Proceedings of Machine Learning Research vol 247:1–25, 2024 37th Annual Conference on Learning Theory

Low-degree phase transitions for detecting a planted clique in sublinear time

Jay Mardia JMARDIA@STANFORD.EDU
Stanford University

Kabir Aladin Verchand KAV29@CAM.AC.UK
University of Cambridge and Georgia Institute of Technology

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

Editors: Shipra Agrawal and Aaron Roth

Abstract

We consider the problem of detecting a planted clique of size k in a random graph on n vertices. When the size of the clique exceeds $\Theta(\sqrt{n})$, polynomial-time algorithms for detection proliferate. We study faster—namely, sublinear time—algorithms in the high-signal regime when $k = \Theta(n^{1/2+\delta})$, for some $\delta > 0$. To this end, we consider algorithms that non-adaptively query a subset M of entries of the adjacency matrix and then compute a low-degree polynomial function of the revealed entries. We prove a computational phase transition for this class of *non-adaptive low-degree algorithms*: under the scaling $|M| = \Theta(n^\gamma)$, the clique can be detected when $\gamma > 3(1/2-\delta)$ but not when $\gamma < 3(1/2-\delta)$. As a result, the best known runtime for detecting a planted clique, $\widetilde{O}(n^{3(1/2-\delta)})$, cannot be improved without looking beyond the non-adaptive low-degree class.

Our proof of the lower bound—based on bounding the conditional low-degree likelihood ratio—reveals further structure in non-adaptive detection of a planted clique. Using (a bound on) the conditional low-degree likelihood ratio as a potential function, we show that for *every* non-adaptive query pattern, there is a highly structured query pattern of the same size that is at least as effective.

Keywords: Sublinear time algorithms, statistical-computational gaps, low-degree polynomials

1. Introduction

Many high-dimensional statistical inference problems (e.g., community detection (Decelle et al., 2011), planted clique (Jerrum, 1992), and tensor PCA (Richard and Montanari, 2014), to name a few) appear to exhibit statistical-computational gaps wherein the amount (or quality) of data required for all known polynomial-time algorithms may be significantly larger than the amount of data required information-theoretically.

Central among these is the planted clique problem which we consider here. In more detail, the planted clique problem consists of observing a graph G on n vertices which may have arisen from one of two distributions: the null distribution in which $G \sim G(n, 1/2)$ (the Erdős–Rényi distribution) and a planted distribution in which k of the n vertices form a clique and the remaining edges in the graph appear with probability $1/2$, independently. The goal of the detection task is to distinguish these two cases. Information-theoretically, it is possible to detect the presence of a planted clique of size $k \geq (2+\epsilon)\log_2(n)$ for any $\epsilon > 0$ (Bollobas and Erdős, 1976), whereas the best known polynomial-time algorithms require $k = \Omega(\sqrt{n})$ (see, e.g., Kučera, 1995; Alon et al., 1998; Deshpande and Montanari, 2015; Barak et al., 2019, and the references therein). Each of

the aforementioned algorithms requires the full observation of the random graph G, which has size $\Theta(n^2)$, and as a result they require runtime at least $\Omega(n^2)$.

When n gets large, even this polynomial running time may prove prohibitively expensive, and it becomes of interest to apply algorithms which run in time sublinear in the input size (see, e.g., the review Rubinfeld and Shapira, 2011, and references therein). In this work we aim to investigate precisely what runtime is required to detect a clique in the "easy" regime $k = \Theta(n^{1/2+\delta})$ for a constant $\delta \in (0, 1/2)$. In this regime, the clique vertices can be identified simply based on their degree in the graph (Kučera, 1995), and therefore the maximum degree suffices as a statistic for distinguishing the null and planted distributions. While a naive computation of the maximum degree requires time $\Omega(n^2)$, a faster detection algorithm of runtime $\widetilde{O}(n^{3(1/2-\delta)})$ was given by Mardia et al. (2020): the idea is to approximately estimate the degrees of some subset of the vertices while only examining $\widetilde{O}(n^{3(1/2-\delta)})$ entries of the adjacency matrix. Is this optimal, or might it be possible to reduce the runtime even further?

