Algorithmic Decorrelation and Planted Clique in Dependent Random Graphs: The Case of Extra Triangles

Guy Bresler
LIDS
MIT
Cambridge, MA, USA
guy@mit.edu

Chenghao Guo
LIDS
MIT
Cambridge, MA, USA
chenghao@mit.edu

Yury Polyanskiy
LIDS
MIT
Cambridge, MA, USA
yp@mit.edu

***Abstract*—We aim to understand the extent to which the noise distribution in a planted signal-plus-noise problem impacts its computational complexity. To that end, we consider the planted clique and planted dense subgraph problems, but in a different ambient graph. Instead of Erdős-Rényi $G(n,p)$, which has independent edges, we take the ambient graph to be the *random graph with triangles* (RGT) obtained by adding triangles to $G(n,p)$. We show that the RGT can be efficiently mapped to the corresponding $G(n,p)$, and moreover, that the planted clique (or dense subgraph) is approximately preserved under this mapping. This constitutes the first average-case reduction transforming dependent noise to independent noise. Together with the easier direction of mapping the ambient graph from Erdős-Rényi to RGT, our results yield a strong equivalence between models. In order to prove our results, we develop a new general framework for reasoning about the validity of average-case reductions based on *low sensitivity to perturbations*.**

***Keywords*—average-case complexity, randomness in computing**

I. Introduction

Most modern statistical inference problems exhibit a striking phenomenon: the best efficient algorithm requires substantially more data, or lower noise level, than the information-theoretic limit achieved by inefficient algorithms. Such problems are said to exhibit a *statistical-computational gap*. In this paper we are interested in *average-case* planted problems, meaning that their inputs are sampled according to some probability distribution with planted structure. In order to develop methodology for rigorously reasoning about the computational complexity of these average-case problems, a vibrant line of research at the interface of statistics, probability, and computational complexity has emerged. There are two main approaches to substantiating computational limits: hardness against specific classes of algorithms and average-case reductions.

The first approach attempts to determine the best possible performance for restricted classes of algorithms, such as low-degree polynomials [1], [2], sum-of-squares (SoS) relaxations [3], [4], statistical query algorithms [5]–[7], first-order methods [8], or classes of circuits [9], [10]. Beyond hardness against specific classes of algorithms, the overlap gap method introduced in [11] and refined in [12] has revealed how structural properties of a solution landscape can rule out multiple classes of algorithms. The power of distinct classes of algorithms for solving families of statistical problems have been related: low-degree polynomials vs. statistical query [13], low-degree polynomials vs. approximate message passing [14], and SoS vs. spectral applied to matrices of low-degree polynomials [4]. This broad approach continues to see intense activity and yields insight into the limits of current algorithms. However, it has the drawbacks that (1) each given class of algorithms is known to be suboptimal in certain settings [15]–[17], which reduces confidence in lower bounds against any specific class, and (2) one must prove fresh lower bounds for each problem of interest and for each new class of algorithms that emerges.

The second approach entails devising average-case reductions that map one statistical problem to another using a polynomial-time algorithm, whereby hardness of the first problem is transferred to the second. The reduction approach is a foundational tool across com-

This paper is supported in part by NSF Career award CCF-1940205, CCF-2131115, NSF TRIPODS grant DMS-2022448 and the MIT-IBM Watson AI Lab

plexity theory and cryptography for elucidating relationships between problems and constructing complexity hierarchies. Moreover, it provides insights applicable to all algorithms. The main drawback of this approach is that in the average-case setting, one must precisely map not just problem instances, but rather entire probability distributions over inputs. This is notoriously challenging due to a lack of techniques, as emphasized by Goldreich [18], Barak [19], and Bogdanov and Trevisan [20].

A common starting hardness assumption in the literature on average-case reductions is the Planted Clique (PC) Conjecture. This conjecture posits that no polynomial-time algorithm can detect a planted clique of size k in an Erdős-Rényi random graph with edge density p, when p is constant and $k = o(\sqrt{n})$. Many problems have been shown to be computationally hard based on PC Conjecture assumption including sparse principle component analysis [21]–[25], submatrix detection or biclustering [24], [26], [27], planted dense subgraph [24], [28]. A variety reduction techniques were introduced in [24] and [29], resulting in a web of reductions from PC to problems including robust sparse mean estimation, tensor PCA, general planted dense subgraph, dense stochastic block model, mixtures of sparse linear regressions, robust sparse mean estimation, and many others. Some works show reductions from hypergraph planted clique or hypergraph planted dense subgraph to other problems, including tensor clustering [30] and SVD for random 3-tensors [31]. Another line of research has also explored reductions to statistical problems based on cryptographic assumptions such as lattice problems [32]–[38] or local pseudorandom generators [39], [40].

A slight generalization of Planted Clique is the Planted Dense Subgraph (PDS) problem. Instead of planting a clique of size k, PDS plants a subgraph of size k with higher density than the rest of the graph. It is conjectured to have the same $k = o(\sqrt{n})$ hardness threshold as PC when the edge probabilities (both inside and outside the dense subgraph) are constant. Most problems that have reductions from PC can also be reduced from PDS, for instance all of the reductions in [24] and [29]. In this paper, we will mostly focus on PDS rather than PC. We remark that the worst-case version of the problem, i.e., detecting the densest k-subgraph, has also studied, with the goal of finding the optimal approximation ratio [41]–[43].

Despite these advances, the current literature on average-case reductions for statistical problems exhibits a notable limitation: the absence of two-way equivalences. Most reductions rely on a single hardness assumption, such as the Planted Clique Conjecture or Learning Parity with Noise (LPN), and demonstrate a statistical-computational gap for other problems. Going beyond this, one might dream of partitioning statistical problems into equivalence classes, just as has been done so fruitfully in worst-case complexity. We mention here, only briefly, that the notion of equivalence is itself nontrivial because each statistical "problem" is itself an entire *parameterized family* of problems. We rigorously define what we mean by strong computational equivalence in Section 3 of the full paper.

The primary obstacle to proving equivalence between statistical problems, just as for one-way reductions, is that average-case reductions are challenging: we must transform one noise distribution into another while preserving the planted signal. To date, all reductions have converted models with independent noise to models with either independent or dependent noise. The challenge in establishing equivalence between planted problems is therefore in developing techniques for removing dependence in the noise of high-dimensional distributions. In this paper, we introduce a general framework for the validity of reductions based on low sensitivity to perturbations and apply it to show the first non-trivial computational equivalence between two statistical problems.

A second (related) motivation for the present paper is towards developing an understanding of how general are the phenomena observed in planted statistical problems. For planted matrices or tensors with independent entries, a reduction was devised from planted clique or planted dense subgraph to a broad class of entry distributions by [27], thereby showing that the observed phenomena are *universal* for this class. Thus, we understand how to change the distribution of the entries, but as of yet there are no general techniques for connecting problems with different *dependence* between the entries[1].

A. The Problems We Consider

a) General Hypothesis Testing Setup: In a simple-versus-simple hypothesis testing problem there is a pair of distributions P_0 and P_1 on space Ω, and one observes X generated according to one of the distributions. The task is to decide between the two hypotheses

$$H_0 : X \sim P_0 \quad \text{versus} \quad H_1 : X \sim P_1 \,.$$

We consider noisy signal detection problems indexed by problem size parameter n, where the space is Ω_n,

[1]The exception is sparse PCA, which has a very specific dependence structure.

