacency matrix of a graph). It is natural to view the target PDS density as a translated rank-1 product of vectors (since it has one $k \times k$ elevated submatrix with uniform signal). Therefore we will simply take the *Kronecker product* to result in a $m^2 \times m^2$ matrix, which creates at the (i,j)th columns $R_i^T R_j$ where R_i are the rows of the $m \times m$ matrix.

However, the canonical rank-1 PDS formulation is *not centered*, having zeroes everywhere outside of the planted submatrix and elevated ones signal inside. To map to this instance, we first need to transform the centered signal $R \to \frac{1}{\gamma}(R + \gamma)$ so that we get $\frac{m}{r}$ ones in an all-zero vector for each row in R before taking the rank-1 product to get a $\frac{m}{r} \times \frac{m}{r}$ submatrix of ones inside $m \times m$ zeroes. This would make sure that the design matrix has exactly two different values. Unfortunately, doing so results in a product matrix that is guaranteed to have a large operator norm (since the output PDS_D is easy), explained intuitively because now our matrix is not centered.

To obtain a tighter spectral radius, it is natural for us to recenter the product matrix so that it has zero mean per column, corresponding to exactly PDS^*. This provides a justification from a design matrix perspective of why PDS^* is probably harder than PDS: *re-centering the design matrix decreases spectral norm, which results in a higher signal strength at the output.*

Lemma 9 (Construction of (fixed size) random $K_m^{1/r}$) *For given* α*, exist absolute constant* $C_\alpha > 0$*, such that for every* $m > r > 2$ *where* $r < m^{1-\alpha}$ *divides* m*, there exist* m *subsets* $A_1, A_2, \dots, A_m$ *of* $[m]$ *such that* $|A_i| = \frac{m}{r}$*, and that the* $m^2 \times m^2$ *matrix* $K_{(ij),(kl)} : i,j,k,l \in [m]$ *defined as* $K_{(ij),(kl)} = \mu\sqrt{\frac{r}{m}} \cdot (1_{k \in A_i \text{ and } l \in A_j} \cdot \frac{r}{m} - \frac{1}{mr})$ *has largest singular value at most* 1*. Specifically,*

$$K_m^{1/r} := K = \mu\sqrt{\frac{r}{m}}\left[(R + \frac{1}{\sqrt{mr}}J) \otimes (R + \frac{1}{\sqrt{mr}}J) - \frac{1}{mr}J \otimes J\right]$$

where J *is the all-one matrix and* R *satisfies the criteria from the previous lemma (*$\mu = (C_\alpha + 1)^{-2} \in \Theta_m(1)$*). With probability* $1 - o_m(1)$ *we can find a satisfying assignment in polynomial time.*

3. Hardness of Detection in Mean-corrected Null

We are now ready to state hardness for the degree-1 corrected null hypothesis testing problem PDS^* by constructing an average case mapping. We refer to Figure 2 (Theorem 27) for the full reduction.

Theorem 10 (Lower bounds for efficient PDS^* detection) *Consider hypothesis testing* PDS^* *for* $H_0 : G(n, p_0)$ *versus* $H_1 : \mathsf{PDS}(n,k,q,p)$ *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 exists a sequence* $\{(N_n, K_n, p_n, q_n)\}$ *of parameters such that:*

- *The parameters are in the regime* $p - q \in \widetilde{\Theta}(N^{-\alpha})$*,* $K \in \widetilde{\Theta}(N^\beta)$*. Formally,*

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

- *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 $\mathsf{PDS}^*(N_n, K_n, p_n, q_n)$ is at least 1 assuming the k-PC conjecture.*

Furthermore, we note that there exists a matching upper bound for PDS^*_D based on the empirical variance of degrees (see Proposition 26 for the precise result).

Recall that as discussed in Section 1.4, any recovery oracle (on $H_1 = \mathsf{PDS}^*_{H_1}$, which is the same as PDS_{H_1}) detects between $H_0 : G(n, p_0)$ versus $H_1 : \mathsf{PDS}(n, k, q, p)$ on its supported parameters, implying a natural upper bound on the decision problem. Combining Theorem 10 with Lemma 3, we obtain our main (lower bound) result for the signal strength required for recovery.

