Is Planted Coloring Easier than Planted Clique?

Pravesh K. Kothari PRAVESHK@CS.CMU.EDU
Computer Science Department, Carnegie Mellon University

Santosh S. Vempala VEMPALA@GATECH.EDU
School of Computer Science, Georgia Tech

Alexander S. Wein ASWEIN@UCDAVIS.EDU
Department of Mathematics, University of California, Davis

Jeff Xu JEFFXUSICHAO@CMU.EDU
Computer Science Department, Carnegie Mellon University

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

We study the computational complexity of two related problems: recovering a planted q-coloring in $G(n, 1/2)$, and finding efficiently verifiable witnesses of non-q-colorability (a.k.a. refutations) in $G(n, 1/2)$. Our main results show hardness for both these problems in a restricted-but-powerful class of algorithms based on computing low-degree polynomials in the inputs.

The problem of recovering a planted q-coloring is equivalent to recovering q disjoint planted cliques that cover all the vertices — a potentially easier variant of the well-studied planted clique problem. Our first result shows that this variant is as hard as the original planted clique problem in the low-degree polynomial model of computation: each clique needs to have size $k \gg \sqrt{n}$ for efficient recovery to be possible. For the related variant where the cliques cover a $(1-\epsilon)$-fraction of the vertices, we also show hardness by reduction from planted clique.

Our second result shows that refuting q-colorability of $G(n, 1/2)$ is hard in the low-degree polynomial model when $q \gg n^{2/3}$ but easy when $q \lesssim n^{1/2}$, and we leave closing this gap for future work. Our proof is more subtle than similar results for planted clique and involves constructing a non-standard distribution over q-colorable graphs. We note that while related to several prior works, this is the first work that explicitly formulates refutation problems in the low-degree polynomial model.

The proofs of our main results involve showing low-degree hardness of hypothesis testing between an appropriately constructed pair of distributions. For refutation, we show *completeness* of this approach: in the low-degree model, the refutation task is precisely as hard as the hardest associated testing problem, i.e., proving hardness of refutation amounts to finding a "hard" distribution.

Keywords: Random graphs, coloring, low-degree polynomials, computational complexity

1. Introduction

The *planted clique* problem, introduced by Jerrum (1992) and Kučera (1995), asks for a polynomial-time algorithm to find a clique of size k added to an Erdős–Rényi random graph $G(n, 1/2)$. The associated task of *refuting* the existence of k-cliques in $G \sim G(n, 1/2)$ asks for a polynomial-time algorithm to compute a certificate that can be efficiently verified to infer the absence of a k-clique in G. Despite a long line of work, state-of-the-art polynomial-time algorithms for both problems (Alon et al., 1998) only succeed when $k = \Omega(\sqrt{n})$. In contrast, the clique number of $G(n, 1/2)$ is at most $\lceil 2\log_2 n \rceil + 1$ with high probability and thus, an added clique of any size $k > \lceil 2\log_2 n \rceil + 1$ is uniquely identifiable. A long line of work proving lower bounds in various restricted models

such as Markov chains (Jerrum, 1992), the Statistical Query model (Feldman et al., 2017), convex relaxations (Feige and Krauthgamer, 2003) and in particular the sum-of-squares hierarchy and the related low-degree polynomial model of computation (Barak et al., 2016; Hopkins, 2018), suggest that the *algorithmic threshold* for both variants — the smallest k for which efficient algorithms can find the added k-clique or refute the existence of a k-clique in $G(n, 1/2)$ — is $\Omega(\sqrt{n})$. In the past two decades, the hypothesis that no polynomial-time procedure can beat the above guarantees of the known algorithms has become a focal point in average-case complexity theory and the root of myriad reductions to average-case problems arising in various domains (e.g., Berthet and Rigollet (2013); Hajek et al. (2015); Brennan et al. (2018); Kothari and Mehta (2018)).

Two motivating problems: recovery and refutation of q-colorings. In this paper, we study the following innocuous-looking (and ostensibly easier than planted clique) question where, in the *recovery* problem, we study the complexity of exactly recovering $\approx n/k$ disjoint planted k-cliques in $G \sim G(n, 1/2)$, with high success probability. If the disjoint planted cliques cover all the vertices of the graph, then the complement of the graph has a planted (n/k)-coloring. Thus, this version of our problem is tantamount to studying whether recovering a *planted q-coloring* in $G(n, 1/2)$ is easier than recovering a single planted clique. In the associated *refutation* problem, the goal is to find an algorithm that takes as input a graph G and outputs NO or MAYBE with the guarantee that (1) whenever it outputs NO, the graph must not admit a valid q-coloring of its vertices, and (2) when $G \sim G(n, 1/2)$, the algorithm should output NO with probability $1 - o(1)$ over the draw of G. For reference, the chromatic number of $G(n, 1/2)$ is $\Theta(n/\log n)$ with high probability (see Heckel (2018)), so the *information-theoretic* threshold for refutation is $q = \Theta(n/\log n)$. Intuitively, the recovery and refutation problems for coloring appear easier than their clique counterparts because the planted structure is more prominent and therefore seemingly easier to find (or refute).

The relation between the recovery and refutation tasks is somewhat subtle: while these two problems appear related, we are not aware of a formal reduction between them in either direction. In this paper, we study the recovery and refutation problems separately, and draw attention to the fact that rather different methods will be needed to prove lower bounds in the two settings. We note that for colorability of *sparse* random regular graphs, there appears to be a constant-factor gap between the recovery and refutation thresholds (Bandeira et al., 2021).

We note that for simplicity we consider the *exact* recovery problem. One can also consider various notions of approximate recovery. This tends not to make a difference in our setting because once a small fraction of the vertices in a clique are known, it is easy to find the rest by examining common neighbors.

Proof strategy: hypothesis testing. One common strategy to understand the complexity of recovery or refutation is to introduce an auxiliary *hypothesis testing* task: given a graph G that is sampled either from some "null" distribution $\mathbb{Q}$ (e.g., $G(n, 1/2)$) or some "planted" distribution $\mathbb{P}$ (e.g., some distribution supported on q-colorable graphs), design an efficiently computable statistical test that decides which of the two distributions generated a given sample G, with high success probability over the draw of G. Note that if there is an efficient refutation algorithm for some distribution $\mathbb{Q}$, then we immediately obtain an efficient distinguisher between $\mathbb{Q}$ and *any* distribution $\mathbb{P}$ supported on q-colorable graphs. Similarly, if there is an efficient recovery algorithm for some distribution $\mathbb{P}$, then we immediately obtain a distinguishing algorithm between $\mathbb{P}$ and *any* distribution $\mathbb{Q}$ supported on non-q-colorable graphs. As a result of this connection, we can conclude:

(I) To show computational hardness of exactly recovering a planted q-coloring in a particular planted distribution $\mathbb{P}$, it suffices to construct a null distribution $\mathbb{Q}$ such that (i) with high probability, $G \sim \mathbb{Q}$ is not q-colorable and (ii) it is computationally hard to distinguish $\mathbb{P}$ from $\mathbb{Q}$.

(II) To show computational hardness of refuting q-colorability for a particular null distribution $\mathbb{Q}$, it suffices to construct a planted distribution $\mathbb{P}$ such that (i) $\mathbb{P}$ is supported on q-colorable graphs and (ii) it is computationally hard to distinguish $\mathbb{P}$ from $\mathbb{Q}$.

Note that we have flexibility to choose either $\mathbb{Q}$ (if studying recovery) or $\mathbb{P}$ (if studying refutation). We will see later that it can be a non-trivial task to construct the right distribution. It need not be the case that the same testing problem arises when studying recovery as when studying refutation.

Strategy (II) has been referred to as constructing a *computationally quiet planted distribution* (Bandeira et al., 2020), where "quiet" pertains to the fact that the planted structure's presence cannot be detected by an efficient algorithm. Similarly, strategy (I) corresponds to constructing a *computationally quiet null distribution*.

Since proving lower bounds for average-case hypothesis testing problems based on standard hardness assumptions is an elusively difficult goal at present (notwithstanding the recent successes of Brennan et al. (2018); Brennan and Bresler (2020) that use the hardness of planted clique and its variants as a starting point in certain limited settings), we will obtain evidence of hardness for testing problems by focusing on a restricted but powerful and well-studied family of tests that we next describe.

Low-degree testing. The low-degree polynomial model of hypothesis testing restricts the class of tests to be polynomial functions in a natural representation of the input, with the complexity of a test captured by the degree of the polynomial. Specifically, viewing graphs as elements of $\{-1,1\}^{\binom{n}{2}}$ with a $\{\pm 1\}$-indicator of presence or absence of every possible edge, the low-degree polynomial tests informally correspond to computing thresholds of arbitrary degree-D polynomials of the edge-indicator variables. Since degree-D polynomials can be computed (when described in the monomial coefficient representation) in time $n^{O(D)}$, constant-degree tests yield polynomial-time distinguishing algorithms. Despite being restricted, these low-degree tests already capture tests based on basic statistics of graphs such as edge counts, triangle counts, and more generally small subgraph counts (the number of edges in the subgraph corresponds to the degree of the polynomial). Various spectral methods (e.g., the leading eigenvalue of the adjacency matrix, or some other symmetric matrix whose entries are low-degree polynomials of the input variables) can also be approximated by polynomial tests of logarithmic degree in the number of variables; see Kunisky et al. (2022), Section 4.2.3. As a result, low-degree tests (with degree $O(\log n)$) already capture the best known polynomial-time algorithms for a wide variety of high-dimensional statistical testing tasks (although we won't attempt to precisely characterize which tasks here; see e.g. Hopkins et al. (2017); Hopkins (2018); Kunisky et al. (2022); Holmgren and Wein (2021); Zadik et al. (2022) for discussion). As a result, if we manage to establish that all degree-D tests provably fail to solve a particular testing problem for some $D = \omega(\log n)$, we say the problem is "low-degree hard." This can be viewed as evidence suggesting computational hardness of the hypothesis testing problem. This is a widely-applicable and by now, commonly-used framework that originated in a line of work on proving lower bounds against the sum-of-squares hierarchy (Barak et al., 2016; Hopkins

and Steurer, 2017; Hopkins et al., 2017); see also Hopkins (2018); Kunisky et al. (2022) for further exposition.

Summary of results. Our main results use strategies (I) and (II) described above to shed light on the computational complexity of recovery and refutation of q-coloring. The formal models and statements are presented in the next section, but here we give a brief overview. Throughout, we will implicitly assume an asymptotic regime $n \to \infty$ where other parameters (e.g., q, k) may scale with n. We say an event occurs "with high probability (w.h.p.)" if it has probability $1 - o(1)$ as $n \to \infty$. Since our focus is on identifying computational thresholds up to the correct power of n, we use the symbol $\ll$ in our informal discussions to hide factors of $n^{o(1)}$.

Our main result for the recovery problem shows that adding $\approx n/k$ disjoint cliques of size k (instead of a single one) does not make the problem of recovering the added planted cliques easier. That is, our lower bounds suggest that each added clique needs to be of size $\gtrsim \sqrt{n}$ for efficient recovery to be possible.

In contrast and perhaps surprisingly, it turns out that adding more cliques makes the problem of *distinguishing* the planted graph from $G(n, 1/2)$ easier, simply by counting the total number of edges. This reveals a *detection-recovery gap*, in contrast to the single planted clique problem (see Section 2.1.2).

More precisely, our results for recovery are as follows:

- In the planted partial-coloring model where some fraction of the vertices are colored (equivalently, many disjoint planted cliques in $G(n, 1/2)$ that cover at most a $(1 - \epsilon)$-fraction of the graph), we show that:
  - (i) If each clique has size $k \gg \sqrt{n}$, a simple algorithm can be used to recover them.
  - (ii) If each clique has size $k \ll \sqrt{n}$, it is computationally hard to recover them assuming the Planted Clique Hypothesis. That is, recovering many planted k-cliques is as hard as recovering a single planted k-clique.
- In the full planted coloring model (q planted cliques of size $k = n/q$ partitioning the entire graph), we are unable to show hardness via reduction, but instead give an indirect argument that supports the same conclusion as above:
  - (i) If each clique has size $k \gg \sqrt{n}$, there is again a simple algorithm to recover them.
  - (ii) If each clique has size $k \ll \sqrt{n}$, we argue that recovery is computationally hard via strategy (I), taking the null distribution $\mathbb{Q}$ to be a planted $(q + 1)$-coloring. In other words, we prove that low-degree tests cannot even distinguish a planted q-colorable graph from a planted $(q + 1)$-colorable graph. This suggests hardness of recovery via a two-stage argument described in Section 2.1.3.

For the problem of refuting q-colorability in $G(n, 1/2)$, it is known that a poly-time algorithm exists when $k := n/q \gg \sqrt{n}$ (Coja-Oghlan, 2005). To explore the complexity of this problem, we explicitly formulate the refutation problem in the low-degree polynomial model (for the first time), and show the following:

- If $k \gtrsim \sqrt{n}$ (i.e., $q \lesssim \sqrt{n}$), then there is a low-degree polynomial that refutes q-colorability in $G(n, 1/2)$.