$P_{0,n}$ represents the pure noise distribution, and $P_{1,n}$ is the distribution with planted signal. We often keep the dependence on n implicit. An algorithm $\Phi : \Omega \to \{0,1\}$ *solves* the hypothesis testing problem if the sum of type I and type II errors asymptotically vanish, i.e.,

$$\mathbb{P}_{X\sim P_0}(\Phi(X)=1) + \mathbb{P}_{X\sim P_1}(\Phi(X)=0) \to 0$$

as $n \to \infty$.

b) Planted Dense Subgraph Problem: In this paper, we consider a slight generalization of Planted Clique known as the *Planted Dense Subgraph (PDS)* problem. Here, the pure noise distribution is the simple Erdős-Rényi random graph, $G(n,p)$. The alternative hypothesis, denoted by $G(n,p,k,q)$, is generated by starting with $G(n,p)$, choosing a uniformly random subset S of size k from $[n]$, erasing all edges in S, and independently resampling the edges in S to be included with probability q. The PDS problem is, given a graph G, to distinguish between the two hypotheses

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

It is believed that the PDS problem with constant p and q has the same $k = o(\sqrt{n})$ computational threshold as the planted clique problem [3], [1].

c) Random Graph with Triangles: Next, instead of the ambient graph being Erdős-Rényi, we consider a simple graph distribution with dependent edges. We emphasize that while this model is interesting in its own right, our main motivation is as a tractable setting to develop techniques for algorithmic decorrelation.

Definition 1 (Random Graph with Triangles)**.** The random graph with triangles distribution $\mathsf{RGT}(n,p,p')$ is the law of a graph generated as follows. Let $G \sim G(n,p)$ be an Erdős-Rényi random graph and for every triple of nodes $\{i,j,k\} \in \binom{[n]}{3}$ with probability p' independently add the triangle consisting of the three edges $(i,j),(j,k)$, and (i,k) to $E(G)$.

Remark 1. The expected edge density of $G(n,p,p')$ is $p+(1-p)\left(1-(1-p')^{n-2}\right)$. It is always assumed that $p' = O(1/n)$ so that the edge density remains bounded away from 1.

Remark 2. In certain parameter regimes, the RGT model is close in total variation distance to the Random Intersection Graph, a model inspired by real-world networks. This will be explained in more depth in Section I-B.

Remark 3. Adding triangles is not only a simple way of introducing dependence among edges, but also a well researched phenomenon in sociology [44]. This phenomenon is commonly referred to as "triadic closure", which says if two people share a social connection, a third person is more likely to form connections with both of them than would be expected otherwise.

d) Planted Dense Subgraph in Random Graph with Triangles: We plant a signal in the random graph with triangles (*planted RGT*) to obtain $\mathsf{RGT}(n,p,p',k,q)$, the law of a graph generated as follows. Start with $\mathsf{RGT}(n,p,p')$, choose a random subset S of k vertices, erase all edges within S and then independently include each edge within S with probability q. The Planted RGT problem is, given a graph G, to distinguish between the two hypotheses

$$H_0 : G \sim \mathsf{RGT}(n,p,p')$$

and

$$H_1 : G \sim \mathsf{RGT}(n,p,p',k,q)\,.$$

Remark 4. There are multiple options for the definition of Planted RGT. We will discuss a substantial generalization in the next section. One could also, for example, plant a smaller RGT with different edge or triangle densities within a larger RGT. We leave this for future work.

B. Our results

Our main result is a strong computational equivalence between the planted dense subgraph (PDS) and planted random graph with triangles (Planted RGT) problems. To achieve this, we must design transformations that connect these problems in both directions. Here, and throughout the paper, for two distributions μ and ν we use $\mu \approx_\epsilon \nu$ as shorthand for $d_{\mathsf{TV}}(\mu,\nu) \le \epsilon$.

Theorem I.1. *For any $k = o(n^{1/4}\log^{-17/4} n), p' \le 1/(n\log n)$ and $0 < p,q < 1$, there are efficient algorithms $\mathcal{A}_\triangle$ and $\mathcal{A}$ that satisfy the following transition properties.*

1) *$\mathcal{A}_\triangle$ maps both H_0 and H_1 hypotheses of PDS to planted RGT with parameter map f:*

$$\mathcal{A}_\triangle(G(n,p)) \approx_{o_n(1)} \mathsf{RGT}(n,p,p')\,,$$

$$\mathcal{A}_\triangle(G(n,p,k,q)) \approx_{o_n(1)} \mathsf{RGT}(n,p,p',k,f(q))\,,$$

2) *$\mathcal{A}$ maps both H_0 and H_1 hypotheses of planted RGT to PDS with parameter map g:*

$$\mathcal{A}(\mathsf{RGT}(n,p,p')) \approx_{o_n(1)} G(n,p)\,,$$

$$\mathcal{A}(\mathsf{RGT}(n,p,p',k,q)) \approx_{o_n(1)} G(n,p,k,g(q))\,.$$

Moreover, the parameter mappings f and g satisfy $f \circ g(q) = q + o_n(1)$ and $g \circ f(q) = q + o_n(1)$.

Part 1 entails transforming from PDS to planted RGT, i.e., an independent noise model to a dependent noise model. In the case of $G(n,p)$ to $\mathsf{RGT}(n,p,p')$, the mapping itself is straightforward based on the definition of RGT: one simply adds each triangle with probability p'. However, the fact that the planted distributions are mapped correctly via the very same procedure does still require a nontrivial argument.

In Part 2, the other direction, we show how to map from RGT to Erdős-Rényi. Here the mapping even in the unplanted pure-noise case is not obvious, and we describe it in Section II-B. It is obtained by viewing the triangle addition procedure as a Markov transition kernel on graphs, and implementing its time-reversal.

To constitute a valid reduction from Planted RGT to PDS, the same mapping must also work when a dense subgraph has been planted. We face a delicate balance between two competing objectives: (1) maintaining the algorithm's behavior outside of the dense subgraph, ensuring that its performance remains approximately invariant to the placement of the dense subgraph, and (2) removing the dependence between triangles, which inherently necessitates dependence in the algorithm's behavior.

Validity of both reductions is proved via application of a general framework that we develop, stated as Theorem II.1 in Section II. Applying our general theorem presents a number of technical challenges in the form of bounding total variation distances between perturbed distributions. Methods for proving the proximity of two high-dimensional distributions in total variation distance are notably limited, and even more so when neither of them is a product distribution. To this end, we introduce several technical innovations that will be discussed in detail in the next section.

We conjecture that the statement of Theorem I.1 holds for k up to $\tilde{O}(\sqrt{n})$, while our result only applies for $\tilde{O}(n^{1/4})$. The range of p' is optimal up to log factors: if $p' \geq 2.1n^{-1}\ln n$, then with high probability the RGT is the complete graph and planted RGT is information-theoretically impossible, so the planted RGT cannot be equivalent to PDS in this regime. We leave open the problem of obtaining a similar result for the full range of k, p, q, and p'.

a) General Planted Signal: Our results not only demonstrate that the detection of planted dense subgraphs problem is equivalent in Erdős-Rényi graphs and RGT, but also reveal that a wide range of planted signal detection problems are equivalent in both models. Let us define the general planted signal (GPS) model.