Corollary 11 (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:*

- *The parameters are in the regime $\gamma := |p - q| \in \widetilde{\Theta}(N^{-\alpha})$, $K \in \widetilde{\Theta}(N^{\beta})$.*
- *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 exact recovery on $\mathsf{PDS}(N_n, K_n, p_n, q_n)$ assuming k-PC.*

We remark that the constraint $\alpha < \beta$ comes from the fact that recovery is statistical impossible at $\alpha \geq \beta$ (see Theorem 24). For completeness, we refer to the appendix (Theorem 31) for an extended discussion on the statistical boundaries associated. Moreover, we remark that in light of our recovery to detection reduction framework, the detection-recovery gap can in fact be viewed as a detection (PDS) - detection (PDS^*) gap.

4. Hardness of Refutation

4.1. Detection hardness for ISBM

As before, given that a refutation blackbox only operates on H_0, we want to find some "quiet" distribution $\widetilde{H_1}$, such that it has the correct valuation but is hard to distinguish from a null instance. We will propose the ISBM model (3) in this section as a qualifying planted distribution. Due to the rank-1 nature of its bias structure, it is easy to construct design matrices by just taking the per-column rank-1 product from Lemma 8, hence hardness result can be proven similar to Theorem 27 with a reduction. As in the proof of reduction to PDS^*, we can then generalize to the complete boundary in ISBM detection, leading to refutation hardness. This is an extension of Theorem 3.2 in Brennan and Bresler (2020) where their (deterministic rotation kernel) reduction only works with a number-theoretic constraint restricting the parameters. Our results extend to the full boundary line by the regular concentration lemma on random matrices.

Theorem 12 (Hardness of detection in ISBM) *Consider hypothesis testing ISBM_D (3) where $k = n/r$ is the planted size. 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 $|P_{11} - P_{22}| \in \widetilde{\Theta}(N^{-\alpha})$, $R \in \widetilde{\Theta}(N^{\beta})$.*
- *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 k-PC.*

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

Equipped with the hardness results in ISBM, which has a large dense subgraph and thus can be used as a candidate $\widetilde{H}_1$ in refutation, we obtain the formal refutation hardness results similar in how we showed recovery hardness from a reduction with refutation (recovery) oracle:

Theorem 13 (Hardness in refutation of PDS in the dense regime) *Consider the 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 $p_1 - p_0 \in \widetilde{\Theta}(N^{-\alpha})$, $K \in \widetilde{\Theta}(N^{\beta})$.*
- *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 semi-definite programming relaxation (see appendix). Moreover, we also show that the statistical boundary for refutation lies exactly as that for recovery from applying a reduction to (statistical) recovery.

Theorem 14 (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 $\mathsf{val}(G) \le p$ with probability $\to 1$.*
- *When $kD_{KL}(p\|p_0) \in \widetilde{o}(1)$, the densest k subgraph $\mathsf{val}(G) \ge p$ with probability $\to 1$.*

Remark 15 *The problem of densest-k-subgraph in $G(n, \frac{1}{2})$ was very recently solved in Cheairi and Gamarnik (2022) with deep techniques from Bernoulli Disorder. However, here we can derive a log-optimal result using statistical reductions from recovery boundaries.*

5. Biclustering and Biased Sparse PCA

We point out a couple of other random models that have a detection hardness gap as an implication of PDS hardness guarantees. Those connections were first observed in Cai et al. (2017); Brennan et al. (2018) but under the conjectural tight hardness bound and Schramm and Wein (2022) with low-degree polynomials.

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 is to refute submatrices with large mean.

Biased SPCA Consider the *spiked covariance model* where v is a k-sparse unit vectors with non-zero entries equal to $\pm\frac{1}{\sqrt{k}}$:

$$H_0 : X_1, X_2, \ldots, X_n \sim \mathcal{N}(0, I_d)^{\otimes n} \quad \text{and}$$
$$H_1 : X_1, X_2, \ldots, X_n \sim \mathcal{N}\left(0, I_d + \theta vv^\top\right)^{\otimes n} \text{ where } \left|\|v\|_0^+ - \frac{k}{2}\right| > \delta \cdot k. \tag{5}$$

The recovery task is to estimate $\text{supp}(v)$ given observations $X_1, X_2, \ldots, X_n$ sampled from H_1. 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.