- If $k \ll n^{1/3}$ (i.e., $q \gg n^{2/3}$), then no low-degree polynomial refutes q-colorability in $G(n, 1/2)$. The proof follows strategy (II) and involves constructing a non-trivial planted distribution $\mathbb{P}$.
- We conjecture $k \sim \sqrt{n}$ is the true low-degree refutation threshold, and we leave this to future work. One way to improve the lower bound would be to construct a "quieter" planted distribution, i.e., a distribution supported on q-colorable graphs that is low-degree hard to distinguish from $G(n, 1/2)$ whenever $k \ll \sqrt{n}$. Our final result is a duality argument showing that in fact, the conjecture is *equivalent* to the existence of such a planted distribution.

2. Results

A central concept in this work will be that of *hypothesis testing* between two high-dimensional distributions. We consider two (sequences of) distributions $\mathbb{P} = \mathbb{P}_n$ and $\mathbb{Q} = \mathbb{Q}_n$. For us, these distributions will always be over n-vertex graphs. We use the following asymptotic notion of successful testing.

Definition 1 (Strong distinguishing) *For two distributions $\mathbb{P}_n$ and $\mathbb{Q}_n$, we say an algorithm A_n* strongly distinguishes *$\mathbb{P}$ and $\mathbb{Q}$ if it takes as input a sample drawn from one of the two distributions and correctly determines which distribution it came from with probability $1 - o(1)$ as $n \to \infty$. In other words, both type I and type II error probabilities must be $o(1)$.*

We will also be interested in the following class of "low-degree" tests. A degree-D test is simply a (multivariate) polynomial in the input variables (or rather a sequence of such polynomials, one for each problem size n). In our case, there will be $\binom{n}{2}$ input variables — one for every possible edge in an n-vertex graph — taking values in $\{\pm 1\}$, where $+1$ indicates the presence of an edge and -1 indicates the absence. We use the following standard notion of "success" for a polynomial test.

Definition 2 (Strong/weak separation of distributions) *Suppose $\mathbb{P}_n$ and $\mathbb{Q}_n$ are distributions on $\mathbb{R}^N$ for some $N = N_n$. A polynomial $f_n : \mathbb{R}^N \to \mathbb{R}$ is said to* strongly separate *$\mathbb{P}$ and $\mathbb{Q}$ if, as $n \to \infty$,*

$$\sqrt{\max\left\{\operatorname*{Var}_{\mathbb{Q}}[f], \operatorname*{Var}_{\mathbb{P}}[f]\right\}} = o\left(\left|\operatorname*{\mathbb{E}}_{\mathbb{P}}[f] - \operatorname*{\mathbb{E}}_{\mathbb{Q}}[f]\right|\right),$$

and weakly separate *$\mathbb{P}$ and $\mathbb{Q}$ if*

$$\sqrt{\max\left\{\operatorname*{Var}_{\mathbb{Q}}[f], \operatorname*{Var}_{\mathbb{P}}[f]\right\}} = O\left(\left|\operatorname*{\mathbb{E}}_{\mathbb{P}}[f] - \operatorname*{\mathbb{E}}_{\mathbb{Q}}[f]\right|\right).$$

Note that strong separation implies that $\mathbb{P}$ and $\mathbb{Q}$ can be strongly distinguished by thresholding the value of the polynomial f. Weak separation implies that the output of f can be used to distinguish better than random guessing; see Bandeira et al. (2022b), Proposition 6.1.

In our case, the input variables will take values in $\{\pm 1\}$ and so the polynomial f can be multilinear without loss of generality.

If all degree-D polynomials *fail* to strongly separate $\mathbb{P}$ and $\mathbb{Q}$ for some $D = \omega(\log n)$, we say the testing problem is "low-degree hard." As explained in the introduction, this can be viewed as evidence for inherent computational hardness of strong distinguishing.

Proofs that rule out strong or weak separation typically proceed by bounding the *advantage*, defined below:

$$\mathsf{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) := \sup_{f \in \mathbb{R}[Y]_{\leq D}} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}}, \tag{1}$$

where $\mathbb{R}[Y]_{\leq D}$ denotes the set of polynomials $\mathbb{R}^N \to \mathbb{R}$ of degree (at most) D. It is well known that $\mathsf{Adv}_{\leq D}$ also admits a characterization as the *norm of the low-degree likelihood ratio*; see Hopkins (2018); Kunisky et al. (2022). If $\mathsf{Adv}_{\leq D} = O(1)$ then strong separation is impossible, and if $\mathsf{Adv}_{\leq D} = 1 + o(1)$ then weak separation is impossible (see Lemma 22).

2.1. Recovery

2.1.1. MODELS

The primary objective of this section will be to understand the recovery problem in two related models for planted coloring and planted partial-coloring. As explained in the introduction, the complement of a q-colorable graph is partitioned into q cliques. To fix notation and compare with the standard planted clique model, we will take the clique perspective here. Thus we study the problem of multiple cliques planted in $G(n, 1/2)$.

The first model $\mathsf{MC}(n, q)$ ("multiple cliques") corresponds to a true planted coloring, i.e., the cliques partition the entire graph.

Definition 3 *In the model $\mathsf{MC}(n, q)$, we observe an n-vertex graph where each vertex is independently assigned a uniformly random label from $[q] := \{1, 2, \ldots, q\}$. Vertices with the same label are always connected, and vertices with different label are connected with probability $1/2$. Given the graph, the goal is to exactly recover the clique partition with probability $1 - o(1)$ as $n \to \infty$, where $q = q_n$ may scale with n.*

The next model is a variation for partial coloring, i.e., the cliques do not partition the entire graph. In the coloring viewpoint, some fraction of the vertices do not belong to any color class (and have no constraints on the colors of their neighbors). One motivation for defining this model is that it is a variant of the original model where we will be able to prove a strong form of hardness via reduction. For technical convenience, the cliques in this model have exactly the same size, unlike $\mathsf{MC}(n, q)$.

Definition 4 *In the model $\mathsf{MC}(n, q, \delta)$, we observe an n-vertex graph where $(1 - \delta)n$ vertices are partitioned into q cliques, each of size exactly $k := (1-\delta)n/q$ (which we assume is an integer). Two vertices in the same clique are always connected, and all remaining edges occur independently with probability $1/2$. Given the graph, the goal is to exactly recover the clique partition (and identify the non-clique vertices) with probability $1 - o(1)$ as $n \to \infty$, where the parameters $q = q_n$ and $\delta = \delta_n$ may scale with n.*

2.1.2. HARDNESS OF PLANTED PARTIAL-COLORING VIA REDUCTION

We now consider the recovery problem in $\mathsf{MC}(n, q, \delta)$. First, we observe that a simple algorithm based on examining degrees and common neighbors can exactly recover the cliques when $k \gg \sqrt{n}$. This matches (up to log factors) the best known algorithms for recovering a single planted k-clique in $G(n, 1/2)$.

Theorem 5 (Upper bound) *If q, δ scale with n such that $k := (1-\delta)n/q = \omega(\sqrt{n \log n})$ then there is a polynomial-time algorithm achieving exact recovery w.h.p. in $\mathsf{MC}(n, q, \delta)$.*

Proof We will use the following standard version of Bernstein's inequality: for independent random variables $X_1, \dots, X_n$ satisfying $\mathbb{E}[X_i] = 0$ and $|X_i| \le M$ almost surely, we have for any $t \ge 0$ that

$$\Pr\left(\sum_{i=1}^{n} X_i \ge t\right) \le \exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^{n} \mathrm{Var}(X_i) + \frac{1}{3}Mt}\right).$$

Fix an arbitrary sequence $\alpha_n = \omega(1)$. The degree d_i of a non-clique vertex i has a binomial distribution $d_i \sim \mathrm{Bin}(n-1, 1/2)$, which by Bernstein's inequality satisfies $d_i \le \frac{n}{2} + \alpha\sqrt{n \log n}$ with probability $1 - n^{-\omega(1)}$. On the other hand, a clique vertex i has degree $d_i \sim (k-1) + \mathrm{Bin}(n-k, 1/2)$, which by Bernstein's inequality satisfies $d_i \ge \frac{n+k}{2} - \alpha\sqrt{n \log n}$ with probability $1 - n^{-\omega(1)}$. By thresholding degrees, this lets us perfectly classify the non-clique vertices with probability $1 - o(1)$, provided $k = \omega(\sqrt{n \log n})$.

It remains to partition the clique vertices. If vertices i, j are in different cliques, their number of common neighbors is $d_{ij} \sim \mathrm{Bin}(2(k-1), 1/2) + \mathrm{Bin}(n-2k, 1/4)$, which satisfies $d_{ij} \le \frac{n}{4} + \frac{k}{2} + \alpha\sqrt{n \log n}$ with probability $1 - n^{-\omega(1)}$. If vertices i, j instead belong to the same clique, their number of common neighbors is $d_{ij} \sim (k-2) + \mathrm{Bin}(n-k, 1/4)$, which satisfies $d_{ij} \ge \frac{n}{4} + \frac{3k}{4} - \alpha\sqrt{n \log n}$ with probability $1 - n^{-\omega(1)}$. By thresholding common neighbors, this allows us to exactly recover the clique partition with probability $1 - o(1)$, again provided $k = \omega(\sqrt{n \log n})$. ■

We next show a matching lower bound: computational hardness of recovering the cliques when $k \ll \sqrt{n}$. This result will be conditional on the *Planted Clique Hypothesis*, a conjecture that is commonly used as the basis for deducing average-case hardness results. In the *planted clique model* $\mathsf{PC}(N, K)$, an N-vertex graph has a clique on K vertices, and all other edges occur independently with probability $1/2$. The following version of the conjecture appears, for instance, as Conjecture 2.1 in Brennan et al. (2018).

Conjecture 6 (Planted Clique Hypothesis) *If $K = K_N$ scales as $K \le N^{1/2-\Omega(1)}$ then no sequence of randomized polynomial-time algorithms B_N can strongly distinguish (Definition 1) between $\mathsf{PC}(N, K)$ and $G(N, 1/2)$.*

Assuming this conjecture, we have the following hardness result for $\mathsf{MC}(n, q, \delta)$.

Theorem 7 (Lower bound) *Assume the Planted Clique Hypothesis (Conjecture 6). If q, δ scale with n such that $k := (1-\delta)n/q$ satisfies $(2 + \Omega(1)) \log_2 n \le k \le (\delta n)^{1/2-\Omega(1)}$ then no sequence of randomized polynomial-time algorithms A_n achieves exact recovery w.h.p. in $\mathsf{MC}(n, q, \delta)$.*

The condition $(2 + \Omega(1)) \log_2 n \le k$ is natural because $2 \log_2 n$ is the size of the maximum clique in $G(n, 1/2)$. To satisfy the condition $k \le (\delta n)^{1/2-\Omega(1)}$, it suffices to have $k = n^{\frac{1}{2}-\Omega(1)}$ and $\delta = n^{-o(1)}$.

The reduction which proves Theorem 7 is very simple but (to our knowledge) has not appeared before in the literature. Intuitively, the idea is the following: in the multiple cliques model, even if

an oracle were to reveal the positions of all cliques but one, the remaining problem is still a hard instance of planted clique.

Proof Let q, δ scale as prescribed. Assume for the sake of contradiction that an algorithm A_n achieves exact recovery in $\mathsf{MC}(n, q, \delta)$. Let $K = k = (1-\delta)n/q$ and $N = K + \delta n$. Note that as $n \to \infty$ we have $N \to \infty$ because

$$N \geq K = k \geq (2 + \Omega(1)) \log_2 n \to \infty,$$

and also $K \leq N^{1/2-\Omega(1)}$ because

$$K = k \leq (\delta n)^{\frac{1}{2}-\Omega(1)} \leq N^{\frac{1}{2}-\Omega(1)}.$$

We will give an algorithm B_N achieving strong detection between $G(N, 1/2)$ and $\mathsf{PC}(N, K)$, contradicting the planted clique conjecture.

The algorithm B_N works as follows. Given an N-vertex graph, add $(q-1)k$ additional vertices (bringing the total to n), partitioned into $q-1$ cliques each of size k. Add all other edges (both among the new vertices and between the old and new vertices) independently with probability $1/2$. Now run A_n on the resulting graph. If it finds q disjoint cliques of size k and one of these cliques lies within the original N vertices, output "$\mathsf{PC}(N, K)$"; otherwise, output "$G(N, 1/2)$."