Definition 2 (General Planted Signal (GPS))**.** For edge sequence $\vec{e} = (e_1, \ldots, e_K)$ and $\vec{p} = (p_1, \cdots, p_K)$ let $\mathsf{GPS}_{\vec{e},\vec{p}}$ take as input a graph, and for each $1 \leq i \leq K$, edge e_i is resampled to be included in the graph with probability p_i. For a pure-noise distribution $\mathsf{P_0}$, let $\mathsf{GPS}(\mathsf{P_0})$ denote the distribution of $\mathsf{GPS}(G)$ where $G \sim \mathsf{P_0}$.

Planted dense subgraph is a special case of GPS with $e_1, \cdots, e_{\binom{k}{2}}$ being the edges inside the dense subgraph, and $p_i = q$ is the planted edge density. The GPS problem in random graph distribution $\mathsf{P_0}$ is the hypothesis testing problem between $\mathsf{P_0}$ and $\mathsf{GPS}(\mathsf{P_0})$.

Our bi-directional reduction result holds also in this general setting.

Theorem I.2. *Let $\mathsf{GPS}_n = \mathsf{GPS}_{\vec{e},\vec{p}}$ be a sequence of general planted signal problems. For any $K = o(\sqrt{n}\log^{-17/2} n), p' \leq 1/(n\log n)$ and $p_1, \ldots, p_K$ each bounded away from 0 and 1 by a constant, there are distributions $\mathsf{GPS}_n^g = \mathsf{GPS}_{\vec{e},g(\vec{p})}$ and $\mathsf{GPS}_n^f = \mathsf{GPS}_{\vec{e},f(\vec{p})}$ (where $f(\vec{p})$ stands for $f : \mathbb{R} \to \mathbb{R}$ applied to all elements of the vector) and efficient algorithms $\mathcal{A}$, $\mathcal{A}'$ satisfying the transition properties.*

1)

$$\mathcal{A}(G(n,p)) \approx_{o_n(1)} \mathsf{RGT}(n,p,p')$$

$$\mathcal{A}(\mathsf{GPS}_n(G(n,p))) \approx_{o_n(1)} \mathsf{GPS}_n^f(\mathsf{RGT}(n,p,p')),$$

and

2)

$$\mathcal{A}'(\mathsf{RGT}(n,p,p')) \approx_{o_n(1)} G(n,p)$$

$$\mathcal{A}'(\mathsf{GPS}_n(\mathsf{RGT}(n,p,p'))) \approx_{o_n(1)} \mathsf{GPS}_n^g(G(n,p))\,.$$

Moreover, the parameter mappings f and g satisfy $f \circ g(q) = q + o_n(1)$ and $g \circ f(q) = q + o_n(1)$.

Remark 5. GPS includes Subgraph Stochastic Block Model (SSBM), which has a rank-1 signal (i.e., a small SBM) on k vertices planted inside an Erdős-Rényi graph. Its complete phase diagram was determined by [24], with computationally hard region obtained via reduction from the planted clique conjecture. Our result shows that for $k = o(n^{1/4}\log^{-17/4} n)$, the computational complexity of SSBM is the same in ambient graph being Erdős-Rényi or RGT.

b) Approximation of Random Intersection Graph: Aside from RGT being intrinsically interesting as a natural random graph model with low-order dependence, it turns out that the RGT non-trivially approximates the random intersection graph.

Definition 3 (Random Intersection Graph)**.** A sample from $\mathsf{RIG}(n, d, \delta)$ is defined by sampling n sets $S_1, \cdots, S_n \subset [d]$ where each S_i includes each element of $[d]$ independently with probability δ. Vertices i and j are connected in G if and only if S_i and S_j have non-empty intersection.

Theorem I.3. *For an RIG with constant edge density, i.e., $d = \Theta(1/\delta^2)$, if $d \gg n^{2.5}$, then*

$$d_{\mathsf{TV}}(\mathsf{RIG}(n, d, \delta), \mathsf{RGT}(n, p, p')) = o(1)\,,$$

where $p = 1 - e^{-d\delta^2 + (n-2)d\delta^3(1-\delta)^{n-3}}$ and $p' = 1 - e^{-d\delta^3(1-\delta)^{n-3}}$.

It was shown in [45] that the threshold for distinguishing RIG from Erdős-Rényi occurs at $d \sim n^3$. Thus, the RGT is close in total variation in the range $n^{2.5} \ll d \ll n^3$, whereas in this range the RIG and Erdős-Rényi have total variation distance close to 1.

II. General Result and Application to RGT

We start by introducing a general theorem that shows how a mapping between unplanted distributions yields a mapping also in the planted case if the mapping satisfies a certain *perturbation invariance*. We then describe our specific mappings between Erdős-Rényi and RGT. After that, in the following section, we give the main ideas for showing that our mappings satisfy the conditions of our general perturbation theorem, which constitutes the bulk of the paper's technical innovation.

A. Reductions and Sensitivity to Perturbation

We introduce a novel and general framework for validity of a reduction from one general graph model to another. The general result will be used to show that planted dense subgraph is computationally equivalent in Erdős-Rényi and in RGT. In this section, the graph models and transformations are abstract and specific constructions will be discussed in Section II-B.

Let P_0 and $\mathsf{P}_0{}'$ be two distributions of random graphs on vertex set $[n]$ and consider the general planted signal $\mathsf{GPS}_{\vec{e},\vec{p}}$ on both random graphs.

Suppose that $\mathcal{A}$ is a randomized algorithm (equivalently, a Markov kernel) satisfying $\mathcal{A}(\mathsf{P}_0) = \mathsf{P}_0{}'$. The following theorem gives a general sufficient condition for $\mathcal{A}$ to approximately correctly map the *planted distribution* $\mathsf{P}_1 := \mathsf{GPS}_{\vec{e},\vec{p}}(\mathsf{P}_0)$ to the planted distribution $\mathsf{P}_1' := \mathsf{GPS}_{\vec{e},\vec{q}}(\mathsf{P}_0{}')$. For each $0 \le i \le K$, let P_{1i} (or P_{1i}') be the planted version of P_0 (or $\mathsf{P}_0{}'$), defined by starting with P_0 (or $\mathsf{P}_0{}'$) and resampling edges $e_1, e_2, \cdots, e_i \in \binom{[n]}{2}$, respectively, with probability $p_1, p_2 \cdots, p_i$ (or $q_1, q_2 \cdots, q_i$).