6. Open Problems

We point out two open problems related to our work:

1. Construct "quiet" H_0 hypotheses without any dense subgraphs that are hard to distinguish from PDS in order to resolve Conjecture 2. This would also imply a *detection-certification gap* as well as Conjecture 2 itself.
2. Can one can construct the inverse of the reduction of Remark 5, from a binomial version of PDS to the fixed sized PDS? This would show equivalence of the binomial and fixed versions.

Acknowledgments

This work was supported in part by NSF CAREER award CCF-1940205.

References

Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998.

Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '07, page 496–505, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595936318. doi: 10.1145/1250790.1250863. URL https://doi.org/10.1145/1250790.1250863.

Venkat Anantharam and Justin Salez. The densest subgraph problem in sparse random graphs. *The Annals of Applied Probability*, 26(1):305 – 327, 2016. doi: 10.1214/14-AAP1091. URL https://doi.org/10.1214/14-AAP1091.

Paul Balister, Béla Bollobás, Julian Sahasrabudhe, and Alexander Veremyev. Dense subgraphs in random graphs. *Discrete Applied Mathematics*, 260:66–74, 2019. ISSN 0166-218X. doi: https://doi.org/10.1016/j.dam.2019.01.032. URL https://www.sciencedirect.com/science/article/pii/S0166218X19300678.

Afonso S. Bandeira, Jess Banks, Dmitriy Kunisky, Cristopher Moore, and Alexander S. Wein. Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. *CoRR*, abs/2008.12237, 2020. URL https://arxiv.org/abs/2008.12237.

Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K. Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019. doi: 10.1137/17M1138236. URL https://doi.org/10.1137/17M1138236.

Gérard Ben Arous, Alexander S. Wein, and Ilias Zadik. Free energy wells and overlap gap property in sparse pca, 2020. URL https://arxiv.org/abs/2006.10689.

Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In Shai Shalev-Shwartz and Ingo Steinwart, editors, *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pages 1046–1066, Princeton, NJ, USA, 12–14 Jun 2013. PMLR. URL https://proceedings.mlr.press/v30/Berthet13.html.

Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage, 2020. URL https://arxiv.org/abs/2005.08099.

Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure, 2018. URL https://arxiv.org/abs/1806.07508.

Matthew Brennan, Guy Bresler, and Wasim Huleihel. Universality of computational lower bounds for submatrix detection, 2019. URL https://arxiv.org/abs/1902.06916.

Matthew S Brennan, Guy Bresler, Sam Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low degree tests are almost equivalent. In *Conference on Learning Theory*, pages 774–774. PMLR, 2021.

Cristina Butucea and Yuri I. Ingster. Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli*, 19(5B):2652 – 2688, 2013. doi: 10.3150/12-BEJ470. URL https://doi.org/10.3150/12-BEJ470.

T. Tony Cai, Tengyuan Liang, and Alexander Rakhlin. Computational and statistical boundaries for submatrix localization in a large noisy matrix. *The Annals of Statistics*, 45(4), aug 2017. doi: 10.1214/16-aos1488. URL https://doi.org/10.1214%2F16-aos1488.

Houssam El Cheairi and David Gamarnik. Densest subgraphs of a dense erdos-renyi graph. asymptotics, landscape and universality. *arXiv preprint arXiv:2212.03925*, 2022.

Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices, 2014. URL https://arxiv.org/abs/1402.1267.

Colin Cooper, Martin Dyer, Catherine Greenhill, and Andrew Handley. The flip markov chain for connected regular graphs, 2017. URL https://arxiv.org/abs/1701.03856.

Persi Diaconis and David Freedman. Finite exchangeable sequences. *The Annals of Probability*, pages 745–764, 1980.

Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2): 1–37, 2017.

David Gamarnik. The overlap gap property: a geometric barrier to optimizing over random structures. *CoRR*, abs/2109.14409, 2021. URL https://arxiv.org/abs/2109.14409.

David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 369–376, 2014.

David Gamarnik and Ilias Zadik. The landscape of the planted clique problem: Dense subgraphs and the overlap gap property. *arXiv preprint arXiv:1904.07174*, 2019.

Catherine Greenhill. A polynomial bound on the mixing time of a markov chain for sampling regular directed graphs, 2011. URL https://arxiv.org/abs/1105.0457.

Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. In Peter Grünwald, Elad Hazan, and Satyen Kale, editors, *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 899–928, Paris, France, 03–06 Jul 2015a. PMLR. URL https://proceedings.mlr.press/v40/Hajek15.html.