To argue correctness of B_N, first suppose the input came from $\mathsf{PC}(N, K)$. Then the n-vertex graph produced is exactly a sample from $\mathsf{MC}(n, q, \delta)$, and so A_n must correctly identify all the cliques with probability $1 - o(1)$, leading B_N to correctly answer "$\mathsf{PC}(N, k)$." Now suppose instead that the input to B_N came from $G(N, 1/2)$. Due to the assumption $k \geq (2 + \Omega(1)) \log_2 n \geq (2 + \Omega(1)) \log_2 N$, with probability $1 - o(1)$ there is no k-clique within the original N vertices, in which case B_N must correctly answer "$G(N, 1/2)$." ■

Remark 8 *We note that the Planted Clique Hypothesis also implies hardness of detecting a constant number of planted k-cliques in $G(n, 1/2)$ when $k \ll \sqrt{n}$. The idea is to first show by reduction from planted clique that distinguishing between q planted cliques and $(q+1)$ planted cliques is hard; the reduction is simply to add q new cliques (on new vertices). Then the classical "hybrid argument" implies that distinguishing between 0 and q cliques is hard for any constant q. (We thank Guy Bresler for pointing out this argument.)*

Detection-recovery gap. In the standard planted clique model (with a single clique), $k \sim \sqrt{n}$ is the best known threshold for both efficiently recovering the clique and efficiently "detecting" it, i.e., distinguishing the planted clique model from $G(n, 1/2)$. While we have shown that adding more cliques does not make recovery any easier, it certainly does make detection easier. For instance, in the extreme case where the cliques cover the whole graph, the total edge count strongly distinguishes $\mathsf{MC}(n, q)$ from $G(n, 1/2)$ provided $q = o(n)$. Thus, the multiple cliques problem exhibits a "detection-recovery gap" that is not present in the single clique case.

We remark that our reduction is a rare (perhaps unique?) example where a detection-recovery gap has been established based on the Planted Clique Hypothesis. For instance, the prior work Brennan et al. (2018) on various planted matrix and graph problems was only able to establish hardness of recovery in a regime where detection is easy if reducing from some starting problem (not planted clique) that is already conjectured to have a detection-recovery gap. While Cai et al. (2017) claims

to overcome this by reducing from planted clique to planted submatrix recovery, the argument is incorrect.[1]

Finally we note that the notion of a "detection-recovery gap" is arguably somewhat artificial in that it assumes we have chosen one "canonical" testing problem to associate with the recovery problem (a perspective we are avoiding in this paper). One might expect that the gap can be closed by choosing a different null distribution whose total edge count matches that of the planted distribution. It turns out that closing the gap is not quite this simple, and the challenge of constructing a better null distribution plays a key role in the next section.

2.1.3. TESTING q-COLORABILITY VERSUS $(q+\ell)$-COLORABILITY

The results of the previous section do not quite cover the case of a true coloring, i.e., where the cliques partition the entire graph. In this case, exact recovery remains easy when $k := n/q \gg \sqrt{n}$, and we expect it to be hard when $k \ll \sqrt{n}$; however, we do not know how to establish this via reduction from planted clique. We will instead follow strategy (I) from the introduction: we fix $\mathbb{P} = \mathsf{MC}(n,q)$ and our goal is to design a null distribution $\mathbb{Q}$ such that w.h.p. $G \sim \mathbb{Q}$ is not q-colorable (or rather, its complement is not), and distinguishing $\mathbb{P}$ versus $\mathbb{Q}$ is low-degree hard. Once we have achieved this goal, this gives an indirect two-stage argument for hardness of recovery: the low-degree hardness leads us to conjecture that no poly-time algorithm can distinguish $\mathbb{P}$ from $\mathbb{Q}$, and this conjecture (if true) formally implies that no poly-time algorithm can recover the cliques in $\mathbb{P}$.

Perhaps the first natural attempt is to choose $\mathbb{Q} = G(n,1/2)$. However, this will not suffice, as $G(n,1/2)$ is too easy to distinguish from $\mathsf{MC}(n,q)$ due to the detection-recovery gap discussed in the previous section. Instead, we will choose $\mathbb{Q} = \mathsf{MC}(n,q+1)$, which w.h.p. is not q-colorable for $q \le \Omega(n/\log n)$; see Appendix D. We will show that testing $\mathbb{P} = \mathsf{MC}(n,q)$ versus $\mathbb{Q} = \mathsf{MC}(n,q+1)$ is low-degree hard when $k := n/q \ll \sqrt{n}$. As discussed above, this suggests hardness of exact recovery in $\mathsf{MC}(n,q)$ when $k \ll \sqrt{n}$.

We will in fact consider a slightly more general testing problem: $\mathbb{P} = \mathsf{MC}(n,q)$ versus $\mathbb{Q} = \mathsf{MC}(n,q+\ell)$ for some $\ell \ge 1$ (which may scale with n). This generality will not cost us much, and we feel it is a question of possible independent interest. The following results establish that (in the low-degree framework) this problem is easy when $q^2 \ll \ell n$ and hard when $q^2 \gg \ell n$.

Theorem 9 (Upper bound) *If q, ℓ scale with n such that $1 \le q < q+\ell \le n$ and $q^2 = o(\ell n)$ then there is a degree-1 polynomial achieving strong separation between $\mathbb{P} = \mathsf{MC}(n,q)$ and $\mathbb{Q} = \mathsf{MC}(n,q+\ell)$.*

Theorem 10 (Lower bound) *Fix an arbitrary constant $\epsilon > 0$, not depending on n. If q, ℓ scale with n such that $1 \le q < q+\ell \le n$ and $q^2 \ge \ell n^{1+\epsilon}$ then there is no degree-$o(\log n/\log\log n)^2$ polynomial achieving weak separation between $\mathbb{P} = \mathsf{MC}(n,q)$ and $\mathbb{Q} = \mathsf{MC}(n,q+\ell)$.*

Testing planted versus planted. On a technical level, this result differs from nearly all existing low-degree lower bounds because here we are testing between two different "planted" distributions. In contrast, most prior work has considered testing between some planted distribution and

1. On pg 21-22 of Cai et al. (2017) (arXiv v2), the bootstrapping construction in Eq. (42) does not actually produce an instance of the submatrix model because the entries of the noise matrix are not mutually independent. An issue occurs near the top of pg 22, where pairwise independence does not imply mutual independence. The reduction does show hardness of some non-standard submatrix model where the noise entries are not mutually independent.

an i.i.d. null distribution, which is much easier to analyze. The first "planted-versus-planted" low-degree lower bounds were given recently by Rush et al. (2022), based on a technique developed by Schramm and Wein (2022). Our proof is based on similar ideas, but differs from Rush et al. (2022) on a technical level; the bounds for dense subgraph problems in Rush et al. (2022) do not work when the subgraph is extremely dense (e.g., a clique), and so we use a somewhat different variation of the argument.

The standard approach to proving low-degree lower bounds is based on relatively straightforward moment calculations, but relies heavily on knowing an orthogonal basis of polynomials with respect to $\mathbb{Q}$; see Section 2.3 of Hopkins (2018). The key technical challenge in planted-versus-planted testing is that, since $\mathbb{Q}$ is not a product measure, we do not know such an orthogonal basis of polynomials that is convenient to work with. Our Proposition 23 overcomes this, showing that it suffices to control certain recursively-defined quantities w_α. This generalizes the standard approach, as discussed in Remark 24. Similarly to Rush et al. (2022), the quantities w_α turn out to have a convenient multiplicative property (Lemma 25) which helps in the analysis. The proof of Proposition 23 takes an approach first used by Schramm and Wein (2022), where we apply Jensen's inequality to the "signal" but not the "noise," and then leverage an orthogonal basis of polynomials for the i.i.d. "noise."

We note that an alternative form of evidence for hardness of our original recovery problem would be to directly formulate a low-degree recovery question in the style of Schramm and Wein (2022), but we have chosen to instead investigate the quiet planting approach.

2.2. Refutation

A common framework for studying the average-case complexity of refutation problems is to prove lower bounds against the sum-of-squares (SoS) hierarchy, a powerful class of methods based on semi-definite programming. For the problem of refuting q-colorability, a particular SoS formulation is known to fail when $q \gg \sqrt{n}$ (Kothari and Manohar, 2021); however, it remains open to characterize the more canonical (and potentially stronger) SoS SDP which has equality constraints instead of inequalities (see Section 1.5 of Kothari and Manohar (2021)).

In this paper, we formulate an alternative type of refutation lower bound based directly on low-degree polynomials, which complements the SoS approach. Some advantages of the new formulation are its simplicity, and the fact that (unlike SoS) there is no ambiguity in the choice of SDP relaxation; we only need to specify how our input is encoded as real-valued variables. To our knowledge, there are no formal implications in either direction between SoS lower bounds and our new framework. Like SoS, our framework captures spectral methods (as illustrated by the proof of Theorem 16 below), a powerful class of refutation algorithms which give the best known poly-time algorithms for a wide variety of average-case refutation tasks.

We note that some prior work has used low-degree lower bounds to give evidence for hardness of refutation, via a two-stage argument that first gives a polynomial-time reduction from a testing problem to refutation (Bandeira et al., 2020, 2022a). Our new framework is similar in spirit but more direct, as we define for the first time a notion of what it means for a polynomial to solve a refutation problem (Definition 11).

2.2.1. FRAMEWORK FOR LOW-DEGREE REFUTATION

We will now define a notion (Definition 11) of what it means for a polynomial to refute a property $\mathcal{R} \subseteq \mathbb{R}^N$ (e.g., the set of q-colorable graphs $X \in \{\pm 1\}^{\binom{n}{2}}$) over a distribution $\mathbb{Q}$ (e.g., $G(n, 1/2)$). We will later argue that this definition is reasonable in that it indeed implies a solution to the refutation problem (Proposition 13). We also illustrate that our definition captures spectral methods, a powerful class of refutation algorithms (see the proof of Theorem 16).

Definition 11 (Strong/weak separation of a distribution and property) *Suppose $\mathbb{Q}_n$ is a distribution on $\mathbb{R}^N$ for some $N = N_n$, and suppose $\mathcal{R} = \mathcal{R}_n \subseteq \mathbb{R}^N$. A polynomial $f_n : \mathbb{R}^N \to \mathbb{R}$ is said to* strongly separate *$\mathbb{Q}$ and $\mathcal{R}$ if*

$$f(X) \geq 1 \quad \forall X \in \mathcal{R} \qquad \text{and} \qquad \mathop{\mathbb{E}}_{\mathbb{Q}}[f^2] = o(1),$$

and weakly separate *$\mathbb{Q}$ and $\mathcal{R}$ if*

$$f(X) \geq 1 \quad \forall X \in \mathcal{R} \qquad \text{and} \qquad \mathop{\mathbb{E}}_{\mathbb{Q}}[f] = 0, \ \mathop{\mathbb{E}}_{\mathbb{Q}}[f^2] = O(1).$$

Remark 12 *The requirement $\mathbb{E}_{\mathbb{Q}}[f] = 0$ can optionally be added to the definition of strong separation: if $f = f_n$ satisfies the original definition it can be shifted and scaled to satisfy the modified one.*

More generally, one could define separation to mean there exists $B = B_n > \mathbb{E}_{\mathbb{Q}}[f]$ such that $f(X) \geq B$ for all $X \in \mathcal{R}$, and $\sqrt{\mathrm{Var}_{\mathbb{Q}}[f]}$ is either $o(B - \mathbb{E}_{\mathbb{Q}}[f])$ (for strong separation) or $O(B - \mathbb{E}_{\mathbb{Q}}[f])$ (for weak separation). This is equivalent in the sense that if $f = f_n$ satisfies the original definition it also satisfies the new one with $B = 1$, and if f satisfies the new definition it can be shifted and scaled to satisfy the original one.

As we see next, strong and weak separation are natural sufficient conditions for refuting $\mathcal{R}$ with high probability or constant probability (respectively) by evaluating f.

Proposition 13 *Suppose f strongly (or weakly, respectively) separates $\mathbb{Q}$ and $\mathcal{R}$. Define a refutation algorithm that, on input $X \in \mathbb{R}^N$, outputs NO if $f(X) < 1$ and outputs MAYBE otherwise. Then this algorithm has the guarantee that (1) whenever it outputs NO, $X \notin \mathcal{R}$, and (2) when $X \sim \mathbb{Q}$, the output is NO with probability $1 - o(1)$ (or $\Omega(1)$, respectively).*

Proof Guarantee (1) is immediate from the property $f(X) \geq 1$ for all $X \in \mathcal{R}$. For strong separation, (2) follows because by Markov's inequality, $\mathbb{E}[f^2] = o(1)$ implies that $|f(X)| < 1$ with probability $1 - o(1)$. It remains to verify (2) for weak separation: letting $\mathbb{E}_{\mathbb{Q}}[f^2] \leq C$ and $p \coloneqq \Pr_{\mathbb{Q}}\{f(X) < 1\}$,

$$0 = \mathbb{E}[f] \geq 1 \cdot \Pr\{f \geq 1\} + \mathbb{E}[f \cdot \mathbb{1}_{f<1}] \geq (1-p) - \sqrt{\mathbb{E}[f^2]} \cdot \sqrt{p} \geq 1 - p - C\sqrt{p} \geq 1 - (C+1)\sqrt{p},$$

implying $p \geq 1/(C+1)^2$. ∎

In line with strategy (II) from the introduction, one way to rule out strong (or weak) separation is to construct a planted distribution and bound the quantity $\mathsf{Adv}_{\leq D}$ defined in (1).