To explore the fundamental limits of sublinear-time computation, we will consider the *query complexity* of algorithms, that is, the number of entries of the adjacency matrix that need to be read. This, after all, is the bottleneck in the runtime of Mardia et al. (2020). Certainly any algorithm of runtime $O(t)$, for some $t = t(n)$, must make at most $O(t)$ queries. These queries can potentially be chosen adaptively, based on the results of previous queries. On the other hand, the algorithm of Mardia et al. (2020) is *non-adaptive*, meaning it specifies upfront a *mask* $M \subseteq \binom{[n]}{2}$, i.e., a subset of entries of the input to be observed (depending only on the problem size n).

This suggests a natural path forward: prove lower bounds on the query complexity, which in turn imply lower bounds on runtime. In fact this has been studied already: for $k = \Theta(n^{1/2+\delta})$ with $\delta \in (-1/2, 1/2)$, it is possible to detect a clique with $\widetilde{O}(n^{2(1/2-\delta)})$ non-adaptive queries, and up to log factors this number of queries is information-theoretically necessary (even if adaptivity is allowed) (Rácz and Schiffer, 2019). This improves the query complexity of Mardia et al. (2020), yet does not lead to a better runtime because a quasipolynomial-time exhaustive search is used to identify a large clique within the queried subgraph. The situation thus proves more subtle than it first appeared: query complexity is not the only bottleneck for runtime.

Our goal will be to show that the runtime of Mardia et al. (2020) is optimal, at least within some broad class of algorithms. In light of the above, we cannot merely study the (information-theoretic) query complexity, but will need to further "tie the hands" of the algorithm. First, for simplicity we will focus on algorithms that non-adaptively query the input. Second, we will ask that the results of the queries are processed via an efficient (say, polynomial-time) computation. Given the current state of average-case complexity theory, we cannot hope to prove negative results for arbitrary poly-time computation, so we follow a line of prior work (Hopkins and Steurer, 2017; Hopkins et al., 2017; Hopkins, 2018) and adopt a popular proxy for this: algorithms that can be represented as $O(\log n)$-degree polynomials. Thus we study the class of *non-adaptive low-degree algorithms*: such an algorithm consists of a sequence (indexed by the problem size n) of *masks* $M \subseteq \binom{[n]}{2}$ along with a sequence of multivariate polynomials $f : \{0,1\}^M \to \mathbb{R}$ of degree $O(\log n)$ whose input variables are the revealed entries of the adjacency matrix. An algorithm of this type is considered successful at detecting the planted clique if the output of f *separates* (in a sense made precise by Definition 5) the null and planted distributions. Logarithmic-degree polynomials are fairly expressive, allowing computation of edge counts, triangle counts, and other small subgraph counts, as well as approximate eigenvalue computations via power iteration (see e.g., Kunisky et al. (2022)). Our lower bound rules out polynomials of even larger degree, namely any $o(\log^2 n)$.

Our main result is to characterize the number of queries $|M| = \Theta(n^\gamma)$ required for a non-adaptive low-degree algorithm to detect a planted clique of size $k = \Theta(n^{1/2+\delta})$ for a constant $\delta \in (0, 1/2)$. In more detail, we show (see Theorem 6 to follow) that—by simulating the degree-counting algorithm of Mardia et al. (2020)—some non-adaptive low-degree algorithm succeeds when $\gamma > 3(1/2 - \delta)$, but conversely, no non-adaptive low-degree algorithm succeeds when $\gamma < 3(1/2-\delta)$. This lets us complete the phase diagram for planted clique detection in the non-adaptive query model, shown in Figure 1. As a result, the runtime $\widetilde{O}(n^{3(1/2-\delta)})$ of Mardia et al. (2020) for detecting a planted clique cannot be significantly improved without looking beyond the non-adaptive low-degree class.