Bruce Hajek, Yihong Wu, and Jiaming Xu. Information limits for recovering a hidden community, 2015b. URL https://arxiv.org/abs/1509.07859.

Bruce Hajek, Yihong Wu, and Jiaming Xu. Semidefinite programs for exact recovery of a hidden community, 2016. URL https://arxiv.org/abs/1602.06410.

Chris Jones, Aaron Potechin, Goutham Rajendran, Madhur Tulsiani, and Jeff Xu. Sum-of-squares lower bounds for sparse independent set. *CoRR*, abs/2111.09250, 2021. URL https://arxiv.org/abs/2111.09250.

Can M. Le, Elizaveta Levina, and Roman Vershynin. Concentration and regularization of random graphs, 2015. URL https://arxiv.org/abs/1506.00669.

Zongming Ma and Yihong Wu. Computational barriers in minimax submatrix detection. *The Annals of Statistics*, 43(3):1089–1116, 2015. ISSN 00905364. URL http://www.jstor.org/stable/43556548.

Cynthia Rush, Fiona Skerman, Alexander S. Wein, and Dana Yang. Is it easier to count communities than find them?, 2022. URL https://arxiv.org/abs/2212.10872.

Tselil Schramm and Alexander S. Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833 – 1858, 2022. doi: 10.1214/22-AOS2179. URL https://doi.org/10.1214/22-AOS2179.

Konstantin Tikhomirov and Pierre Youssef. The spectral gap of dense random regular graphs. *The Annals of Probability*, 47(1):362 – 419, 2019. doi: 10.1214/18-AOP1263. URL https://doi.org/10.1214/18-AOP1263.

Alexander Veremyev, Vladimir Boginski, Pavlo Krokhmal, and David Jeffcoat. Dense percolation in large-scale mean-field random networks is provably "explosive". *PloS one*, 7:e51883, 12 2012. doi: 10.1371/journal.pone.0051883.

Appendix A. Notations and Preliminaries

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 $\mathit{Unif}_n(k)$ 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 $f(P)$ the push-forward. We also use sets in $V \in 2^{[n]}$ and vectors $v \in \{0, 1\}^n$ interchangeably, and $\mathsf{PDS}(n, S, p, q)$ to be the planted dense subgraph instance conditioned on planted location at set S.

Lemma 16 (Data Processing Inequality) *Let $\{$ be a Markov transition kernel and A, B be two distributions, then:*

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

Lemma 17 (Tensorization of TV) *Let P_i, Q_i be distributions for $i = 1, 2, \ldots, n$. Then:*

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

Lemma 18 (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:*

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

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

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

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

Lemma 19 (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 20 (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. ■

Appendix B. 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 Hajek et al. (2015a); Barak et al. (2019); Bandeira et al. (2020); Brennan et al. (2018). Here we re-formulate the necessary proofs:

Lemma 21 (Refutation implies detection) *Consider two hypotheses 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.*

Proof Consider the canonical form of refutation as described in Section 1.4 with a polynomial-timed refutation blackbox A having success probability $p > 0$. We show that A:

- When $G \sim H_0$, $A(G) = 0$ with probability at least $p \cdot P(\mathsf{val}(G) < \epsilon - \delta | G \sim H_0)$, thus the Type I error is at most $1 - p \cdot P(\mathsf{val}(G) < \epsilon - \delta | H_0)$.

- When $\mathsf{val}(G) > \epsilon + \delta$ the output is always 1, thus the Type II error is at most the probability of a low valuation $P(\mathsf{val}(G) > \epsilon + \delta | H_1)$.

Therefore, the sum of errors is bounded above by:

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

Note that as $n \to \infty$,

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

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 22 (Recovery implies detection) *Fix two hypothesis H_0, H_1 with valuation function val_s that can be computed in polynomial time given key s and define $\mathsf{val}(G) = \max_s \mathsf{val}_s(G)$.*

Assuming $\mathsf{val}(G \sim H_0) \leq \epsilon - \delta$ with high probability and suppose that a polynomial time oracle (on input G) generates a key k such that $\mathsf{val}_k(G) > \epsilon$ with high probability over $G \sim H_1$, then detection is possible in polynomial time equipped with such key oracle.