Proposition 14 *Suppose that on an infinite subsequence of n values we have a distribution $\mathbb{P} = \mathbb{P}_n$ supported on $\mathcal{R}$. If $\mathsf{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) = O(1)$ (respectively, $1 + o(1)$) for some $D = D_n$, then no degree-D polynomial strongly (resp., weakly) separates $\mathbb{Q}$ and $\mathcal{R}$.*

Proof Since $\mathbb{P}$ is supported on $\mathcal{R}$, the separation condition implies $\mathbb{E}_{\mathbb{P}}[f] \geq 1$. The proof is now nearly identical to that of Lemma 22. ■

Remark 15 *We note that for the well-studied problem of refuting a single k-clique in $G(n, 1/2)$, existing work implies sharp upper and lower bounds in our new framework. For the lower bound, let $\mathbb{P}$ be the standard planted k-clique model and combine Proposition 14 with the low-degree analysis of planted clique (Hopkins, 2018, Section 2.4) to conclude: if $k \leq n^{1/2-\epsilon}$ for a constant $\epsilon > 0$ then no degree-$o(\log n/\log\log n)^2$ polynomial weakly separates $G(n, 1/2)$ from the property of containing a k-clique. The upper bound follows from the proof of Theorem 16 below: if $k \geq 2.1\sqrt{n}$ then there is an $O(\log n)$-degree polynomial that strongly separates $G(n, 1/2)$ from the property of containing a k-clique.*

2.2.2. LOW-DEGREE REFUTATION OF q-COLORABILITY

We now apply the framework from the previous section to the problem of refuting q-colorability in $G(n, 1/2)$. Throughout, we represent graphs as elements of $\{\pm 1\}^{\binom{n}{2}}$ as usual, take $\mathbb{Q} = G(n, 1/2)$, and use $\mathcal{R}_q \subseteq \{\pm 1\}^{\binom{n}{2}}$ to denote the property of q-colorability (i.e., the set of graphs that are q-colorable).

First, we give an upper bound: low-degree polynomials can refute q-colorability for $q \lesssim \sqrt{n}$. The proof proceeds by taking a standard spectral refutation algorithm (based on the maximum eigenvalue of the adjacency matrix) and approximating it by a polynomial.

Theorem 16 (Upper bound) *Suppose $q \leq b\sqrt{n}$ for a constant $b < 1/2$ (not depending on n). Then there exists a constant $C = C(b) > 0$ and a polynomial $f = f_n$ of degree at most $C \log n$ that strongly separates $G(n, 1/2)$ and $\mathcal{R}_q$.*

We also give a lower bound: no low-degree polynomial can refute q-colorability for $q \gg n^{2/3}$. Note there is a gap between our upper and lower bounds, and we leave closing this gap as an interesting direction for future work.

Theorem 17 (Lower bound) *If $q \geq n^{2/3+\epsilon}$ for a constant $\epsilon > 0$, then no degree-$o(\log n/\log\log n)^2$ polynomial weakly separates $G(n, 1/2)$ and $\mathcal{R}_q$.*

The proof of the lower bound will use Proposition 14, which is a rigorous incarnation of strategy (II) from the introduction. In other words, our goal is to construct a planted distribution $\mathbb{P}$ supported on q-colorable graphs that is hard to distinguish from $\mathbb{Q} = G(n, 1/2)$ in the sense $\mathsf{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) = 1 + o(1)$.

Constructing this planted distribution is non-trivial. The naive choice would be the "canonical" planted model $\mathsf{MC}(n, q)$ (or rather, its complement), but this is not a good choice because it can be easily distinguished from $G(n, 1/2)$ by counting the total number of edges whenever $q \ll n$. A next attempt is to modify $\mathsf{MC}(n, q)$ to have a slightly lower probability for non-clique edges so as to correct the total edge count. This gives a quieter planting that is hard to distinguish from $\mathbb{Q}$ when

$q \gg n^{3/4}$, but easy when $q \ll n^{3/4}$ by counting signed triangles (each of the $\binom{n}{3}$ triangles in the complete graph counts for $+1$ if an even number of its edges are present or -1 if an odd number are present). Our final construction, defined below, that reaches the threshold $q \sim n^{2/3}$, is more complicated and involves planting both cliques and independent sets.

Definition 18 (Quiet planting for $q \gg n^{2/3}$) *Suppose n, q are positive integers. To each of the n vertices, independently assign a label $(a, b) \in [q] \times [q]$ uniformly at random. Conditioned on the labels, do the following independently for each pair of distinct vertices $\{u, v\}$: denote the two vertex labels by (a_1, b_1) and (a_2, b_2); if $a_1 = a_2$ then do not include the edge (u, v); if $a_1 \neq a_2$ and $b_1 = b_2$ then include the edge (u, v); otherwise include the edge (u, v) with probability $1/2$.*

Note that all the vertices with a given a value form an independent set, and thus the distribution is supported on q-colorable graphs. Also, the vertices with a given b value nearly form a clique, aside from the non-edges required for the independent sets. In the proof of Theorem 17, we show that this distribution is low-degree indistinguishable from $G(n, 1/2)$ when $q \gg n^{2/3}$. Our analysis of this distribution is tight, as the count of signed 4-cycles distinguishes it from $G(n, 1/2)$ when $q \ll n^{2/3}$.

Although we have not proven it, we expect the true threshold for low-degree refutation of colorability to be $q \sim \sqrt{n}$.

Conjecture 19 *Fix an arbitrary $\epsilon > 0$, not depending on n. If $q \geq n^{1/2+\epsilon}$ then no degree-D polynomial weakly separates $\mathbb{Q} = G(n, 1/2)$ and $\mathcal{R}_q$, for some $D = \omega(\log n)$.*

2.2.3. COMPLETENESS OF THE QUIET PLANTING APPROACH

A natural approach to prove Conjecture 19 would be to construct a quieter planted distribution $\mathbb{P}$ that is supported on q-colorable graphs but hard to distinguish from $G(n, 1/2)$ when $q \gg \sqrt{n}$. One might worry, however, that this may not even be possible: conceivably, such a planted distribution might not exist, even if the true low-degree refutation threshold is at $q \sim \sqrt{n}$ like we expect. If this were the case, we would need to find an alternative approach to prove the conjecture without relying on quiet planting.

We show in high generality that the hypothetical scenario above actually cannot occur: for every low-degree hard refutation problem, there is a planted distribution that can be used to prove its hardness. Put another way, Conjecture 19 is *equivalent* to the existence of a quiet planted distribution for $q \gg \sqrt{n}$.

Theorem 20 *Fix sequences $N = N_n$, $D = D_n$, $\mathbb{Q} = \mathbb{Q}_n$ a distribution on $\mathbb{R}^N$, and $\mathcal{R} = \mathcal{R}_n \subseteq \mathbb{R}^N$. Assume that for each n, $\mathbb{Q}$ is supported on a finite set and $\mathcal{R}$ is a finite set (but the cardinality of these sets may depend on n). The following are equivalent:*

(1) No degree-D polynomial strongly separates $\mathbb{Q}$ and $\mathcal{R}$.

(2) For an infinite subsequence of n values, there exists a distribution $\mathbb{P} = \mathbb{P}_n$ supported on $\mathcal{R}$ such that $\mathsf{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) = O(1)$.

Similarly, the following are equivalent:

(1) No degree-D polynomial weakly separates $\mathbb{Q}$ and $\mathcal{R}$.

(2) *For an infinite subsequence of n values, there exists a distribution $\mathbb{P} = \mathbb{P}_n$ supported on $\mathcal{R}$ such that* $\mathsf{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) = 1 + o(1)$.

Note that we have already shown that (2) implies (1); see Proposition 14. The proof that (1) implies (2) uses von Neumann's min-max principle.

Remark 21 *We have assumed* $\mathrm{supp}(\mathbb{Q})$ *and $\mathcal{R}$ are finite (the relevant setting for q-coloring) to simplify the analytic conditions needed for the min-max principle, but these assumptions can be relaxed; see Remark 33.*

Adapted to the context of q-coloring, while formally we do not know whether there exists a low-degree polynomial to refute q-coloring when $q \gg \sqrt{n}$, it would be surprising in light of the sum-of-squares lower bound of Kothari and Manohar (2021) for refuting $\tilde{O}(\sqrt{n})$-colorability of $G(n, 1/2)$. Hence, we interpret this argument as suggesting the existence of a computationally quiet planted q-coloring for $G(n, 1/2)$ when $q \approx \sqrt{n}$ even though we do not know an explicit construction of such a distribution. If this construction were known, it may allow for SoS lower bounds in stronger SDP formulations to be proved via the *pseudo-calibration* approach (Barak et al., 2016).

Acknowledgments

P.K.K. was supported by NSF CAREER Award #2047933, Alfred P. Sloan Fellowship and a Google Research Scholar Award. S.S.V. was supported in part by NSF awards CCF-2007443 and CCF-2106444. Part of this work was done while A.S.W. was with the Algorithms and Randomness Center at Georgia Tech, supported by NSF awards CCF-2007443 and CCF-2106444. J.X. was supported in part by NSF CAREER Award #2047933.

We thank the anonymous reviewers for their helpful comments.

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.

Afonso Bandeira, Dmitriy Kunisky, and Alexander Wein. Average-case integrality gap for non-negative principal component analysis. In *Mathematical and Scientific Machine Learning*, pages 153–171. PMLR, 2022a.

Afonso S Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Annals of Probability*, 44(4):2479–2506, 2016.

Afonso S Bandeira, Dmitriy Kunisky, and Alexander S Wein. Computational hardness of certifying bounds on constrained PCA problems. In *11th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 151, 2020.

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

Afonso S Bandeira, Ahmed El Alaoui, Samuel B Hopkins, Alexander S Wein, Tselil Schramm, and Ilias Zadik. The Franz–Parisi criterion and computational trade-offs in high dimensional statistics. In *36th Conference on Neural Information Processing Systems (NeurIPS 2022)*, 2022b.

Boaz Barak, Samuel B. Hopkins, Jonathan Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 428–437, 2016. doi: 10.1109/FOCS.2016.53.

Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on learning theory*, pages 1046–1066. PMLR, 2013.

Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory*, pages 648–847. PMLR, 2020.

Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory*, pages 48–166. PMLR, 2018.

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):1403, 2017.

Amin Coja-Oghlan. The Lovász number of random graphs. *Comb. Probab. Comput.*, 14(4): 439–465, 2005. ISSN 0963-5483. doi: 10.1017/S0963548305006826. URL https://doi.org/10.1017/S0963548305006826.

Uriel Feige and Robert Krauthgamer. The probable value of the Lovász–Schrijver relaxations for maximum independent set. *SIAM J. Comput.*, 32(2):345–370, 2003. URL http://dblp.uni-trier.de/db/journals/siamcomp/siamcomp32.html#FeigeK03.

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.

Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory*, pages 899–928. PMLR, 2015.

Annika Heckel. The chromatic number of dense random graphs. *Random Structures & Algorithms*, 53(1):140–182, 2018.

Justin Holmgren and Alexander S Wein. Counterexamples to the low-degree conjecture. In *12th Innovations in Theoretical Computer Science Conference (ITCS)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

Samuel Hopkins. *Statistical Inference and the Sum of Squares Method*. PhD thesis, Cornell University, 2018.

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

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

Mark Jerrum. Large cliques elude the metropolis process. *Random Struct. Algorithms*, 3:347–360, 1992.

Hidetoshi Komiya. Elementary proof for Sion's minimax theorem. *Kodai mathematical journal*, 11 (1):5–7, 1988.

Pravesh K. Kothari and Peter Manohar. A stress-free sum-of-squares lower bound for coloring. In *Proceedings of the 36th Computational Complexity Conference*, CCC '21, Dagstuhl, DEU, 2021. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 9783959771931. doi: 10.4230/LIPIcs.CCC.2021.23. URL https://doi.org/10.4230/LIPIcs.CCC.2021.23.

Pravesh K. Kothari and Ruta Mehta. Sum-of-squares meets nash: Lower bounds for finding any equilibrium. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, page 1241–1248, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450355599. doi: 10.1145/3188745.3188892. URL https://doi.org/10.1145/3188745.3188892.

Luděk Kučera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57(2-3):193–212, 1995.

Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2022.

Cynthia Rush, Fiona Skerman, Alexander S Wein, and Dana Yang. Is it easier to count communities than find them? *arXiv preprint arXiv:2212.10872*, 2022.

Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022.

Maurice Sion. On general minimax theorems. *Pacific J. Math.*, 8(4):171–176, 1958.