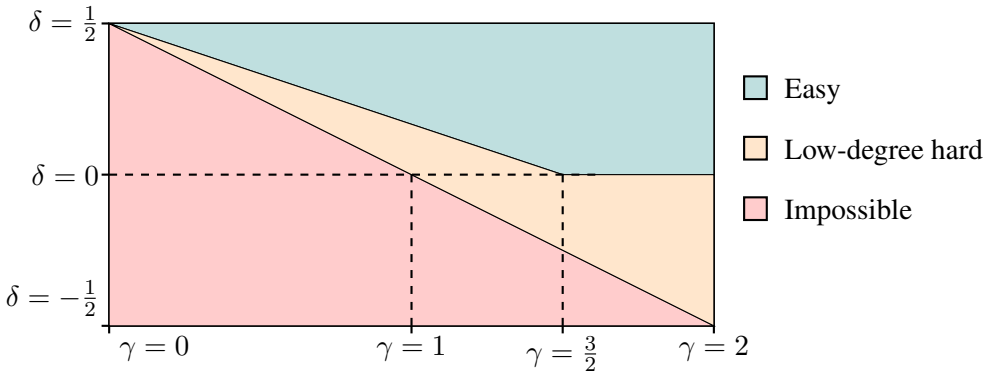

Figure 1: Phase diagram for detecting a clique of size $k = \Theta(n^{1/2+\delta})$ using $|M| = \Theta(n^\gamma)$ non-adaptive queries to the adjacency matrix. If arbitrary computation is allowed on the query results, detection is impossible in the red region and possible otherwise (Rácz and Schiffer, 2019). If a low-degree test must be applied to the query results, our upper bound Theorem 6(b) achieves detection in the green (easy) region; in this region, there is also an algorithm for clique detection whose runtime is dominated by the query complexity $\Theta(n^\gamma)$ (Mardia et al., 2020). Below the line $\delta = 0$, it is known that low-degree polynomials cannot detect the clique, even if the entire input is revealed (Barak et al., 2019; Hopkins, 2018). Our lower bound Theorem 6(a) fills in the rest of the hard (yellow) region.

1.1. Further related work

Computational complexity of statistics. Statistical-computational gaps are ubiquitous throughout high-dimensional testing and inference problems. These gaps call for a theory of computational lower bounds (hardness results), as otherwise we can never be sure whether the "possible but hard" regime fundamentally admits no efficient algorithm or whether there is a better algorithm waiting to be discovered. For *average-case* computational tasks—where the input is random—we unfortunately lack tools to prove complexity results conditional on standard assumptions such as $P \neq NP$. It is therefore common to resort to one of two tactics: (i) average-case reductions which establish hardness conditional on the hardness of some "standard" problem such as planted clique (e.g., Brennan et al. (2018)) or (ii) proving unconditional failure of particular families of algorithms. Within

the latter viewpoint, some popular classes of algorithms to rule out include statistical query (SQ) algorithms (e.g., Feldman et al. (2017)), the sum-of-squares (SoS) hierarchy (e.g., Barak et al. (2019)), and low-degree polynomials (the subject of this work).

We discuss briefly the prospect of applying some of the other frameworks mentioned above to our problem of interest—planted clique detection with non-adaptive queries. Average-case reductions (starting from the basic planted clique problem) have in fact already been applied to our setting, showing hardness in the same regime as us but only for certain highly structured masks (Mardia et al., 2020); addressing arbitrary masks appears to be beyond the reach of current techniques and we consider this an interesting question for future work. The SQ framework is not directly applicable to our setting because our input (a random graph) does not consist of i.i.d. samples; it may be possible to formulate a bipartite variant of our problem in the SQ model, similar to Feldman et al. (2017). SoS lower bounds tend to be rather unwieldy to prove, even for the basic planted clique problem (Barak et al., 2019), and they only show hardness of the *refutation* problem (which in general need not imply hardness of detection; see Banks et al. (2021); Bandeira et al. (2021)).

Low-degree polynomials as a model of computation. The idea to consider low-degree polynomials as a restricted class of statistical tests first arose from the sum-of-squares literature (Barak et al., 2019; Hopkins and Steurer, 2017; Hopkins et al., 2017; Hopkins, 2018) and has by now found success in a wide variety of settings (see Kunisky et al. (2022) for a survey), including extensions beyond hypothesis testing (Gamarnik et al., 2020; Schramm and Wein, 2022). For instance, in the planted clique problem, $O(\log n)$-degree polynomials can detect a clique of size $k \gtrsim \sqrt{n}$ but provably fail to detect a clique of size $k \ll \sqrt{n}$ (Barak et al., 2019; Hopkins, 2018), suggesting that this threshold is a fundamental barrier for efficient computation (or more conservatively, a barrier for certain known approaches). For planted clique and various other inference problems of this style, low-degree polynomials capture the best known poly-time algorithms and give a rigorous explanation for apparent computational barriers.