Proof Consider the alternate valuation function $\mathsf{val}' = \mathsf{val}_k$, which can be computed in polynomial time by first asking $k(G)$ from the given oracle. From the previous lemma and the separation conditions we know that:

1. For $G \sim H_0$, $\mathsf{val}'(G) \leq \mathsf{val}(G) \leq \epsilon - \delta$ with high probability.
2. For $G \sim H_1$, $\mathsf{val}'(G) > \epsilon$ 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 21, we have the desired conclusion. ■

Appendix C. 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 both statistically and computationally; we will use it as a crucial reduction step to Theorem 14.

Lemma 23 (Minimal recovery implies exact recovery) *For the $\mathsf{PDS}(n, k, p, q)$ 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 a subroutine GRAPH-CLONE (Lemma 5.2 in Brennan et al. (2019)) 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 $\mathcal{M}_{[N]\times[N]}(\eta\times\eta, \mathrm{Bern}(P), \mathrm{Bern}(Q'))$.

Our assumption also 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 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

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

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 25 means that the direct reduction in Lemma 23 does not apply. In fact, it will be easy to see that partial recovery for our instance in Theorem 25 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 in below.

Theorem 24 (PDS recovery – Theorem 2 in Hajek et al. (2015b)) *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.

C.1. Evidence of recovery hardness

We further remark that with some *relaxed* condition on the signal, one can prove tight recovery hardness for some models on DkS (Densest k-Subgraph) valuation that resembles PDS in structure.

Theorem 25 *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. This result also suggest that, if PDS recovery is indeed easier than conjectured, the algorithm must use heavily the community structure.

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 22. Consider the following procedure applied on a graph $G \sim \mathsf{PDS}(n, k, 1, 1/2)$ 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 $k \in \widetilde{\omega}(\sqrt{n})$. ■

Appendix D. Proofs for Design Matrices

D.1. Proof of Lemma 8

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 Tikhomirov and Youssef (2019) 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 $1 - o_n(1)$. ■

D.2. Proof of Lemma 9

Note that we can sample from directed regular graphs efficiently by the ergodicity of a simple edge-flipping Markov process (Greenhill (2011); Cooper et al. (2017)), and hence we have the following.
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:

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

because the top eigenvalue of $J = \mathbb{1}\mathbb{1}^T$ is exactly n. Therefore $\sigma(K) \leq 1$ for any R satisfying the criteria of Lemma 8.

Finally, note Theorem 1 in Greenhill (2011) 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 8 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 independently, the probability of failure becomes exponentially small. ■

Appendix E. Proofs for Recovery

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 Schramm and Wein (2022).

Proposition 26 (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.
Proof 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 $\mathrm{Binom}(n, p_0)$ with expectation given by

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

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

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

and the k nodes planted has:

$$\begin{aligned}
&\mathbb{E}(\sum_{i\in v} d_i^2) = k\mathbb{E}_{x\sim\mathrm{Binom}(n-k,\,q_1),y\sim\mathrm{Binom}(k,\,p_1)}(x+y)^2 \\
&= k\left(\left((n-k)q_1(1-q_1)+(n-k)^2q_1^2\right)+kp_1(1-p_1)+k^2p_1^2+2q_1p_1k(n-k)\right) \\
&= k(n-k)q_1+k^2p_1+k((n-k)q_1+kp_1)^2-k(n-k)q_1^2-k^2p_1^2
\end{aligned}$$

The difference between expectations in H_0 and H_1 is thus:

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

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

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

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

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

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

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 $(\mathbb{E}_{H_1}(f)+\mathbb{E}_{H_0}(f))/2$). ■

E.1. Proof of reduction to PDS*

Theorem 27 (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 $0 < q < p \le 1$ 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 (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:*

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

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 28 (TO-k-PARTITE-SUBMATRIX $-$ Lemma 7.5 in Brennan and Bresler (2020)) *With the given assumptions, step 1 (denote as* $\mathcal{A}_1$*) of the reduction runs in poly*(N) *time and it follows that:*

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

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

Lemma 29 (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* $K_{n/k_0}^{1/r}$ *be the design matrix obtained from Lemma 9 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*:*

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

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 $E_1, E_2, \ldots, E_t$. 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 $S_1, S_2, \ldots, S_{k_0}$ 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:

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

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

2. *Flattened Bernoulli Rotations:* Let S be a partition of $[n]$ into k_0 equal parts $S_1, S_2, \ldots, S_{k_0}$ 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 $(K^{1/r}_{n/k_0})^T$, 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*.