Theorem II.1. *Let $\mathcal{G}$ be a set of graphs on n vertices such that for any $0 \le i \le K$, $\mathbb{P}_{G\sim\mathsf{P}_{1i}}(G \in \mathcal{G}) \ge 1 - \delta$. Let $\mathcal{A}$ be a randomized mapping between graphs on n vertices satisfying $\mathcal{A}(\mathsf{P}_0) = \mathsf{P}_0{}'$. Suppose that for each $e \in \binom{[n]}{2}$, there exists $p_-^e, p_+^e \in [0,1]$ such that for every $G \in \mathcal{G}$:*

C1 *Presence of edge e in the input to $\mathcal{A}$ has little influence on the other edges in the output of $\mathcal{A}$:*

$$\mathcal{A}(G - e)|_{\sim e} \approx_\epsilon \mathcal{A}(G + e)|_{\sim e}\,.$$

C2 *In the output graph, edge e is approximately independent of the rest of the edges:*

$$\mathcal{A}(G - e) \approx_\epsilon \mathcal{A}(G - e)|_{\sim e} \times \mathcal{A}(G - e)|_e \quad \textit{and}$$

$$\mathcal{A}(G + e) \approx_\epsilon \mathcal{A}(G + e)|_{\sim e} \times \mathcal{A}(G + e)|_e\,.$$

C3 *The marginal probability on edge e is approximately constant as a function of the input graph:*

$$\left|\,\mathbb{P}(e \in \mathcal{A}(G - e)) - p_-^e\,\right| \le \epsilon \quad \textit{and}$$

$$\left|\,\mathbb{P}(e \in \mathcal{A}(G + e)) - p_+^e\,\right| \le \epsilon\,.$$

Then we have

$$d_{\mathsf{TV}}\big(\mathcal{A}(\mathsf{GPS}_{\vec{e},\vec{p}}(\mathsf{P}_0)), \mathsf{GPS}_{\vec{e},\vec{q}}(\mathsf{P}_0{}')\big) = O(K(\epsilon + \delta))\,,$$

where $q_i = p_i p_+^{e_i} + (1 - p_i)p_-^{e_i}$ for each $i \in [K]$.

The main technical challenge of this paper is to prove that our proposed mapping $\mathcal{A}$ from RGT to Erdős-Rényi satisfies conditions **C1** and **C2**. The mapping is given in Section II-B.

Remark 6. It turns out that the conditions in the theorem are equivalent (up to constant factors) to the following more compact conditions:

$$d_{\mathsf{TV}}\big(\mathcal{A}(G - e), \mathcal{A}(G - e)|_{\sim e} \times \mathsf{Bern}(p_-^e)\big) \le \epsilon \quad \text{and}$$

$$d_{\mathsf{TV}}\big(\mathcal{A}(G + e), \mathcal{A}(G - e)|_{\sim e} \times \mathsf{Bern}(p_+^e)\big) \le \epsilon\,.$$

We state Theorem II.1 in the form above because it corresponds to how we apply it.

1) Proof of Theorem II.1 via Single Edge Lemma: The idea is to resample edges of the input one at a time and bound the effect of each step on the output of $\mathcal{A}$. Let Res_e^p be the operation of resampling edge e to be present with probability p.

Lemma II.1 (Single Edge Lemma)**.** *Under the conditions of Theorem II.1, for any $1 \le i \le K$,*

$$d_{\mathsf{TV}}\big(\mathcal{A} \circ \mathsf{Res}_{e_i}^{p_i}(\mathsf{P}_{1\,i-1}), \mathsf{Res}_{e_i}^{q_i} \circ \mathcal{A}(\mathsf{P}_{1\,i-1})\big) = O(\epsilon + \delta)\,.$$

Thus, the operations of applying $\mathcal{A}$ and resampling edge e_i *approximately commute*, with p_i being replaced by q_i.

Note that $\mathsf{P}_{1\,i} = \mathsf{Res}_{e_i}^{p_i}\mathsf{P}_{1\,i-1}$, so Lemma II.1 equivalently states that

$$d_{\mathsf{TV}}\big(\mathcal{A}(\mathsf{P}_{1\,i}), \mathsf{Res}_{e_i}^{q_i} \circ \mathcal{A}(\mathsf{P}_{1\,i-1})\big) = O(\epsilon + \delta)\,.$$

Applying the triangle inequality K times and using the last display for each term yields

$$d_{\mathsf{TV}}\big(\mathsf{Res}_{e_1}^{q_1} \circ \cdots \circ \mathsf{Res}_{e_K}^{q_K} \circ \mathcal{A}(\mathsf{P}_0), \mathcal{A}(\mathsf{P}_{1\,K})\big) = O(K(\epsilon+\delta))\,.$$

Note that $\mathcal{A}(\mathsf{P}_0) = \mathsf{P}_0{}'$, and by definition $\mathsf{Res}_{e_1}^{q_1} \cdots \mathsf{Res}_{e_K}^{q_K}\mathsf{P}_0{}' = \mathsf{P}_{1\,K}{}'$, so Theorem II.1 is proved. It remains to prove Lemma II.1.

B. Mapping between Erdős-Rényi and RGT

To use Theorem II.1 to relate ordinary PDS to the version in RGT, we first need to specify transformations between the ambient graph models. Our mapping from $\mathsf{G}(n,p)$ to $\mathsf{RGT}(n,p,p')$ simply applies the definition of RGT.

Definition 4 (Forward Transition)**.** Given any graph G, let $\mathcal{A}_\triangle(G)$ be the graph obtained from G by the following process: independently for each set of three vertices, add the three edges between them with probability p'. This defines a Markov transition kernel on the space of graphs.

By the definition of random graph with triangles, $\mathcal{A}_\triangle$ transfers $\mathsf{G}(n,p)$ to $\mathsf{RGT}(n,p,p')$. The reverse transition is more complicated. We will first describe the distribution of the set of triangles that were added to G_0, conditioned on observing $G = \mathcal{A}_\triangle(G_0)$. Let $X \in \{0,1\}^{\binom{[n]}{3}}$ be an indicator of a set of triangles. We use $|X|$ to denote the size of this set, $E(X)$ to denote the set of edges included in at least one of the triangles, and $\mathsf{e}(X) = |E(X)|$ to denote the total number of edges.

Definition 5 (Triangle Distribution)**.** For a given graph G, the *triangle distribution* μ_G is a distribution over subsets of triangles $x \in \{0,1\}^{\binom{[n]}{3}}$. The probability mass function μ_G is given by

$$\mu_G(x) = \frac{1}{Z_G}\Big(\frac{p'}{1-p'}\Big)^{|x|} p^{-\mathsf{e}(x)}\,,$$

if $E(x) \subseteq G$ and $\mu_G(x) = 0$ if $E(x) \not\subseteq G$. Here $Z_G = \sum_{x: E(x)\subset G}\big(\frac{p'}{1-p'}\big)^{|x|} p^{-e(x)}$ normalizes μ_G.

It will be convenient to let $Y \in \{0,1\}^{\binom{[n]}{2}}$ be the indicator vector of the edge set $E(X)$,

$$Y_e = \mathbb{1}\{e \in E(X)\} \quad \text{for each } e \in \binom{[n]}{2}\,.$$

To distinguish between different graphs, we use $\mathcal{L}_G(Y)$ to denote the law of Y for a given graph G.

Definition 6 (Reverse Transition)**.** Given any graph G, let $G' = \mathcal{A}(G)$ be the graph obtained from G by the following process:

1) Sample a set of triangles $X \sim \mu_G$
2) Let G' be equal to G on the set $E(X)^c$
3) For each $e \in E(X)$, include e in G' independently with probability p.

Remark 7. The reverse transition is only analyzed for $p' = 1/(n\log n)$, not for any $p' \le 1/(n\log n)$. Nonetheless, we can still construct a reverse map that works for any $p' \le 1/(n\log n)$ by first adding triangles to increase p' to $1/(n\log n)$, and then applying $\mathcal{A}$. This is formalized in Corollary 7.1 of the full version.