Ilias Zadik, Min Jae Song, Alexander S Wein, and Joan Bruna. Lattice-based methods surpass sum-of-squares in clustering. In *Conference on Learning Theory*, pages 1247–1248. PMLR, 2022.

Appendix A. Testing the Number of Cliques

A.1. Upper Bound

We restate the theorem for the reader's convenience.

Theorem 9 (Upper bound) *If q, ℓ scale with n such that $1 \le q < q + \ell \le n$ and $q^2 = o(\ell n)$ then there is a degree-1 polynomial achieving strong separation between $\mathbb{P} = \mathsf{MC}(n, q)$ and $\mathbb{Q} = \mathsf{MC}(n, q + \ell)$.*

Proof Let f be the degree-1 polynomial that counts the total number of signed edges in the graph: $f(Y) = \sum_{1 \le i < j \le n} Y_{ij}$, where recall $Y_{ij} \in \{\pm 1\}$. Using linearity of expectation,

$$\mathop{\mathbb{E}}_{Y \sim \mathsf{MC}(n,q)} f(Y) = \binom{n}{2} \frac{1}{q}$$

and so

$$\left| \mathop{\mathbb{E}}_{Y \sim \mathbb{P}} f(Y) - \mathop{\mathbb{E}}_{Y \sim \mathbb{Q}} f(Y) \right| = \binom{n}{2} \left(\frac{1}{q} - \frac{1}{q + \ell} \right) = \binom{n}{2} \frac{\ell}{q(q + \ell)}. \tag{2}$$

For the second moment,

$$\mathop{\mathbb{E}}_{Y \sim \mathsf{MC}(n,q)} f(Y)^2 = \sum_{i<j} \sum_{i'<j'} \mathbb{E}[Y_{ij} Y_{i'j'}].$$

There are a few different terms to consider depending on how the edges (i, j) and (i', j') interact.

- If $(i, j) = (i', j')$ then $\mathbb{E}[Y_{ij} Y_{i'j'}] = \mathbb{E}[Y_{ij}^2] = 1$.
- If (i, j) and (i', j') have no vertices in common then Y_{ij} and $Y_{i'j'}$ are independent, and so $\mathbb{E}[Y_{ij} Y_{i'j'}] = \mathbb{E}[Y_{ij}] \, \mathbb{E}[Y_{i'j'}] = \frac{1}{q^2}$.
- If (i, j) and (i', j') have one vertex in common then we again have that Y_{ij} and $Y_{i'j'}$ are independent: if say $i = i'$ then the event that i, j have the same label is independent from the event that i, j' have the same label, due to symmetry among the possible labels for i. Therefore $\mathbb{E}[Y_{ij} Y_{i'j'}] = \frac{1}{q^2}$.

Putting it together,

$$\mathop{\mathbb{E}}_{Y \sim \mathsf{MC}(n,q)} f(Y)^2 = \binom{n}{2} \cdot 1 + \binom{n}{2} \left[\binom{n}{2} - 1 \right] \cdot \frac{1}{q^2}$$

and so

$$\begin{aligned} \mathop{\mathrm{Var}}_{Y \sim \mathsf{MC}(n,q)} f(Y) &= \binom{n}{2} \cdot 1 + \binom{n}{2} \left[\binom{n}{2} - 1 \right] \cdot \frac{1}{q^2} - \left[\binom{n}{2} \frac{1}{q} \right]^2 \\ &= \binom{n}{2} \left(1 - \frac{1}{q^2} \right) \\ &\le \binom{n}{2}. \end{aligned} \tag{3}$$

Combining (2) and (3), f achieves strong separation provided

$$\sqrt{\binom{n}{2}} = o\left(\binom{n}{2}\frac{\ell}{q(q+\ell)}\right), \qquad \text{i.e.,} \qquad q\left(\frac{q}{\ell}+1\right) = o(n).$$

It therefore suffices to have $q = o(n)$ and $q^2 = o(\ell n)$. Note that $q = o(n)$ is implied by $q^2 = o(\ell n)$ together with $\ell \le n$. ■

A.2. Lower Bound

We restate the theorem for the reader's convenience.

Theorem 10 (Lower bound) *Fix an arbitrary constant $\epsilon > 0$, not depending on n. If q, ℓ scale with n such that $1 \le q < q + \ell \le n$ and $q^2 \ge \ell n^{1+\epsilon}$ then there is no degree-$o(\log n/\log\log n)^2$ polynomial achieving weak separation between $\mathbb{P} = \mathsf{MC}(n, q)$ and $\mathbb{Q} = \mathsf{MC}(n, q+\ell)$.*

A.2.1. PROOF OVERVIEW

We first perform a standard manipulation, showing that it suffices to bound the quantity $\mathsf{Adv}_{\le D}$.

Lemma 22 *Let $\mathbb{P} = \mathbb{P}_n$ and $\mathbb{Q} = \mathbb{Q}_n$ be distributions on $\mathbb{R}^N$ for some $N = N_n$. For some $D = D_n$, let $\mathbb{R}[Y]_{\le D}$ denote the set of polynomials $\mathbb{R}^N \to \mathbb{R}$ of degree (at most) D. If*

$$\mathsf{Adv}_{\le D}(\mathbb{P}, \mathbb{Q}) := \sup_{f \in \mathbb{R}[Y]_{\le D}} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} = 1 + o(1),$$

then no degree-D polynomial $f : \mathbb{R}^N \to \mathbb{R}$ weakly separates $\mathbb{P}$ and $\mathbb{Q}$. Similarly, if $\mathsf{Adv}_{\le D}(\mathbb{P}, \mathbb{Q}) = O(1)$ then no degree-D polynomial strongly separates $\mathbb{P}$ and $\mathbb{Q}$.

It is always the case that $\mathsf{Adv}_{\le D} \ge 1$, by taking $f = 1$.

Proof Assume for the sake of contradiction that some degree-D polynomial $g : \mathbb{R}^N \to \mathbb{R}$ weakly separates $\mathbb{P}$ and $\mathbb{Q}$. By shifting and scaling, we can assume without loss of generality that $\mathbb{E}_{\mathbb{Q}}[g] = 0$ and $\mathbb{E}_{\mathbb{P}}[g] = 1$. For sufficiently large n, weak separation guarantees $\mathrm{Var}_{\mathbb{Q}}[g] = \mathbb{E}_{\mathbb{Q}}[g^2] \le C$ for some constant $C > 0$. Define $f = g + C$ and compute

$$\frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} = \frac{1+C}{\sqrt{\mathbb{E}_{\mathbb{Q}}[g^2] + C^2}} \ge \frac{1+C}{\sqrt{C + C^2}} = \sqrt{\frac{1+C}{C}},$$

which is a constant strictly greater than 1, contradicting $\mathsf{Adv}_{\le D} = 1 + o(1)$. The proof for strong separation is similar, now with $C = o(1)$. ■

A key ingredient in the proof will be an upper bound on $\mathsf{Adv}_{\le D}$ in the following generic setting (of which our problem is a special case). Suppose $\mathbb{Q}$ takes the form $Y = X \vee Z$ where $X, Z \in \{\pm 1\}^N$ with "noise" Z i.i.d. Rademacher and "signal" X having an arbitrary distribution (independent from Z), and $\vee$ denotes entrywise maximum. (In our case $N = \binom{n}{2}$ and X is the ± 1-valued indicator for clique edges.)

Proposition 23 *Suppose $\mathbb{Q}$ takes the form $Y = X \vee Z$ as described above and $\mathbb{P}$ is any distribution on $\{\pm 1\}^N$. For $\alpha, \beta \subseteq [N]$, define*

$$c_\alpha = \mathop{\mathbb{E}}_{Y \sim \mathbb{P}}[Y^\alpha] := \mathop{\mathbb{E}}_{Y \sim \mathbb{P}} \prod_{i \in \alpha} Y_i$$

and

$$M_{\beta\alpha} = \Pr_X(\alpha \setminus X = \beta).$$

Here and throughout, we abuse notation and use X to refer to the set $\{i \in [N] : X_i = 1\}$.

Suppose $M_{\alpha\alpha} > 0$ for all $|\alpha| \le D$. Then

$$\mathsf{Adv}^2_{\le D} \le \sum_{\alpha \subseteq [N],\, |\alpha| \le D} w_\alpha^2 \tag{4}$$

where w_α is defined recursively by

$$w_\alpha = \frac{1}{M_{\alpha\alpha}} \left(c_\alpha - \sum_{\beta \subsetneq \alpha} w_\beta M_{\beta\alpha} \right).$$

No explicit base case is needed for the recursion above, but one can think of $w_\emptyset = 1$ as the base case.

We pause to give some remarks on the origin of the above formula. The proof (given in Section A.2.4) follows a strategy based on Schramm and Wein (2022): apply Jensen's inequality to X (but not Z) and then the result can be explicitly calculated by solving an upper-triangular linear system. The original work Schramm and Wein (2022) gave a similar formula in the setting of *estimation*, and more recently Rush et al. (2022) was first to demonstrate that related techniques can also be used for testing between two "planted" distributions (which is also the setting of the current work). In contrast, previous low-degree lower bounds for testing problems had always required the "null" distribution $\mathbb{Q}$ to have independent coordinates; see the remark below for comparison.

Remark 24 *We note that Proposition 23 generalizes a well known formula for low-degree testing between "signal" and "pure noise." Specifically, consider the case where $X = -\mathbb{1}$ so that $\mathbb{Q}$ is i.i.d. Rademacher, and $\mathbb{P}$ is any distribution on $\{\pm 1\}^N$. In this case $M_{\beta\alpha} = \mathbb{1}_{\beta=\alpha}$ and so Proposition 23 reduces to the bound*

$$\mathsf{Adv}^2_{\le D} \le \sum_{|\alpha| \le D} \left(\mathop{\mathbb{E}}_{Y \sim \mathbb{P}}[Y^\alpha] \right)^2, \quad Y^\alpha := \prod_{i \in \alpha} Y_i,$$

which is standard (and in fact holds with equality); see Section 2.3 of Hopkins (2018).

Returning to the proof, a more convenient parametrization for w_α will be $\hat{w}_\alpha = M_{\alpha\alpha} w_\alpha$. In this case, since $M_{\emptyset\alpha} = \mathbb{E}_{\mathbb{Q}}[Y^\alpha]$, the recurrence can be written as

$$\hat{w}_\emptyset = 1,$$

$$\hat{w}_\alpha \;=\; c_\alpha - \sum_{\beta \subsetneq \alpha} \hat{w}_\beta \frac{M_{\beta\alpha}}{M_{\beta\beta}} \;=\; \mathop{\mathbb{E}}_{\mathbb{P}}[Y^\alpha] - \mathop{\mathbb{E}}_{\mathbb{Q}}[Y^\alpha] - \sum_{\emptyset \subsetneq \beta \subsetneq \alpha} \hat{w}_\beta \frac{M_{\beta\alpha}}{M_{\beta\beta}} \qquad \text{for } |\alpha| \geq 1. \tag{5}$$

The ratio of M's can be thought of as a conditional probability:

$$R_{\beta\alpha} \coloneqq \frac{M_{\beta\alpha}}{M_{\beta\beta}} = \frac{\Pr_X(\alpha \setminus X = \beta)}{\Pr_X(\beta \cap X = \emptyset)} = \Pr_X(\alpha \setminus X = \beta \mid \beta \cap X = \emptyset). \tag{6}$$

From this point onward, we specialize to our testing problem of interest: $\mathbb{P} = \mathsf{MC}(n, q)$ versus $\mathbb{Q} = \mathsf{MC}(n, q + \ell)$. As discussed above, our goal is to show $\mathsf{Adv}_{\leq D} = 1 + o(1)$ by bounding the formula in (4). The "1" comes from the $\alpha = \emptyset$ term, and we need to show that the rest of the sum is $o(1)$.

The following property of $\hat{w}$ will be key to the analysis; it is used crucially in the proof of Lemma 28. Note that we can think of α as a subset of edges of the complete graph on n vertices, and in this sense we can talk about α being connected or having connected components.

Lemma 25 *If α has connected components $\alpha_1, \ldots, \alpha_t$ then $\hat{w}_\alpha = \prod_{i=1}^t \hat{w}_{\alpha_i}$.*

Proof It suffices to prove the claim in the case where α is comprised of two non-empty disjoint edge sets α_1, α_2 with no vertices in common (i.e., each α_i is a union of connected components). Once we establish $\hat{w}_\alpha = \hat{w}_{\alpha_1}\hat{w}_{\alpha_2}$ in this case, the general statement follows by induction.