Lemma 30 (Thresholding from Gaussians) *Let $\mathcal{A}_3$ be the final step from the above reduction, with the same notations as the previous lemma, then:*

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

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

As an important pre-processing step, we single out the steps in Lemma 28 first, the proof in its exact form is deferred to Brennan and Bresler (2020). 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 28, we arrive at a bi-partite k-PDS instance with slightly different parameters, and we prove the following lemmas to complete the reduction.

Proof [Lemma 29] 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

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

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):

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

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 $O((\frac{n}{k_0})^2R_{rk}^{-3})$ by Lemma 7.

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 $O(n^2R_{rk}^{-3})$ in d_{TV}, which results in at most $O(n^4R_{rk}^{-3})$ 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 [Lemma 30] 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 $\mathsf{PDS}(n, n/r, P_2, P_1)$ as the output. ■

Proof [Theorem 27] 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:

$$\begin{aligned}
\mathcal{P}_0 &= G_E(N, k, p, q) \\
\mathcal{P}_1 &= \mathcal{M}_{[n]\times[n]}(S \times S, \mathrm{Bern}(p), \mathrm{Bern}(Q)) \quad \text{where } S \sim \mathcal{U}_n(F) \\
\mathcal{P}_2 &= \gamma \cdot \mathbb{1}_S \otimes \mathbb{1}_S + \mathcal{N}(-\frac{\gamma}{r^2}, 1)^{\otimes n \times n} \quad \text{where } S \sim \mathit{Unif}_n(k) \\
\mathcal{P}_4 &= \mathsf{PDS}(n, k, P_1, P_2)
\end{aligned}$$

Applying Lemma 28 before, we can take

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

where $C_Q = \max\left\{\frac{Q}{1-Q}, \frac{1-Q}{Q}\right\}$. For ϵ_2, Lemma 29 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 $d_{\mathrm{TV}}(\mathcal{A}_i(\cdot), \mathcal{A}_i(\cdot')) \le d_{\mathrm{TV}}(\cdot, \cdot')$ so each step the total variation loss at most accumulates (Lemma 18), thus by the triangle inequality on TV we get

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

Under H_0, consider the distributions

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

As above, Lemmas Lemma 28, Lemma 29 and Lemma 30 imply that we can take

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

Again by the data processing inequality (Lemma 18), 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. ■

E.2. Proof of Theorem 10

Suppose we now have a direct reduction to $\mathsf{PDS}(n,k,P_1,P_2)$ 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:

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

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

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 $\mathrm{Bern}(s)$ instance get transferred (independently) directly to $\mathrm{Bern}(2P_0s)$ 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 CDF) 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 $\mathcal{A}_4(\mathsf{PDS}(n,k,P_1,P_2)) = \mathsf{PDS}(n,k,p_1,2P_2P_0)$, and we only need to show that $d_{\mathrm{TV}}(\mathsf{PDS}(n,k,p_1,2P_2P_0),\mathsf{PDS}(n,k,p_1,p_2)) = o(1)$. The trick here is to use the data processing inequality again: because the distribution $\mathsf{PDS}(n,k,p,q)$ is obtained by applying the (random) planted dense subgraph over $G(n,q)$, thus the total variation:

$$\begin{aligned} d_{\mathrm{TV}}(\mathsf{PDS}(n,k,p_1,2P_2P_0),\mathsf{PDS}(n,k,p_1,p_2)) &\le d_{\mathrm{TV}}(G(n,2P_2P_0),G(n,p_2)) \\ &= d_{\mathrm{TV}}(\mathrm{Bern}(2P_0P_2)^{\otimes\binom{n}{2}},\mathrm{Bern}(p_2)^{\otimes\binom{n}{2}}). \end{aligned}$$

Moreover, by Lemma 20 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:

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

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 $\mathsf{PDS}(n, k, p_1, p_2)$ 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 Proposition 26.

E.3. Proof of Corollary 11

Proof By Lemma 22, 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 37 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 37 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 22 and Theorem 10 we are done with the proof. ■

E.4. Statistical Optimal 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 Butucea and Ingster (2013); Hajek et al. (2015a); Ma and Wu (2015) 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 31 (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 $P_\theta = \mathsf{PDS}(n, \theta, p, q)$ with planted set at θ and $\theta \sim \binom{n}{k}$ be uniformly distributed.