Our work is the first to employ low-degree polynomials to probe the precise limits of sublinear computation. Prior work has addressed coarser questions about runtime by taking polynomial degree as a proxy for runtime, e.g., with degree n^δ corresponding to time $\exp(n^{\delta \pm o(1)})$ (see e.g., Ding et al. (2023b)). In our regime of sublinear runtime, we cannot hope for a meaningful correspondence between polynomial degree and runtime, since even a degree-1 polynomial can already read the entire input. Instead, our approach relies on explicitly restricting the algorithm to a small fraction of the input variables.

On a technical level, we use the standard *low-degree likelihood ratio* (see Hopkins (2018)) as a tool for ruling out all low-degree polynomial tests. For testing between a specific pair of planted and null distributions, this often boils down to a relatively straightforward computation. However, since we allow an arbitrary choice of mask, we effectively need to prove *many* such hardness results all at once. To complicate things further, our setting requires a *conditional* variant of the low-degree likelihood ratio (Bandeira et al., 2022; Coja-Oghlan et al., 2022; Dhawan et al., 2023; Ding et al., 2023a). We give an overview of the proof in Section 2.1.

Average-case fine-grained hardness. In analogy with classical worst-case to average-case reductions (see e.g., Ajtai (1996); Regev (2010)), a recent line of literature establishes fine-grained notions of average-case complexity (Ball et al., 2017; Dalirrooyfard et al., 2020) via worst-case to average-case reductions for problems such as counting cliques in random hypergraphs (Goldreich and Rothblum, 2018; Boix-Adserà et al., 2021) and counting bicliques in random bipartite

graphs (Hirahara and Shimizu, 2021). Our work departs from these along two axes: First, our hardness results are unconditional but restricted to algorithms which can be expressed as low-degree polynomials. Second, as opposed to a counting problem, here we consider a testing problem which appears to exhibit a statistical-computational gap. To the best of our knowledge, our work (along with Mardia et al. (2020)) is the first to provide such evidence for fine-grained hardness of testing.

Restricting algorithms via query complexity. While restricting algorithms via query complexity forms a dominant theme in the study of sublinear-time algorithms and property testing (see, e.g., Rubinfeld and Shapira (2011) and Goldreich (2017) for comprehensive accounts), we note that the restriction to algorithms which make non-adaptive queries can be alternatively motivated in its own right (without reference to sublinear runtime). In particular, this models a scenario where the statistician must decide upfront which data to collect. This is relevant in (for instance) the *group testing* problem where the goal is to identify which individuals are afflicted by a disease based on "pooled" tests (see, e.g., Aldridge et al. (2019) for a survey). It is realistic to assume that the subset of individuals included in each test must be chosen non-adaptively (without knowledge of other test results) for purposes of practical implementation. Unlike our problem, there is no statistical-computational gap in group testing: a particular choice for the "design" (the choice of subsets to test, analogous to our "mask") succeeds using the information-theoretic minimum number of tests (Coja-Oghlan et al., 2020).

Turning to query complexity in problems on random graphs, Ferber et al. (2016); Conlon et al. (2020); Alweiss et al. (2021) consider the *subgraph query problem*, that is, the query complexity of finding a fixed subgraph in a sufficiently large Erdős–Rényi random graph. Following this line of work, Feige et al. (2020) study the query complexity of finding a large clique in an Erdős–Rényi graph. Importantly, Feige et al. (2020) allow for a limited amount of adaptivity in their queries, allowing a constant number of rounds in which the queries in each round may depend on the result of previous rounds. Tighter bounds on the query complexity upon restricting the number of rounds of adaptivity were later obtained by Feige and Ferster (2021); Csóka and Pongrácz (2023).

Information-theoretic (potentially adaptive) query complexity limits were established for the planted variant of the problem which we study here by Rácz and Schiffer (2019) (see also Rashtchian et al. (2021) for an alternate proof of the lower bound via communication complexity). The more general problem of random subgraph detection was later studied by Huleihel et al. (2021) in which an analogous information-theoretic threshold of $\widetilde{\Theta}(n^2/k^2)$ was established as well as a polynomial-time algorithm which requires $\widetilde{\Omega}(n^3/k^3)$ queries to succeed. In contrast with this last line of work, we consider the planted variant of the problem and provide *restricted* query complexity lower bounds over the family of algorithms which can be expressed as low-degree polynomial functions of the input. These lower bounds in turn match existing algorithmic upper bounds in the literature (Mardia et al., 2020).