Note that due to independence across connected components, $c_\alpha = c_{\alpha_1}c_{\alpha_2}$. Any $\beta \subseteq \alpha$ can be uniquely decomposed as $\beta = \beta_1 \cup \beta_2$ with $\beta_1 \subseteq \alpha_1$ and $\beta_2 \subseteq \alpha_2$. Again by independence, $R_{\beta\alpha} = R_{\beta_1\alpha_1}R_{\beta_2\alpha_2}$. We will also need the fact $R_{\alpha\alpha} = 1$. We proceed by induction on $|\alpha|$. If either α_1 or α_2 is empty, the result follows immediately because $\hat{w}_\emptyset = 1$. Otherwise, assume by induction that $\hat{w}_\beta = \hat{w}_{\beta_1}\hat{w}_{\beta_2}$ for any $\beta \subsetneq \alpha$. We have

$$\begin{aligned}
\hat{w}_\alpha &= c_\alpha - \sum_{\beta \subsetneq \alpha} \hat{w}_\beta R_{\beta\alpha} \\
&= c_{\alpha_1}c_{\alpha_2} - \sum_{\substack{\beta_1 \subsetneq \alpha_1 \\ \beta_2 \subsetneq \alpha_2}} \hat{w}_{\beta_1}\hat{w}_{\beta_2}R_{\beta_1\alpha_1}R_{\beta_2\alpha_2} - \sum_{\substack{\beta_1 \subsetneq \alpha_1 \\ (\beta_2 = \alpha_2)}} \hat{w}_{\beta_1}\hat{w}_{\alpha_2}R_{\beta_1\alpha_1}R_{\alpha_2\alpha_2} - \sum_{\substack{\beta_2 \subsetneq \alpha_2 \\ (\beta_1 = \alpha_1)}} \hat{w}_{\alpha_1}\hat{w}_{\beta_2}R_{\alpha_1\alpha_1}R_{\beta_2\alpha_2} \\
&= c_{\alpha_1}c_{\alpha_2} - \left(\sum_{\beta_1 \subsetneq \alpha_1} \hat{w}_{\beta_1}R_{\beta_1\alpha_1}\right)\left(\sum_{\beta_2 \subsetneq \alpha_2} \hat{w}_{\beta_2}R_{\beta_2\alpha_2}\right) - \hat{w}_{\alpha_2}\sum_{\beta_1 \subsetneq \alpha_1} \hat{w}_{\beta_1}R_{\beta_1\alpha_1} - \hat{w}_{\alpha_1}\sum_{\beta_2 \subsetneq \alpha_2} \hat{w}_{\beta_2}R_{\beta_2\alpha_2}.
\end{aligned}$$

Using the recurrence (5), this becomes

$$\hat{w}_\alpha = c_{\alpha_1}c_{\alpha_2} - (c_{\alpha_1} - \hat{w}_{\alpha_1})(c_{\alpha_2} - \hat{w}_{\alpha_2}) - \hat{w}_{\alpha_2}(c_{\alpha_1} - \hat{w}_{\alpha_1}) - \hat{w}_{\alpha_1}(c_{\alpha_2} - \hat{w}_{\alpha_2}),$$

which simplifies to $\hat{w}_{\alpha_1}\hat{w}_{\alpha_2}$ as desired. ■

A.2.2. BOUNDING $\hat{w}_\alpha$

In the remainder of the proof we need to bound the values w_α and plug this into (4). Recall that when α is thought of as a graph, $|\alpha|$ is the number of edges. We also define $V(\alpha)$ to be the set of vertices of α, i.e., the vertices $i \in [n]$ incident to at least one edge of α.

Lemma 26 *For any α we have $M_{\alpha\alpha} \geq 1 - \frac{|\alpha|}{q+\ell}$.*

Proof Recall that $M_{\alpha\alpha}$ is the probability (under $\mathbb{Q}$) that α contains no clique edges. The probability that any specific edge is a clique edge is $1/(q+\ell)$, so the result follows by a union bound. ■

Lemma 27 *If $|\alpha| \geq 1$ and α is connected then*

$$0 \leq \mathop{\mathbb{E}}_{\mathbb{P}}[Y^\alpha] - \mathop{\mathbb{E}}_{\mathbb{Q}}[Y^\alpha] \leq \frac{\ell}{q^{|V(\alpha)|}}\left(|V(\alpha)| - 1\right).$$

Proof Since α is connected, $\mathbb{E}_{\mathbb{P}}[Y^\alpha]$ is the probability that all vertices of α are assigned the same label in $[q]$ (and similarly for $\mathbb{E}_{\mathbb{Q}}[Y^\alpha]$), i.e.,

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbb{P}}[Y^\alpha] - \mathop{\mathbb{E}}_{\mathbb{Q}}[Y^\alpha] &= \left(\frac{1}{q}\right)^{|V(\alpha)|-1} - \left(\frac{1}{q+\ell}\right)^{|V(\alpha)|-1} \\
&= \left(\frac{1}{q}\right)^{|V(\alpha)|-1}\left[1 - \left(\frac{q}{q+\ell}\right)^{|V(\alpha)|-1}\right] \\
&= \left(\frac{1}{q}\right)^{|V(\alpha)|-1}\left[1 - \left(1 - \frac{\ell}{q+\ell}\right)^{|V(\alpha)|-1}\right] \\
&\leq \left(\frac{1}{q}\right)^{|V(\alpha)|-1}\left[1 - \left(1 - \frac{\ell}{q+\ell}\left(|V(\alpha)| - 1\right)\right)\right] \\
&= \left(\frac{1}{q}\right)^{|V(\alpha)|-1}\frac{\ell}{q+\ell}\left(|V(\alpha)| - 1\right) \\
&\leq \frac{\ell}{q^{|V(\alpha)|}}\left(|V(\alpha)| - 1\right).
\end{aligned}$$

■

Lemma 28 *If $|\alpha| \geq 1$ then*

$$|\hat{w}_\alpha| \leq \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\alpha)|}(|\alpha| + 1)^{|\alpha|}.$$

Proof Proceed by induction on $|\alpha|$. First consider the case where α is not connected. Write α as the union of two non-empty disjoint edge sets α_1, α_2 with no vertices in common. By Lemma 25 and the induction hypothesis,

$$\begin{aligned}
|\hat{w}_\alpha| = |\hat{w}_{\alpha_1}| \cdot |\hat{w}_{\alpha_2}| &\leq \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\alpha_1)|}(|\alpha_1| + 1)^{|\alpha_1|} \cdot \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\alpha_2)|}(|\alpha_2| + 1)^{|\alpha_2|} \\
&\leq \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\alpha_1)|+|V(\alpha_2)|}(|\alpha_1| + |\alpha_2| + 1)^{|\alpha_1|+|\alpha_2|} \\
&= \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\alpha)|}(|\alpha| + 1)^{|\alpha|}
\end{aligned}$$

as desired.

Now consider the case where α is connected. Using (5) and Lemma 27,

$$|\hat{w}_\alpha| \le \left|\mathop{\mathbb{E}}_{\mathbb{P}}[Y^\alpha] - \mathop{\mathbb{E}}_{\mathbb{Q}}[Y^\alpha]\right| + \sum_{\emptyset \subsetneq \beta \subsetneq \alpha} |\hat{w}_\beta| \cdot |R_{\beta\alpha}| \le \frac{\ell}{q^{|V(\alpha)|}} \left(|V(\alpha)| - 1\right) + \sum_{\emptyset \subsetneq \beta \subsetneq \alpha} |\hat{w}_\beta| \cdot |R_{\beta\alpha}|.$$

Using the definition (6) and the connectivity of α, we can deduce (for any $\emptyset \subsetneq \beta \subsetneq \alpha$)

$$0 \le R_{\beta\alpha} \le \left(\frac{1}{q+\ell}\right)^{|V(\alpha)|-|V(\beta)|} \le \left(\frac{1}{q}\right)^{|V(\alpha)|-|V(\beta)|},$$

because once we condition on the labels in $V(\beta)$, each vertex in $V(\alpha) \setminus V(\beta)$ has at most one possible label that would allow the event $\alpha \setminus X = \beta$ to occur. (More formally, any vertex $i \in V(\alpha) \setminus V(\beta)$ is connected to some vertex $j \in V(\beta)$ by a path using edges from $\alpha \setminus \beta$. Since every edge on this path must be a clique edge in order for $\alpha \setminus X = \beta$ to occur, i must have the same label as j.) Now using the above bounds and the induction hypothesis,

$$\begin{aligned}
|\hat{w}_\alpha| &\le \frac{\ell}{q^{|V(\alpha)|}} \left(|V(\alpha)| - 1\right) + \sum_{\emptyset \subsetneq \beta \subsetneq \alpha} \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\beta)|} (|\beta|+1)^{|\beta|} \cdot \left(\frac{1}{q}\right)^{|V(\alpha)|-|V(\beta)|} \\
&\le \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\alpha)|} \left[|V(\alpha)| - 1 + \sum_{\emptyset \subsetneq \beta \subsetneq \alpha} (|\beta|+1)^{|\beta|}\right] \qquad \text{since } |V(\alpha)| \ge 2 \text{ and } |V(\beta)| \le |V(\alpha)| \\
&= \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\alpha)|} \left[|V(\alpha)| - 1 + \sum_{m=1}^{|\alpha|-1} \binom{|\alpha|}{m} (m+1)^m\right] \\
&\le \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\alpha)|} \left[|V(\alpha)| - 1 + \sum_{m=1}^{|\alpha|-1} \binom{|\alpha|}{m} |\alpha|^m\right] \\
&= \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\alpha)|} \left[|V(\alpha)| - 1 + (|\alpha|+1)^{|\alpha|} - 1 - |\alpha|^{|\alpha|}\right] \qquad \text{by the Binomial theorem} \\
&\le \left(\frac{\sqrt{\ell}}{q}\right)^{|V(\alpha)|} (|\alpha|+1)^{|\alpha|},
\end{aligned}$$

where the last step used $|V(\alpha)| \le 2|\alpha| \le |\alpha|^{|\alpha|} + 1$. ■

A.2.3. PUTTING IT TOGETHER

The rest of the proof is similar to the low-degree analysis of planted clique; see Section 2.4 of Hopkins (2018). We now complete the proof of Theorem 10.

Proof For any $|\alpha| \le D$, we have from Lemma 26 that

$$M_{\alpha\alpha} \ge 1 - \frac{|\alpha|}{q+\ell} \ge 1 - \frac{D}{q} = 1 - o(1),$$

due to our assumptions on q and D. Applying Proposition 23,

$$\mathsf{Adv}^2_{\le D} \le \sum_{|\alpha|\le D} w_\alpha^2 = 1 + \sum_{1\le|\alpha|\le D} \left(\frac{\hat{w}_\alpha}{M_{\alpha\alpha}}\right)^2 \le 1 + (1+o(1)) \sum_{1\le|\alpha|\le D} \hat{w}_\alpha^2.$$

Since our goal (by Lemma 22) is to show $\mathsf{Adv}_{\le D} = 1 + o(1)$, it remains to show

$$\sum_{1\le|\alpha|\le D} \hat{w}_\alpha^2 = o(1).$$

This follows from Proposition 30 below, using the bound on $|\hat{w}_\alpha|$ from Lemma 28 together with the assumption $q^2 \ge \ell n^{1+\epsilon}$. ■

Lemma 29 *For integers $t \ge 2$ and $D \ge 1$, the number of graphs $\alpha \subseteq \binom{n}{2}$ such that $|\alpha| \le D$ and $|V(\alpha)| = t$, is at most $n^t \min\{2^{t^2}, t^{2D}\}$.*

Proof The number of ways to choose t vertices is $\binom{n}{t} \le n^t$. Once the vertices are chosen, we can upper-bound the total number of graphs with $\le D$ edges in two different ways: $2^{\binom{t}{2}} \le 2^{t^2}$ or $\left(\binom{t}{2}+1\right)^D \le (t^2)^D$. ■

Proposition 30 *Suppose there exist fixed constants $\delta > 0$ and $C > 0$ such that for $\alpha \subseteq \binom{n}{2}$ with $1 \le |\alpha| \le D$, we have a quantity ϕ_α bounded by $|\phi_\alpha| \le n^{-\frac{1}{2}(1+\delta)\cdot|V(\alpha)|}(|\alpha|+1)^{C\cdot|\alpha|}$. If $D = D_n$ satisfies $D = o(\log n/\log\log n)^2$ then*

$$\sum_{1\le|\alpha|\le D} \phi_\alpha^2 = o(1)$$

as $n \to \infty$.