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

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 Hajek et al. (2015a)), which concludes our proof.

A better way to view it (from reductions) is as follows: we know that when the inequality condition holds, $\mathsf{PDS}(n, k, p, q)$ is in-distinguishable from $G(n, k, q)$ by the PDS boundary in Brennan et al. (2018). However, in this case we have:

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

by Lemma 20. Therefore, $d_{\mathrm{TV}}(G(n,q), G(n,p_0)) \to 0$ 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 32 *We also remark that the reverse direction for the above lower bound is still open. 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 a stronger version of Lemma 6 in Hajek et al. (2015a), then χ^2 between two hypotheses of PDS_D^* diverges, which is still a necessary (insufficient) condition for the upper bound. It is of interest to show a tight statistical detection upper bound for models like ISBM *and* PDS^* *at the regime when it is computationally infeasible and statistically impossible to recover.*

Appendix F. Proofs for Refutation

F.1. Proof of Theorem 12

Theorem 33 (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 ISBM 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
$$\mu \le \frac{1}{12C\sqrt{\log(N) + \log(p-Q)^{-1}}} \cdot \min\{\log(\frac{p}{Q}), \log(\frac{1-Q}{1-p})\}.$$
- *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 28) 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 3: 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 8. Therefore:

Lemma 34 (Bernoulli Rotation for ISBM$_D$) *Consider the second step $\mathcal{A}_2$ applied on the output of Lemma 28 and assuming notations through Lemma 29, we have:*

$$d_{\mathrm{TV}}(\mathcal{A}_2(\mathrm{Bern}(Q)^{\otimes n \times n}, \mathcal{N}(0,1)^{\otimes n \times n}) = o(n^{-1})$$
$$d_{\mathrm{TV}}(\mathcal{A}_2(\mathcal{M}_{[n]\times[n]}(\mathcal{U}_n(S), p, Q)),\ \mathcal{L}(\frac{\gamma}{r^2} \cdot (rv-1)^T(rv-1) + \mathcal{N}(0,1)^{\otimes n \times n})) = o(n^{-1})$$

where $v \sim$ Unif$_n(k)$.

Proof Similar to Lemma 29, 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 Brennan and Bresler (2020)). 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 27. We defer a formal full proof in thresholding Gaussians and aligning the precise density to Corollary 14.5 and Theorem 3.2 in Brennan and Bresler (2020), 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.

F.2. Proof of Theorem 13

Proof [Refutation hardness] Consider applying Lemma 21, 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 $\mathsf{val} = \mathsf{DkS}$, 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 37 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 12 actually reads $(P_{11} - P_0)^2 \in \widetilde{\Theta}(n/k^2)$.

Therefore, from Theorem 12 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. ■

F.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 Hajek et al. (2016); Chen and Xu (2014).

Consider the following relaxation of the densest-subgraph:

$$\begin{aligned} \widehat{Z}_{SDP} = \arg\max_{Z} & \langle E, Z\rangle \\ \text{s.t. } & Z \succeq 0, \quad Z \geq 0 \\ & Z_{ii} \leq 1, \quad \forall i \in [n] \\ & \langle I, Z\rangle = k \\ & \langle J, Z\rangle = k^2. \end{aligned} \tag{6}$$

It is not hard to see that:

- A feasible solution of a true subgraph is also feasible for (6), thus the latter will always return objective at least the true density.
- (6) is a semi-definite programming problem, and can be efficiently solved.

With the sufficiency results given in Hajek et al. (2016) (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 $\langle E, Z\rangle \leq k^2 p + ck\sqrt{p(1-p)}$ with probability $\to 1$ as constant $c \in \Omega_{n,k}(1)$ by Markov Inequality).

Theorem 35 *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 $H_1 = \mathsf{PDS}(n, k, p, q)$, the objective will be at least $\frac{q+4p}{5}$ with probability $\to 1$. This means that (6) will successfully refute the densest k-subgraph valuation problem in $G(n, q)$ vs $\mathsf{PDS}(n, k, p, q)$.*

Proof We start with the following lemma from stochastic domination:

Lemma 36 *Let $F(P)$ be the distribution of objective (6) 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 $E_P - E_Q \geq 0$ (entry-wise), then for any $x > 0$, $\mathbb{P}(F(P) > x) \geq \mathbb{P}(F(Q) > x)$. 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 (6).
2. Update G in the following way: for any $e_G = 0$, flip $e_G = 1$ with probability $\frac{E_p(e) - E_q(e)}{1 - E_q(e)}$.