Lemma II.2. *The reverse transition $\mathcal{A}$ maps $\mathsf{RGT}(n,p,p')$ to the Erdős-Rényi $\mathsf{G}(n,p)$ distribution.*

For the purpose of reasoning about polynomial-time algorithms, it is crucial that our reduction $\mathcal{A}$ can be implemented in polynomial time. Fortunately, producing a sample $X \sim \mu_G$ can be done efficiently via the Glauber dynamics Markov chain.

Lemma II.3. *For any fixed graph G over n vertices, $p' \ll 1/n$ and constant p, the Glauber dynamics on μ_G mixes in $O(n^3 \log n)$ time.*

The definition of Glauber dynamics and proof of the lemma are in Section 9 of the full paper.

C. Applying Theorem II.1 to Triangle Removal Algorithm

Having Theorem II.1 and the transformations in hand, we are ready to prove Theorem I.1. In this section we focus on the reverse transition, $\mathcal{A}$, as it contains a wider range of technical ideas. The proof for the forward transition $\mathcal{A}_\triangle$, provided in Section 5 of the full paper,

is significantly easier and follows the same high-level outline.

To begin, for the unplanted case Lemma II.2 states that $\mathcal{A}(\mathsf{RGT}(n,p,p')) = G(n,p)$, so it remains to prove that $\mathcal{A}(\mathsf{RGT}(n,p,p',k,q)) \approx_{o_n(1)} G(n,p,k,g(q))$. We focus here on the case $p' = \tilde{\Theta}(1/n)$.

Theorem II.2 (Triangle Removal in Planted Case)**.** *For $k = o(n^{1/4}\log^{-17/4} n)$, $p' = 1/(n\log n)$ and $0 < p < q < 1$ being constant,*

$$d_{\mathsf{TV}}\big(\mathcal{A}(\mathsf{RGT}(n,p,p',S,q)), G(n,p,S,q\cdot p_e)\big) = o_n(1)\,,$$

where $p_e = \mathbb{E}_{G\sim\mathsf{RGT}(n,p,p')}\,\mathbb{P}_{\mathcal{A}}(e \in \mathcal{A}(G+e))$ for an arbitrary edge e.

Here $\mathsf{RGT}(n,p,p',S,q)$ stands for the random graph generated by planting a dense subgraph $G(k,q)$ at vertex set S in $\mathsf{RGT}(n,p,p')$. Similarly, $G(n,p,S,q\cdot p_e)$ stands for the random graph generated by planting a dense subgraph $G(k,q\cdot p_e)$ at vertex set S in $G(n,p)$.

To apply our general Theorem II.1, we need to specify two items: the intermediate planted RGT models and the class of graphs $\mathcal{G}$. We will then check conditions **C1**, **C2**, and **C3**.

1) Intermediate Planted Models: Define $\mathsf{RGT}_i(n,p,p'S,q)$ to be a random graph generated by starting with $\mathsf{RGT}(n,p)$ and independently resampling each edge $e_1, e_2, \cdots, e_i$ to be included with probability q.

2) Defining Class of Graphs $\mathcal{G}$: It is not hard to devise examples of graphs G for which each of the conditions **C1**, **C2**, **C3** are violated. We define the class of graphs $\mathcal{G}$ to avoid these bad examples.

The class of graphs $\mathcal{G}$ is chosen to be $\mathcal{G}_1 \cap \mathcal{G}_2$, where these are as follows:

- $\mathcal{G}_1$ is the set of graphs that are $p^2/3$-uniformly 2-star dense, where a graph is *c-uniformly 2-star dense* if for any pair of nodes i,j, there are at least $c(n-2)$ nodes k such that both (i,k) and (j,k) are in the graph; and
- $\mathcal{G}_2$ is the set of graphs G that satisfy for every $e \in \binom{[n]}{2}$ that

$$\big|\mu_{G+e}(Y_e = 1) - p_e^+\big| = C_p n^{-5/2} p'^{-2}\sqrt{\log n}\,,$$

where $p_e^+ = \mathbb{E}_{G'\sim\mathsf{RGT}(n,p,p')}[\mu_{G'+e}(Y_e=1)]$, and C_p is a fixed constant depending on p.

The following lemma states that $\mathcal{G}$ is a high probability set for each RGT_i as required by Theorem II.1. It is proved in Section 7.5 of the full paper.

Lemma II.4. *For any RGT_i, $\mathbb{P}_{G\sim\mathsf{RGT}_i}(G\in\mathcal{G}) = 1 - o(1/n)$.*

3) Checking the Conditions of Theorem II.1: We now state the main lemma needed to verify the conditions in Theorem II.1. The lemma shows insensitivity of $\mathcal{A}$ to perturbing the input by a single edge, and contains the bulk of our technical contributions. The key ideas will be presented in Section II-D with the full proof deferred to Section 7 of the full paper.

Lemma II.5 (Perturbation Insensitivity)**.** *If $G\in\mathcal{G}_1$, then*

$$\begin{aligned}\mathcal{L}_{G-e}(Y_{\sim e}) &= \mathcal{L}_{G+e}(Y_{\sim e}|Y_e = 0)\\ &\approx_{O(\log^{17/2} n/\sqrt{n})} \mathcal{L}_{G+e}(Y_{\sim e}|Y_e=1)\,,\end{aligned}$$

and from this it follows that

$$\mathcal{L}_{G+e}(Y) \approx_{O(\log^{17/2} n/\sqrt{n})} \mathcal{L}_{G+e}(Y_{\sim e})\times\mathcal{L}_{G+e}(Y_e)\,.$$

We first check that condition **C1** follows from the lemma. Since $\mathcal{L}_{G+e}(Y_{\sim e})$ is a mixture of the laws $\mathcal{L}_{G+e}(Y_{\sim e}|Y_e=0)$ and $\mathcal{L}_{G+e}(Y_{\sim e}|Y_e=1)$, Lemma II.5 implies $\mathcal{L}_{G+e}(Y_{\sim e}) \approx_{\tilde{O}(1/\sqrt{n})} \mathcal{L}_{G-e}(Y_{\sim e})$. Note that $\mathcal{A}$ simply resamples edges in Y, so by the data processing inequality for TV,

$$\mathcal{A}(G+e)|_{\sim e} \approx_{\tilde{O}(1/\sqrt{n})} \mathcal{A}(G-e)|_{\sim e}\,.$$

Condition **C2** for $G-e$ is trivial: $\mathcal{A}(G-e)$ has no edge in position e, so for any G,

$$\mathcal{A}(G-e) = \mathcal{A}(G-e)|_{\sim e}\times\mathsf{Bern}(0)\,.$$

For $G+e$, the second display of Lemma II.5 implies, again by data processing inequality, that

$$\mathcal{A}(G+e) \approx_{\tilde{O}(1/\sqrt{n})} \mathcal{A}(G+e)|_{\sim e}\times\mathcal{A}(G+e)|_e\,.$$