Proof

Using Lemma 29 and the fact $|\alpha| \le \binom{|V(\alpha)|}{2} \le |V(\alpha)|^2$,

$$\sum_{1\le|\alpha|\le D} \phi_\alpha^2 \le \sum_{2\le t\le\sqrt{D}} n^t 2^{t^2} \cdot n^{-(1+\delta)t}(t^2+1)^{2Ct^2} + \sum_{\sqrt{D}\le t\le 2D} n^t t^{2D} \cdot n^{-(1+\delta)t}(D+1)^{2CD}.$$

Consider the first sum on the right-hand side above. The initial term $t = 2$ is $O(n^2 \cdot n^{-2(1+\delta)}) = o(1)$, and the ratio between terms $t+1$ and t is

$$n^{-\delta} \cdot 2^{2t+1} \cdot ((t+1)^2+1)^{2C(2t+1)} \left(\frac{(t+1)^2+1}{t^2+1}\right)^{2Ct^2} \le t^{O(t)} n^{-\delta} \le \sqrt{D}^{O(\sqrt{D})} n^{-\delta} \le \frac{1}{2}$$

for sufficiently large n, using the assumption $D = o\left(\frac{\log n}{\log\log n}\right)^2$. Now consider the second sum. The initial term $t = \left\lceil\sqrt{D}\right\rceil$ is at most

$$n^{-\delta\sqrt{D}}(\sqrt{D}+1)^{2D}(D+1)^{2CD} \le n^{-\delta\sqrt{D}}(D+1)^{2(C+1)D} = o(1),$$

and the ratio between terms $t+1$ and t is

$$n^{-\delta} \cdot \left(\frac{t+1}{t}\right)^{2D} \le n^{-\delta}\left(1+\frac{1}{\sqrt{D}}\right)^{2D} \le n^{-\delta} \cdot e^{O(\sqrt{D})} \le \frac{1}{2}$$

for sufficiently large n. ■

A.2.4. PROOF OF PROPOSITION 23

The proof is similar to the lower bound for planted clique in Schramm and Wein (2022), Section 3.5. We give the details here for convenience.

Any degree-D polynomial $f : \{\pm 1\}^N \to \mathbb{R}$ has a unique expansion $f(Y) = \sum_{\alpha \subseteq [N], |\alpha| \le D} \hat{f}_\alpha Y^\alpha$. Write

$$\mathop{\mathbb{E}}_{\mathbb{P}}[f(Y)] = \sum_{|\alpha| \le D} \hat{f}_\alpha \mathop{\mathbb{E}}_{\mathbb{P}}[Y^\alpha] = \langle c, \hat{f} \rangle$$

where, recall, the vector $c = (c_\alpha)$ is defined by

$$c_\alpha = \mathop{\mathbb{E}}_{\mathbb{P}}[Y^\alpha].$$

By Jensen's inequality,

$$\mathop{\mathbb{E}}_{\mathbb{Q}}[f(Y)^2] \ge \mathop{\mathbb{E}}_{Z}\left(\mathop{\mathbb{E}}_{X} f(X \vee Z)\right)^2 =: \mathop{\mathbb{E}}_{Z} g(Z)^2 = \|\hat{g}\|^2$$

where

$$\begin{aligned} g(Z) &= \mathop{\mathbb{E}}_{X} f(X \vee Z) \\ &= \sum_{|\alpha| \le D} \hat{f}_\alpha \mathop{\mathbb{E}}_{X} (X \vee Z)^\alpha \\ &= \sum_{|\alpha| \le D} \hat{f}_\alpha \sum_{0 \subseteq \beta \subseteq \alpha} Z^\beta \Pr_X\{\alpha \setminus X = \beta\} \\ &= \sum_{\beta} Z^\beta \sum_{\alpha \supseteq \beta} \hat{f}_\alpha \Pr_X\{\alpha \setminus X = \beta\}. \end{aligned}$$

In other words, $\hat{g} = M\hat{f}$ where, recall, the matrix $M = (M_{\beta\alpha})$ is defined by

$$M_{\beta\alpha} = \mathbb{1}_{\beta \subseteq \alpha} \Pr_X\{\alpha \setminus X = \beta\}.$$

Note that M is upper triangular and (by assumption) has positive entries on the diagonal, so M is invertible. We have now shown $\mathbb{E}_{\mathbb{Q}}[f]^2 \ge \|\hat{g}\|^2 = \|M\hat{f}\|^2$ and so

$$\mathsf{Adv}_{\le D} = \sup_{f \in \mathbb{R}[Y]_{\le D}} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \le \sup_{\hat{f}} \frac{\langle c, \hat{f} \rangle}{\|M\hat{f}\|} = \sup_{\hat{g}} \frac{c^\top M^{-1}\hat{g}}{\|\hat{g}\|},$$

which has optimizer $\hat{g} = (c^\top M^{-1})^\top$, yielding

$$\mathsf{Adv}_{\le D} \le \|c^\top M^{-1}\| =: \|w\|$$

where w is the solution to $w^\top M = c^\top$. Solving for w using the upper-triangular structure of M gives the recurrence

$$w_\alpha = \frac{1}{M_{\alpha\alpha}} \left(c_\alpha - \sum_{\beta \subsetneq \alpha} w_\beta M_{\beta\alpha} \right), \tag{7}$$

completing the proof.

Appendix B. Refuting Colorability

B.1. Upper Bound

We restate the theorem for the reader's convenience.

Theorem 16 (Upper bound) *Suppose $q \le b\sqrt{n}$ for a constant $b < 1/2$ (not depending on n). Then there exists a constant $C = C(b) > 0$ and a polynomial $f = f_n$ of degree at most $C \log n$ that strongly separates $G(n, 1/2)$ and $\mathcal{R}_q$.*

Proof Let A denote the $\{\pm 1\}$-valued adjacency matrix of the *complement* graph, with 0's on the diagonal; if the graph is q-colorable then A has value 1 within each color class. For an integer $m \ge 1$ to be chosen later, consider the polynomial $f(X) = (n/q - 1)^{-2m} \operatorname{Tr}(A^{2m})$, which has degree $2m$ in the input variables $X \in \{\pm 1\}^{\binom{n}{2}}$.

First we let $X \in \mathcal{R}_q$ and aim to show $f(X) \ge 1$. Let $S \subseteq [n]$ be the largest color class, so $|S| \ge n/q$. Let $\mathbb{1}_S \in \{0,1\}^n$ denote the indicator vector for S. Letting $\lambda_{\max} = \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ denote the eigenvalues of A,

$$\lambda_{\max} \ge \frac{\mathbb{1}_S^\top A \mathbb{1}_S}{\|\mathbb{1}_S\|^2} = \frac{|S|(|S|-1)}{|S|} = |S| - 1 \ge \frac{n}{q} - 1$$

and

$$\lambda_{\max}^{2m} \le \sum_{i=1}^n \lambda_i^{2m} = \operatorname{Tr}(A^{2m}).$$

Combining these yields $\operatorname{Tr}(A^{2m}) \ge (n/q - 1)^{2m}$ and so $f(X) \ge 1$.

It remains to show $\mathbb{E}[f^2] = o(1)$ when $X \sim G(n, 1/2)$. Let Y be an $n \times n$ symmetric matrix where $\{Y_{ij} : i \le j\}$ are i.i.d. $\mathcal{N}(0,1)$. By direct expansion and comparison of Rademacher moments to Gaussian ones, $\mathbb{E}[\operatorname{Tr}(A^{2m})^2] \le \mathbb{E}[\operatorname{Tr}(Y^{2m})^2]$. Using $\|Y\|$ to denote the spectral norm of Y, the bound in Lemma 2.2 of Bandeira and van Handel (2016) gives

$$\mathbb{E}[\operatorname{Tr}(Y^{2m})^2] \le \mathbb{E}[n^2 \|Y\|^{4m}] \le n^2 (2\sqrt{n} + 2\sqrt{4m})^{4m}.$$

Putting it together,

$$\mathbb{E}[f^2] \le \left(\frac{n}{q} - 1\right)^{-4m} n^2 (2\sqrt{n} + 2\sqrt{4m})^{4m} = n^2 \left(\frac{2q(\sqrt{n} + \sqrt{4m})}{n - q} \right)^{4m},$$

which is $o(1)$ under the conditions of the theorem. ∎

B.2. Lower Bound

We restate the theorem for the reader's convenience.

Theorem 17 (Lower bound) *If $q \geq n^{2/3+\epsilon}$ for a constant $\epsilon > 0$, then no degree-$o(\log n/\log\log n)^2$ polynomial weakly separates $G(n,1/2)$ and $\mathcal{R}_q$.*

In light of Proposition 14, our goal is to show $\mathsf{Adv}_{\leq D}(\mathbb{P},\mathbb{Q}) = 1 + o(1)$ where $\mathbb{Q} = G(n,1/2)$ (and $Y \sim \mathbb{Q}$ is encoded by an element of $\{\pm 1\}^{\binom{n}{2}}$) and $\mathbb{P}$ is the planted distribution defined in Definition 18. Our starting point is the well-known formula from Remark 24:

$$\mathsf{Adv}^2_{\leq D} = \sum_{|\alpha| \leq D} \left(\mathop{\mathbb{E}}_{Y\sim\mathbb{P}} [Y^\alpha] \right)^2,$$

where $\alpha \subseteq \binom{n}{2}$. We identify α with the graph whose edge set is α, and write $V(\alpha) \subseteq [n]$ for the vertex set, i.e., the vertices incident to at least one edge in α. Our first step is to bound the coefficients $\lambda_\alpha := \mathbb{E}_{Y\sim\mathbb{P}}[Y^\alpha]$.

B.2.1. BOUNDING THE COEFFICIENTS

Lemma 31 (Bounding λ_α) *For any graph $\alpha \subseteq \binom{n}{2}$ we have*

$$|\lambda_\alpha| := \left| \mathop{\mathbb{E}}_{Y\sim\mathbb{P}} [Y^\alpha] \right| \leq O(q^{-3/4})^{|V(\alpha)|}$$

where $O(\cdot)$ hides an absolute constant factor.

Proof If $\alpha = \cup_i \alpha_i$ is the decomposition of α into connected components, we have $\lambda_\alpha = \prod_i \lambda_{\alpha_i}$ due to independence across components. It therefore suffices to prove the result in the case where α is connected.

Let $c : V(\alpha) \to [q] \times [q]$ denote the latent assignment of labels (a,b) to vertices from the definition of $\mathbb{P}$ (Definition 18). We have

$$\lambda_\alpha = \mathop{\mathbb{E}}_{c} \mathop{\mathbb{E}}_{Y\sim\mathbb{P}|c} [Y^\alpha] = \sum_c \Pr[c] \cdot \mathbb{E}[Y^\alpha | c].$$

Note that $\mathbb{E}[Y^\alpha|c] = 0$ unless every edge in α is either an independent set edge or clique edge in c, and in this case,

$$\mathbb{E}[Y^\alpha|c] = (-1)^{\#\text{ ind-set edges}}.$$

As a result, one possible upper bound on $|\lambda_\alpha|$ is the probability over c that every edge in α is either an ind-set edge or clique edge. We can bound this probability as follows. Recall we are assuming α is connected, and explore the vertices of α according to a breadth-first search. The first vertex's label is unconstrained. Each edge that leads to a new vertex must be an ind-set edge or clique edge, giving at most $2q$ possibilities for the new vertex's label. Since there are q^2 possible labels in total, we conclude

$$|\lambda_\alpha| \leq \left(\frac{2q}{q^2}\right)^{|V(\alpha)|-1} = \left(\frac{2}{q}\right)^{|V(\alpha)|-1} \tag{8}$$

for any connected α.

The bound (8) implies the desired result $|\lambda_\alpha| \le O(q^{-3/4})^{|V(\alpha)|}$ provided $|V(\alpha)| \ge 4$, as in this case we have $|V(\alpha)| - 1 \ge |V(\alpha)| - \frac{1}{4}|V(\alpha)| = \frac{3}{4}|V(\alpha)|$. For $|V(\alpha)| \le 3$ we will manually verify the result by checking all the possible graphs:

- If α has no edges then $\lambda_\alpha = 1$.
- If α is a single edge, the cases to consider for c are $\{(a,b),(a,b)\}$, $\{(a,b),(a,b')\}$, and $\{(a,b),(a',b)\}$ (where $a \ne a'$, $b \ne b'$). This gives

$$\lambda_\alpha = -\frac{1}{q^2} - \frac{1}{q}\left(1 - \frac{1}{q}\right) + \frac{1}{q}\left(1 - \frac{1}{q}\right) = -q^{-2}.$$

- If α is a length-2 path then conditioned on any label for the middle vertex, the two edges are independent. Reusing the calculation for the single edge, we have $\lambda_\alpha = (-q^{-2})^2 = q^{-4}$.
- If α is a triangle, we first claim that the only labelings c that contribute to λ_α are those in which a label (a,b) is repeated. This follows from the symmetry between c and the reversed labeling $\overline{c}$ where each pair is reversed: $(a,b) \mapsto (b,a)$. If c has no repeated labels, c and $\overline{c}$ contribute the same term but with opposite signs, as every ind-set edge becomes a clique edge and vice versa. In light of this, the remaining cases to consider for c are $\{(a,b),(a,b),(a,b)\}$, $\{(a,b),(a,b),(a,b')\}$, and $\{(a,b),(a,b),(a',b)\}$. This gives

$$\lambda_\alpha = -\frac{1}{q^4} - 3\cdot\frac{1}{q^3}\left(1 - \frac{1}{q}\right) - 3\cdot\frac{1}{q^3}\left(1 - \frac{1}{q}\right) = O(q^{-3}).$$

We have now verified $|\lambda_\alpha| \le O(q^{-3/4})^{|V(\alpha)|}$ for every connected α. As discussed previously, this implies the result for all α. ■

B.2.2. PUTTING IT TOGETHER

We now combine the results from above in order to bound $\mathsf{Adv}_{\le D}$ and complete the proof of Theorem 17.