The objective never decreases because we only add edges in the second step, whereas the unconditional distribution of the graph generated from 2 is exactly P. Hence we find a coupling between two distributions 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 (6) 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 $F(\mathsf{PDS}(n,k,p,q))$ dominates $F(G(n,q))$ for any density $p > q$. Consider the alternative $\mathsf{PDS}(n,k,(p+q)/2,q)$ for (6), which also satisfies the conditions for successful recovery by (6). Note that in this case, we know that the objective:

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

with probability $\to 1$ 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 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 $X \le k^2 \cdot \frac{4p+q}{5}$ with probability at most

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

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

F.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 Veremyev et al. (2012); Anantharam and Salez (2016); Balister et al. (2019). However, our regime of interest ($k = \Theta(n^\alpha), q \in \Omega(1)$) remains largely unstudied in past literature.

Theorem 37 *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 $\to 1$ if $\frac{kd}{\log n} \to \infty$ and $k \to \infty$. Thus statistical refutation is possible.*

Proof Firstly we need a tail bound on the Binomial distribution (for $r := \lceil pN \rceil$):

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

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:

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

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 38 (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*[5]*. 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 $p - q \in \Omega(\frac{\log^2 n}{n})$. In fact, from the statistical boundary on exact recovery given in Hajek et al. (2015b), we know that recovery is impossible in this regime, even in the minimal variant (from the reduction given in Lemma 23).

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 planting the 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'):

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

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 $E_T^{G_1'} < r_{|T|} \binom{|T|}{2}$.

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

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:

$$(r_{|T|} - s_{|T|})\binom{|T|}{2} \le \frac{1}{6}(p-q)\binom{k}{2}.$$

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 $k^2(p-q) \in \Theta(k^{3/2}\sqrt{\log^{-1.01} n})$.[6]

Now note that $r - s = p - q + (d + f)$, thus what remains to show is the local inequality

$$|T| \in o(\frac{k}{\log^{0.7} k}), \qquad \text{when } k(p-q) \in \Theta(\log^{-1} n)$$

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 24), below which it is impossible to perform (even weak) recovery, so the above bound on $|T|$ follows immediately from Lemma 23 with strength $\alpha = 1.004$ (so the expected size of $|T|$ cannot be greater than $\frac{k}{(\log k)^{1.004}} < \frac{k}{(\log k)^{0.7}}$ by minimal recovery). ■

As a conclusion to this section, we note that a version of the refutation bound was very recently closed in Cheairi and Gamarnik (2022), which states that the exact refutation boundary lies in $k(p-q)^2 \in \Theta(\log n)$ (below which a dense subgraph exist with high probability). Though our theorem above is a weaker version of their result, the goal is to provide insights into reductions via recovery.

Appendix G. 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 Brennan et al. (2018), 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 11. Denote the H_1 hypothesis distributions in Section 5 as $\mathrm{BC}(n, k, \mu)$ and $\mathrm{BSPCA}(m = n, k, d, \theta)$, respectively.

G.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.

6. 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 $\widehat{E}$.

Lemma 39 (Reduction to Bi-clustering – Lemma 6.7 in Brennan et al. (2018)) *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

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

With this, we can now state the lower bound for recovery and refutation in BC by Lemma 22:

Corollary 40 (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 k-PC 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 $\beta > \frac{1}{2} + \frac{1}{2}\alpha$.

G.2. Detection-Recovery gap in BSPCA

Theorem 41 (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 k-PC 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 $\text{BSPCA}(N_n, K_n, D_n, \theta_n)$ is greater than 0 asymptotically, where the supports are independently distributed from the uniform prior.*

Proof The proof follows from the following lemma:

Lemma 42 (Random Rotation – Lemma 8.7 in Brennan et al. (2018)) *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*

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

We defer the proof to Brennan et al. (2018) and focus on the reduction forward. Note that the left-hand side can be viewed as the asymmetric biclustering distribution, thus combining Lemma 39 and Lemma 42 with Lemma 18 we get a polynomial-time map $\mathcal{A}$ such that:

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

Now the only thing left is to define precise parameter correspondence to apply Theorem 11. 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 11, which concludes that no polynomial black-box can successfully recover the planted instance u here. ■