Lastly, condition **C3** is immediate for $G-e$, since $\mathcal{A}(G-e)|_e = \mathsf{Bern}(0)$ for all G. For $G+e$, it follows from the definition of $\mathcal{G}_2$: For any $G\in\mathcal{G}_2$, we have by conditioning on the value of Y_e that

$$\begin{aligned}&\mathcal{A}(G+e)|_e\\ &\sim \mu_{G+e}(Y_e=1)\cdot\mathsf{Bern}(p) + [1-\mu_{G+e}(Y_e=1)]\cdot\mathsf{Bern}(1)\\ &= \mathsf{Bern}\big(1-(1-p)\mu_{G+e}(Y_e=1)\big)\\ &\approx_{\tilde{O}(1/\sqrt{n})} \mathsf{Bern}\Big(1-(1-p)\cdot\mathop{\mathbb{E}}_{G'\sim\mathsf{RGT}(n,p,p')}[\mu_{G'+e}(Y_e=1)]\Big)\,.\end{aligned}$$

The last step used the fact that $d_{\mathsf{TV}}(\mathsf{Bern}(a),\mathsf{Bern}(b)) = |a-b|$. Let $p_e = \mathbb{E}_{G'\sim\mathsf{RGT}(n,p,p')}\,\mathbb{P}(e\in\mathcal{A}(G'+e)) = 1-(1-p)\cdot\mathbb{E}_{G'\sim\mathsf{RGT}(n,p,p')}[\mu_{G'+e}(Y_e=1)]$. Combining the last two displayed equations,

$$\mathcal{A}(G+e) \approx_{\tilde{O}(1/\sqrt{n})} \mathcal{A}(G+e)|_{\sim e}\times\mathsf{Bern}(p_e)\,.$$

Therefore, the conditions of Theorem II.1 hold with $q' = q\cdot p_e$, proving Theorem II.2.

D. Showing Perturbation Insensitivity: Main Ideas Behind Lemma II.5

Let us examine more closely the generation of variables in Lemma II.5. We fix G and let $X \sim \mu_{G+e}(\cdot|Y_e = 0) = \mu_{G-e}$ and $X^+ \sim \mu_{G+e}(\cdot|Y_e = 1)$, where recall that Y and Y^+ are the corresponding edge indicator vectors as defined in Section II-B. Without loss of generality, assume $e \notin G$, so $G-e = G$. Our objective is to show that $Y_{\sim e}$ and $Y^+_{\sim e}$ are close in total variation.

Let $X_{T(e)}$, $X^+_{T(e)}$, and $X_{\sim T(e)}$, $X^+_{\sim T(e)}$ be the triangle indicator vectors restricted to the set of triangles that contain and do not contain e, respectively. The proof of Lemma II.5 is divided into two conceptual parts: (1) insensitivity to perturbing Y_e of triangles $X_{\sim T(e)}$ that do not include e, and (2) conditioning on $X_{\sim T(e)}$, addressing the difference of edges brought about by triangles $X_{T(e)}$ that include e. We decompose $d_{\mathsf{TV}}(Y_{\sim e}, Y^+_{\sim e})$ into two terms (using Lemma 4.3 in Section 4.3 of the full paper):

$$\begin{aligned} &d_{\mathsf{TV}}(Y_{\sim e}, Y^+_{\sim e}) \le d_{\mathsf{TV}}(X_{\sim T(e)}, X^+_{\sim T(e)}) \\ &+ d_{\mathsf{TV}}\big(Y_{\sim e}, \mathop{\mathbb{E}}_{X' \sim X_{\sim T(e)}} \mathcal{L}(Y^+_{\sim e}|X^+_{\sim T(e)} = X')\big)\,. \end{aligned} \tag{1}$$

We next describe how to bound each of the two terms on the right-hand side.

1) Edge e Has Low Influence on Non-Containing Triangles: We first bound $d_{\mathsf{TV}}(X_{\sim T(e)}, X^+_{\sim T(e)})$. Only in Section II-D1 we will pretend that $X^+ \sim \mu_{G+e}$ rather than $\mu_{G+e}(\cdot|Y_e = 1)$, and this turns out to be valid as the two distributions have roughly the same variation distance to $X_{\sim T(e)}$, since μ_{G+e} is a constant-weight mixture of $\mu_{G+e}(\cdot|Y_e = 1)$ and μ_{G-e}.

Note that because $e \notin G$, $X_{T(e)}$ is always 0, so $X_{\sim T(e)}$ has the same distribution as X, defined by μ_G. As for $X^+_{\sim T(e)}$, letting T_e be the set of triangles containing e,

$$\mathcal{L}(X^+_{\sim T(e)}) = \sum_{x \in \{0,1\}^{T_e}} \mathbb{P}(X^+_{T(e)} = x)\mathcal{L}(X^+_{\sim T(e)}|X^+_{T(e)} = x)$$

is a mixture of Gibbs distributions indexed by the value of $X^+_{T(e)}$. By Pinsker's inequality we have $d_{\mathsf{TV}}(X_{\sim T(e)}, X^+_{\sim T(e)})^2 \le \chi^2(X_{\sim T(e)} \| X^+_{\sim T(e)})$, and we bound the latter quantity via Ingster's 2nd moment method [46]. We emphasize that both $\mathcal{L}(X_{\sim T(e)})$ and $\mathcal{L}(X^+_{\sim T(e)})$ are complicated dependent distributions, while all prior works to the best of our knowledge have always shown bounds between mixtures of *product distributions* (see, for instance, [28], [45]–[47]). We show that despite this dependence it is still possible to derive a tractable bound.

Lemma II.6 (χ^2-divergence for mixture of Gibbs measures)**.** *Let P be a distribution defined by*

$$P(X) = f(X)/Z.$$

Let U be a discrete random variable and Q be a mixture of Gibbs distributions defined by

$$Q(X) = \mathop{\mathbb{E}}_{U}[f_U(X)/Z_U]\,.$$

Letting $\rho_U(X) = f_U(X)/f(X)$, we have that

$$\chi^2(Q\|P) + 1 = \mathop{\mathbb{E}}_{U,U'} \frac{\mathbb{E}_{X\sim P}\, \rho_U(X)\rho_{U'}(X)}{\mathbb{E}_{X\sim P}\, \rho_U(X)\, \mathbb{E}_{X\sim P}\, \rho_{U'}(X)}\,,$$

where U' is an independent and identically distributed copy of U.

It turns out that when applying Lemma II.6 to $X_{\sim T(e)}$ versus $X^+_{\sim T(e)}$, we can bound the χ^2-divergence via marginal influences of the edge distribution $\mathcal{L}(Y)$. Here the marginal influence $\mathbf{I}^M_{A'\to e}$ characterizes how much configurations on A' can affect the conditional marginal probability on edge e. A formal definition of marginal influence can be found in Defn. 9 in Section 6 of the full paper.

Lemma II.7. *Suppose $G \in \mathcal{G}$ and consider $\mathcal{L}_G(Y)$ for Y as in Section II-B and $p' = o(1/n)$. Let $\mathbf{I}^M_{A'\to e}$ be the marginal influence of A' on e for $\mathcal{L}_G(Y)$. Letting $\tilde{X}^+_{T(e)}$ be an independent copy of $X^+_{T(e)}$, we have that $\chi^2(X^+_{\sim T(e)}\|X_{\sim T(e)})$ is upper bounded by*

$$\Big(1 + \frac{1-p}{p} \sup_{\substack{A' \subset A\cup B \\ e \in (A\cup B)\setminus A'}} \mathbf{I}^M_{A'\to e}\Big)^{|B|} (1/p)^{|A\cap B|} - 1$$

and

$$(1/p)^{|A|+|B|} - 1\,,$$

where $A = E(\tilde{X}^+_{T(e)}) - e$ and $B = E(X^+_{T(e)}) - e$.