Proof Due to our assumption $q \ge n^{2/3+\epsilon}$, Lemma 31 gives

$$|\lambda_\alpha| \le O(n^{-\frac{3}{4}(\frac{2}{3}+\epsilon)})^{|V(\alpha)|} = O(n^{-\frac{1}{2}-\frac{3}{4}\epsilon})^{|V(\alpha)|} \le n^{-\frac{1}{2}(1+\epsilon)\cdot|V(\alpha)|}$$

for sufficiently large n. Using Proposition 30, we have for any $D = o(\log n/\log\log n)^2$,

$$\mathsf{Adv}_{\le D}^2 - 1 = \sum_{1\le|\alpha|\le D} \lambda_\alpha^2 = o(1).$$

As discussed at the beginning of Section B.2, this completes the proof. ■

Appendix C. Completeness of Quiet Planting

In this section, we give a simple argument showing that the absence of a computationally quiet planted distribution implies the existence of a low-degree refutation algorithm, in high generality. Our proof is elementary and only needs a simple application of von Neumann's min-max principle. We restate the theorem for the reader's convenience.

Theorem 20 *Fix sequences $N = N_n$, $D = D_n$, $\mathbb{Q} = \mathbb{Q}_n$ a distribution on $\mathbb{R}^N$, and $\mathcal{R} = \mathcal{R}_n \subseteq \mathbb{R}^N$. Assume that for each n, $\mathbb{Q}$ is supported on a finite set and $\mathcal{R}$ is a finite set (but the cardinality of these sets may depend on n). The following are equivalent:*

(1) No degree-D polynomial strongly separates $\mathbb{Q}$ and $\mathcal{R}$.

(2) For an infinite subsequence of n values, there exists a distribution $\mathbb{P} = \mathbb{P}_n$ supported on $\mathcal{R}$ such that $\mathsf{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) = O(1)$.

Similarly, the following are equivalent:

(1) No degree-D polynomial weakly separates $\mathbb{Q}$ and $\mathcal{R}$.

(2) For an infinite subsequence of n values, there exists a distribution $\mathbb{P} = \mathbb{P}_n$ supported on $\mathcal{R}$ such that $\mathsf{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) = 1 + o(1)$.

Proof Let $\mathcal{P}$ denote the space of probability distributions on $\mathcal{R}$. Let $\mathcal{F}$ denote the space of degree-D polynomials $f : \mathbb{R}^N \to \mathbb{R}$ such that $\mathbb{E}_{\mathbb{Q}}[f] = 0$ and $\mathbb{E}_{\mathbb{Q}}[f^2] \leq 1$. Consider

$$\mathrm{val}_n = \inf_{\mathbb{P} \in \mathcal{P}} \sup_{f \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[f]. \tag{9}$$

By von Neumann's min-max principle (see below for discussion of the technical conditions required), the supremum and infimum can be exchanged:

$$\mathrm{val}_n = \sup_{f \in \mathcal{F}} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[f] = \sup_{f \in \mathcal{F}} \inf_{X \in \mathcal{R}} f(X). \tag{10}$$

A degree-D polynomial strongly (respectively, weakly) separates $\mathbb{Q}$ and $\mathcal{R}$ if and only if the value of (10) is $\omega(1)$ (resp., $\Omega(1)$). The negation of this statement is that $\mathrm{val}_n = O(1)$ (resp., $o(1)$) for an infinite subsequence of n, which from (9) is equivalent to having $\mathbb{P}_n$ defined on an infinite subsequence such that $\sup_{f \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[f] = O(1)$ (resp., $o(1)$). Now the result follows due to the identity $(\sup_{f \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[f])^2 + 1 = \mathsf{Adv}^2_{\leq D}(\mathbb{P}, \mathbb{Q})$; see Lemma 34 below.

It remains to verify the technical conditions for the min-max principle. Formally we use the following variant, which is a special case of Sion's min-max theorem (Sion, 1958; Komiya, 1988).

Theorem 32 *Let $\mathcal{P}$ be a compact convex subset of a linear topological space and $\mathcal{F}$ a convex subset of a linear topological space. If $\phi(x, y)$ is a continuous real-valued function on $\mathcal{P} \times \mathcal{F}$ with $\phi(x, \cdot)$ concave for all $x \in \mathcal{P}$, and $\phi(\cdot, y)$ convex for all $y \in \mathcal{F}$, then $\min_{x \in \mathcal{P}} \sup_{y \in \mathcal{F}} \phi(x, y) = \sup_{y \in \mathcal{F}} \min_{x \in \mathcal{P}} \phi(x, y)$.*

In our setting, the linear topological spaces will simply be $\mathbb{R}^d$ for some d. Recall that our choice of $\mathcal{P}$ is the space of probability distributions on a finite set $\mathcal{R} = \{r_1, r_2, \ldots, r_{|\mathcal{R}|}\}$. We can identify $\mathcal{P}$ with a compact convex subset of $\mathbb{R}^{|\mathcal{R}|}$ by encoding a distribution $\mathbb{P}$ as the vector of probabilities $(\mathbb{P}(r_1), \ldots, \mathbb{P}(r_{|\mathcal{R}|}))$. Recall that our choice of $\mathcal{F}$ is the space of degree-D polynomials $f : \mathbb{R}^N \to \mathbb{R}$ such that $\mathbb{E}_{\mathbb{Q}}[f] = 0$ and $\mathbb{E}_{\mathbb{Q}}[f^2] \leq 1$. Letting $\mathcal{X} = \mathrm{supp}(\mathbb{Q}) \cup \mathcal{R} = \{x_1, \ldots, x_{|\mathcal{X}|}\}$, we can identify $\mathcal{F}$ with a convex subset of $\mathbb{R}^{|\mathcal{X}|}$ (note that $\mathcal{F}$ is not required to be compact) by encoding a function $f : \mathbb{R}^N \to \mathbb{R}$ as the vector $(f(x_1), \ldots, f(x_{|\mathcal{X}|}))$. Finally, note that $\phi(\mathbb{P}, f) \coloneqq \mathbb{E}_{\mathbb{P}}[f]$ is continuous, convex in $\mathbb{P}$, and concave in f; in fact, it is linear in both variables. This justifies our earlier exchange of inf and sup, completing the proof. ■

Remark 33 *Above we have assumed* $\mathrm{supp}(\mathbb{Q})$ *and* $\mathcal{R}$ *are finite to simplify the analytic conditions needed for the min-max principle, but these assumptions can be relaxed. For instance, one can alternatively assume that* $\mathbb{Q}_n$ *is any distribution on* $\mathbb{R}^N$ *with all moments finite and that* $\mathcal{R}_n \subseteq \mathbb{R}^N$ *is compact. Since* $\mathcal{R}$ *is compact, the space* $\mathcal{P}$ *of probability distributions on* $\mathcal{R}$ *is compact in the weak-* topology.*

Lemma 34 $\sup_{f \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[f]^2 + 1 = \mathsf{Adv}^2_{\leq D}(\mathbb{P}, \mathbb{Q})$.

Proof If the likelihood ratio $LR = d\mathbb{P}/d\mathbb{Q}$ exists, this fact follows from standard characterizations of these quantities as $L^2(\mathbb{Q})$-norms of projections of likelihoods (see Section 2.3 of Hopkins (2018)); namely, the left-hand side is $\|LR^{\leq D} - 1\|^2_{\mathbb{Q}} + 1$ and the right-hand side is $\|LR^{\leq D}\|^2_{\mathbb{Q}}$. We also give a self-contained proof below.

Recalling the definition of $\mathsf{Adv}_{\leq D}$, our goal is to show

$$\sup_{f \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[f]^2 + 1 = \sup_{g \in \mathbb{R}[Y]_{\leq D}} \frac{\mathbb{E}_{\mathbb{P}}[g]^2}{\mathbb{E}_{\mathbb{Q}}[g^2]}.$$

Note that the value 1 is achievable on both sides by taking $f = 0$ or $g = 1$. To show "$\leq$," suppose we have $f \in \mathcal{F}$ such that $\mathbb{E}_{\mathbb{P}}[f] = a > 0$, achieving value $a^2 + 1$ on the left-hand side. Then $g = f + 1/a$ achieves the same value $a^2 + 1$ on the right-hand side.

To show "$\geq$," suppose g achieves value $b^2 > 1$ on the right-hand side, and scale g so that $\mathbb{E}_{\mathbb{Q}}[g^2] = 1$ and $\mathbb{E}_{\mathbb{P}}[g] = b > 1$. Define $\Delta = \mathbb{E}_{\mathbb{Q}}[g]$ and note that $\mathbb{E}_{\mathbb{Q}}(g - \Delta)^2 = 1 - \Delta^2 \geq 0$ and $\mathbb{E}_{\mathbb{P}}(g - \Delta) = b - \Delta > 0$. If $\Delta = 1$ then the left-hand side is unbounded by taking f to be an arbitrary multiple of $g - \Delta$. Otherwise set $f = (g - \Delta)/\sqrt{1 - \Delta^2} \in \mathcal{F}$ and compute the left-hand side value

$$\mathbb{E}_{\mathbb{P}}[f]^2 + 1 = \frac{(b - \Delta)^2}{1 - \Delta^2} + 1 = b^2 + \frac{(b\Delta - 1)^2}{1 - \Delta^2} \geq b^2,$$

completing the proof. ■

Appendix D. Planted $(q+1)$-coloring is not q-colorable

Here we work with the complement graph and consider a partition into cliques rather than a coloring. Recall the multiple cliques model (Definition 3).

Proposition 35 *If $1 \le q \le \Omega(n/\log n)$ then with probability $1 - o(1)$, $\mathsf{MC}(n, q+1)$ does not admit a partition of the vertices into q cliques.*

Proof Fix an absolute constant $\epsilon > 0$, to be chosen later. Assume $q \le cn/\log n$ for a constant $c = c(\epsilon) > 0$ to be chosen later. The proof hinges on 3 basic facts, which hold w.h.p.:

(i) $G(n, 1/2)$ does not contain the complete bipartite graph $K_{m,m}$ as a subgraph, for $m \ge (2+\epsilon)\log_2 n$.

(ii) Letting $S_1, \dots, S_{q+1}$ denote the color classes of $\mathsf{MC}(n, q+1)$, we have $|S_i| \in (1 \pm \epsilon)\frac{n}{q+1}$ for all $i \in [q+1]$.

(iii) In $\mathsf{MC}(n, q+1)$, any vertex $v \in S_i$ has at most $(1/2 + 2\epsilon)\frac{n}{q+1}$ neighbors in S_j, for $i \ne j$.

Standard arguments show that (i)–(iii) hold with probability $1 - o(1)$, and we omit the details. The proof of (i) is a first moment calculation (compute the expected number of copies of $K_{m,m}$ and apply Markov's inequality), and the proof of (ii) and (iii) uses Bernstein's inequality along with a union bound.

Suppose $G \sim \mathsf{MC}(n, q+1)$. To complete the proof, it suffices to show that properties (i)–(iii) deterministically imply that G has no partition into q cliques (where property (i) applies to the underlying random graph $G' \sim G(n, 1/2)$ used to generate G, before the $q+1$ cliques were added). Assume (i)–(iii) hold, and suppose for contradiction that G admits a partition $V(G) = T_1 \sqcup T_2 \sqcup \cdots \sqcup T_q$ into cliques.

We first claim that for every $i \in [q]$, we either have (Case I) $|T_i| \le \frac{3}{4} \cdot \frac{n}{q+1}$ or (Case II) for some $j \in [q+1]$, $T_i \subseteq S_j$ and $|T_i| > \frac{1}{2}|S_j|$. To see this, note that if Case I fails then $|T_i| > \frac{3}{4} \cdot \frac{n}{q+1}$, and so to avoid violating property (i) there must exist j such that $|T_i \cap S_j| \ge \frac{3}{4} \cdot \frac{n}{q+1} - (2+\epsilon)\log_2 n > (1/2 + 2\epsilon)\frac{n}{q+1}$. Now property (iii) implies $T_i \subseteq S_j$. Also, property (ii) implies $|T_i| > \frac{1}{2}|S_j|$. This proves the claim.

Using the above claim, we can now construct an injective map $\phi : [q] \to [q+1]$ such that $|T_i| \le |S_{\phi(i)}|$. First, for i in Case II, set $\phi(i)$ to be the corresponding j; then for i in Case I, set $\phi(i)$ to be any unused j value. Since some $j \in [q+1]$ is not in the image of ϕ, we have $\sum_{i \in [q]} |T_i| < \sum_{j \in [q+1]} |S_j| = n$, a contradiction. ■