The marginal influence $\mathbf{I}^M_{A'\to e}$ can be bounded by exploiting the fact that μ_G has small marginal probabilities under arbitrary conditioning, and we prove a general bound to this effect in Section 8 of the full paper.

Lemma II.8. *Suppose $G \in \mathcal{G}$ and consider $\mathcal{L}_G(Y)$ for Y as in Section II-B and $p' = o(1/n)$. Let $\mathbf{I}^M_{A\to e}$ be the marginal influence of A on e for $\mathcal{L}_G(Y)$. For an edge set $A \subset E(G)$ with $|A| = O(n)$,*

$$\mathbf{I}^M_{A\to e} = \tilde{O}(|A|/n)\,.$$

With high probability, $|A|$ and $|B|$ each have size $\tilde{O}(1)$ and $|A \cap B| = 0$, so the right hand of Lemma II.7 is $1 + \tilde{O}(1/n)$ with high probability. Of course, the

tail distribution of $|A|$ and $|B|$ is important, and we will make the bound rigorous in Section 7 of the full paper to get that $\chi^2(X_{\sim T(e)} \| X^+_{\sim T(e)}) = \tilde{O}(1/n)$. By Pinsker's inequality, this shows $d_{\mathsf{TV}}(X_{\sim T(e)}, X^+_{\sim T(e)}) = \tilde{O}(1/\sqrt{n})$.

2) Triangles Including e and TV Between Projections of Distributions: Recalling the decomposition (1), we now aim to bound the second term,

$$d_{\mathsf{TV}}\big(Y_{\sim e}, \mathop{\mathbb{E}}_{X' \sim X_{\sim T(e)}} \mathcal{L}(Y^+_{\sim e} | X^+_{\sim T(e)} = X')\big)\,. \tag{2}$$

This can be rewritten in a more symmetric way as the TV distance between $\mathbb{E}_{X' \sim X_{\sim T(e)}} \mathcal{L}(Y_{\sim e} | X_{\sim T(e)} = X')$ and $\mathbb{E}_{X' \sim X_{\sim T(e)}} \mathcal{L}(Y^+_{\sim e} | X^+_{\sim T(e)} = X')$. Since $Y_{\sim e}|X$ and $Y^+_{\sim e}|X^+$ are projections of triangle variables X and X^+ onto the edge space, the most natural approach to establish their closeness is to show that $\mathcal{L}(X_{T(e)}|X_{\sim T(e)})$ and $\mathcal{L}(X^+_{T(e)}|X^+_{\sim T(e)})$ are close and appeal to data processing inequality. However, this is not true: As mentioned earlier, $X_{T(e)} = \vec{0}$ since $e \notin G$, and in contrast $X^+_{T(e)}$ is non-zero with non-negligible probability, since there are $\Theta(n)$ triangles containing to e and each is selected with probability approximately $p' = \tilde{\Theta}(1/n)$.

Nevertheless, the fact that we are proving a statement about projection onto the edge space, $Y_{\sim e}$ and $Y^+_{\sim e}$, allows us to carry out manipulations in the triangle space before projection. We will design an auxiliary distribution X^{aux} over the same support as $X_{\sim T(e)}$ that when added to $X^+_{\sim T(e)}$ results in the *identical* edge projection as X^+_e, i.e.,

$$E(X^{\mathsf{aux}} \vee X^+_{\sim T(e)}) - e = E(X^+_{T(e)} \vee X^+_{\sim T(e)}) - e\,.$$

This means the edge indicator vector of $X^{\mathsf{aux}} \vee X^+_{\sim T(e)}$, which we denote by $\tilde{Y}_{\sim e}$, is the same as $Y^+_{\sim e}$. Thus, instead of comparing $Y_{\sim e}$ and $Y^+_{\sim e}$ in (2), it suffices to compare $Y_{\sim e}$ and $\tilde{Y}_{\sim e}$, and by data processing inequality it in turn suffices to compare $X_{\sim T(e)}$ and $X^+_{\sim T(e)} \vee X^{\mathsf{aux}}$:

$$d_{\mathsf{TV}}\big(Y_{\sim e}, Y^+_{\sim e}\big) \le d_{\mathsf{TV}}\big(X_{\sim T(e)}, X^+_{\sim T(e)} \vee X^{\mathsf{aux}}\big) \tag{3}$$

We now define X^{aux}. Each triangle in $X^+_{T(e)}$ adds at most two edges to $Y^+_{\sim e}$. To simulate this change without using any triangles containing e, we add a triangle to X^{aux} for each new edge introduced by $X^+_{T(e)}$. Crucially, this is done without adding other new edges. Avoiding adding new edges is possible with high probability as long as the graph G is sufficiently dense (here c-uniformly 2-star dense plays a role) and $p' = \tilde{\Theta}(1/n)$, which implies that the edge set of $X^+_{\sim T(e)}$ has sufficient coverage of relevant triangles that we might potentially add to X^{aux}.

From the previous section, we have $X_{\sim T(e)} \approx_{\tilde{O}(1/\sqrt{n})} X^+_{\sim T(e)}$, so to show (3) it remains to prove that $X_{\sim T(e)} \approx X_{\sim T(e)} \vee X^{\mathsf{aux}}$, i.e., the addition of X^{aux} must be undetectable. We show this via concentration of the likelihood ratio between $X_{\sim T(e)} \vee X^{\mathsf{aux}}$ and $X_{\sim T(e)}$. In Section 8 of the full paper, we prove that Lipschitz functions over μ_G concentrate. The likelihood ratio under consideration is not quite Lipschitz – it is Lipschitz only over a high probability subset – but this turns out to suffice for it to concentrate.

References

[1] S. B. Hopkins, "Statistical inference and the sum of squares method," Ph.D. dissertation, Cornell University, 2018.

[2] A. S. Wein, "Average-case complexity of tensor decomposition for low-degree polynomials," *arXiv preprint arXiv:2211.05274*, 2022.

[3] B. Barak, S. B. Hopkins, J. Kelner, P. Kothari, A. Moitra, and A. Potechin, "A nearly tight sum-of-squares lower bound for the planted clique problem," in *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*. IEEE, 2016, pp. 428–437.

[4] S. B. Hopkins, P. K. Kothari, A. Potechin, P. Raghavendra, T. Schramm, and D. Steurer, "The power of sum-of-squares for detecting hidden structures," *Proceedings of the fifty-eighth IEEE Foundations of Computer Science*, pp. 720–731, 2017.

[5] V. Feldman, E. Grigorescu, L. Reyzin, S. Vempala, and Y. Xiao, "Statistical algorithms and a lower bound for detecting planted cliques," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 2013, pp. 655–664.

[6] V. Feldman, W. Perkins, and S. Vempala, "On the complexity of random satisfiability problems with planted solutions," pp. 77–86, 2015.

[7] I. Diakonikolas, D. M. Kane, and A. Stewart, "Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures," in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2017, pp. 73–84.

[8] M. Celentano, A. Montanari, and Y. Wu, "The estimation error of general first order methods," in *Conference on Learning Theory*. PMLR, 2020, pp. 1078–1141.

[9] B. Rossman, "On the constant-depth complexity of k-clique," in *Proceedings of the fortieth annual ACM symposium on Theory of computing*. ACM, 2008, pp. 721–730.

[10] ——, "The monotone complexity of k-clique on random graphs," *SIAM Journal on Computing*, vol. 43, no. 1, pp. 256–279, 2014.

[11] D. Gamarnik and M. Sudan, "Limits of local algorithms over sparse random graphs," in *Proceedings of the 5th conference on Innovations in theoretical computer science*. ACM, 2014, pp. 369–376.

[12] B. Huang and M. Sellke, "Tight lipschitz hardness for optimizing mean field spin glasses," in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2022, pp. 312–322.

[13] M. Brennan, G. Bresler, S. B. Hopkins, J. Li, and T. Schramm, "Statistical query algorithms and low-degree tests are almost equivalent," *arXiv preprint arXiv:2009.06107*, 2020.

[14] A. Montanari and A. S. Wein, "Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation," *arXiv preprint arXiv:2212.06996*, 2022.

[15] F. Koehler and E. Mossel, "Reconstruction on trees and low-degree polynomials," *Advances in Neural Information Processing Systems*, vol. 35, pp. 18 942–18 954, 2022.

[16] A. S. Wein, A. El Alaoui, and C. Moore, "The kikuchi hierarchy and tensor pca," in *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2019, pp. 1446–1468.

[17] I. Zadik, M. J. Song, A. S. Wein, and J. Bruna, "Lattice-based methods surpass sum-of-squares in clustering," in *Conference on Learning Theory*. PMLR, 2022, pp. 1247–1248.

[18] O. Goldreich, "Notes on Levin's theory of average-case complexity," in *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*. Springer, 2011, pp. 233–247.

[19] B. Barak, *The Complexity of Public-Key Cryptography*. Cham: Springer International Publishing, 2017, pp. 45–77. [Online]. Available: https://doi.org/10.1007/978-3-319-57048-8_2

[20] A. Bogdanov and L. Trevisan, "Average-case complexity," *Foundations and Trends® in Theoretical Computer Science*, vol. 2, no. 1, pp. 1–106, 2006.

[21] Q. Berthet and P. Rigollet, "Complexity theoretic lower bounds for sparse principal component detection." in *COLT*, 2013, pp. 1046–1066.

[22] T. Wang, Q. Berthet, and R. J. Samworth, "Statistical and computational trade-offs in estimation of sparse principal components," *The Annals of Statistics*, vol. 44, no. 5, pp. 1896–1930, 2016.

[23] C. Gao, Z. Ma, and H. H. Zhou, "Sparse cca: Adaptive estimation and computational barriers," *The Annals of Statistics*, vol. 45, no. 5, pp. 2074–2101, 2017.

[24] M. Brennan, G. Bresler, and W. Huleihel, "Reducibility and computational lower bounds for problems with planted sparse structure," in *COLT*, 2018, pp. 48–166.

[25] M. Brennan and G. Bresler, "Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness," *arXiv preprint arXiv:1902.07380*, 2019.

[26] Z. Ma and Y. Wu, "Computational barriers in minimax submatrix detection," *The Annals of Statistics*, vol. 43, no. 3, pp. 1089–1116, 2015.

[27] M. Brennan, G. Bresler, and W. Huleihel, "Universality of computational lower bounds for submatrix detection," *arXiv preprint arXiv:1902.06916*, 2019.

[28] B. E. Hajek, Y. Wu, and J. Xu, "Computational lower bounds for community detection on random graphs." in *COLT*, 2015, pp. 899–928.

[29] M. Brennan and G. Bresler, "Reducibility and statistical-computational gaps from secret leakage," in *Conference on Learning Theory*. PMLR, 2020, pp. 648–847.

[30] Y. Luo and A. R. Zhang, "Tensor clustering with planted structures: Statistical optimality and computational limits," *The Annals of Statistics*, vol. 50, no. 1, pp. 584–613, 2022.

[31] A. Zhang and D. Xia, "Tensor svd: Statistical and computational limits," *arXiv preprint arXiv:1703.02724*, 2017.

[32] J. Bruna, O. Regev, M. J. Song, and Y. Tang, "Continuous lwe," in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, 2021, pp. 694–707.

[33] A. Gupte, N. Vafa, and V. Vaikuntanathan, "Continuous lwe is as hard as lwe & applications to learning gaussian mixtures," in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, 2022, pp. 1162–1173.

[34] M. J. Song, I. Zadik, and J. Bruna, "On the cryptographic hardness of learning single periodic neurons," *Advances in neural information processing systems*, vol. 34, pp. 29 602–29 615, 2021.

[35] S. Chen, A. Gollakota, A. Klivans, and R. Meka, "Hardness of noise-free learning for two-hidden-layer neural networks," *Advances in Neural Information Processing Systems*, vol. 35, pp. 10 709–10 724, 2022.

[36] I. Diakonikolas, D. Kane, P. Manurangsi, and L. Ren, "Cryptographic hardness of learning halfspaces with massart noise," *Advances in Neural Information Processing Systems*, vol. 35, pp. 3624–3636, 2022.

[37] S. Tiegel, "Hardness of agnostically learning halfspaces from worst-case lattice problems," in *The Thirty Sixth Annual Conference on Learning Theory*. PMLR, 2023, pp. 3029–3064.

[38] I. Diakonikolas, D. Kane, and L. Ren, "Near-optimal cryptographic hardness of agnostically learning halfspaces and relu regression under gaussian marginals," in *International Conference on Machine Learning*. PMLR, 2023, pp. 7922–7938.

[39] A. Daniely and G. Vardi, "From local pseudorandom generators to hardness of learning," in *Conference on Learning Theory*. PMLR, 2021, pp. 1358–1394.

[40] A. Daniely, N. Srebro, and G. Vardi, "Efficiently learning neural networks: What assumptions may suffice?" *arXiv preprint arXiv:2302.07426*, 2023.

[41] A. Bhaskara, M. Charikar, E. Chlamtac, U. Feige, and A. Vijayaraghavan, "Detecting high log-densities: an $o(n^{1/4})$ approximation for densest k-subgraph," *Proceedings of the forty-second ACM symposium on Theory of computing*, pp. 201–210, 2010.

[42] U. Bhaskar, Y. Cheng, Y. K. Ko, and C. Swamy, "Hardness results for signaling in bayesian zero-sum and network routing games," in *Proceedings of the 2016 ACM Conference on Economics and Computation*. ACM, 2016, pp. 479–496.

[43] B. Applebaum, B. Barak, and A. Wigderson, "Public-key cryptography from different assumptions," in *Proceedings of the forty-second ACM symposium on Theory of computing*. ACM, 2010, pp. 171–180.

[44] M. S. Granovetter, "The strength of weak ties," *American journal of sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[45] M. Brennan, G. Bresler, and D. Nagaraj, "Phase transitions for detecting latent geometry in random graphs," *arXiv preprint arXiv:1910.14167*, 2019.

[46] Y. Ingster and I. Suslina, *Nonparametric goodness-of-fit testing under Gaussian models*. Springer Science & Business Media, 2003, vol. 169.

[47] Q. Berthet and P. Rigollet, "Optimal detection of sparse principal components in high dimension," *The Annals of Statistics*, vol. 41, no. 4, pp. 1780–1815, 2013.