1.2. Open problems

A number of interesting directions remain open and we detail a few here.

- *Reaching the computational threshold.* While our results indicate a smooth tradeoff between clique size and runtime above the computational threshold, it is less clear what happens near the computational threshold $k \approx \sqrt{n}$. When $k = \omega(\sqrt{n \log n})$, the algorithm of Mardia et al. (2020) runs in time $\widetilde{O}(n^{3/2})$. However, when the clique has size $k = o(\sqrt{n \log n})$ so that

degree counting no longer works, but is still above the computational threshold $k = \Omega(\sqrt{n})$, it is unknown whether detection is possible in *strongly sublinear* time, i.e., time $O(n^{2-\epsilon})$ for a constant $\epsilon > 0$.

- *Adaptivity.* Our results provide a lower bound on the capability of non-adaptive low-degree sublinear-time algorithms to detect a planted clique. It is natural to wonder whether adaptivity helps, or whether our lower bounds can be extended to adaptive algorithms. One fruitful direction may be to define a restricted family of adaptive algorithms—perhaps one where low-degree polynomials govern how queries are selected adaptively—and prove lower bounds against this family. We note that while adaptivity does not appear to help for detection, it does appear to help for the related problem of recovering the clique vertices. That is, the degree counting algorithms of Mardia et al. (2020) for detection are non-adaptive, but their natural counterparts for recovery require adaptive queries.

- *Beyond low-degree polynomials.* Our result provides evidence for a query-complexity based statistical-computational gap for the planted clique problem. It would be nice to gain more evidence for this by proving similar hardness results for algorithmic classes beyond low-degree polynomials. We believe the series of steps involved in our impossibility result should be useful even when considering other algorithmic classes. At the end of Section 2.1 we briefly mention which aspect of our approach is specific to low-degree polynomials. It would be very interesting to implement this step for other algorithmic classes. Our ideal goal would be to show that the algorithm of Mardia et al. (2020) is essentially optimal, conditional on the planted clique conjecture.

2. Problem setup and main results

As previously alluded to, we will consider *non-adaptive low-degree algorithms*. When the full input is a graph on n vertices, a non-adaptive algorithm must specify the subgraph to be queried before any observations are made. We formalize this through the notion of a mask, defined presently.

Definition 1 (Mask and mask degree) *A mask M over a ground set $[n]$ is a subset of $\binom{[n]}{2}$. We interpret this subset as specifying a graph on vertex set $V(M) := \{v \in [n] : \exists u \in [n] \text{ such that } (u, v) \in M\}$ with edge set M. Whenever the ground set is clear from context, we will not mention it and suppress it in our notation. For any vertex $v \in V(M)$ we refer to its mask degree as $\mathsf{deg}^M(v) := |\{u \in V(M) : (u, v) \in M\}|$.*

Equipped with this definition, we turn to our null distributions $G(n, 1/2)$ and $G(n, M)$, which denote the Erdős–Rényi distribution and its masked counterpart, respectively.

Definition 2 (Erdős–Rényi distribution $G(n, 1/2)$ and masked Erdős–Rényi distribution $G(n, M)$) *$G(n, 1/2)$ is the uniform distribution on $\{+1, -1\}^{\binom{n}{2}}$. We interpret a sample from this distribution as describing a graph G with vertex set $[n]$ in which the edge (i, j) is present if and only if the entry indexed by the unordered pair (i, j) is $+1$. Moreover, given a mask M, the masked Erdős–Rényi distribution $G(n, M)$ denotes the marginal distribution of $G(n, 1/2)$ restricted to the coordinates in M.*

Turning to our planted distributions, we first define the clique distribution and its conditional counterpart (which will be important as a proof device).

Definition 3 (Clique indicator distributions: $\mathsf{Clique}(n,k)$ and $\mathsf{Clique}(n,k,S)$) *$\mathsf{Clique}(n,k)$ denotes the uniform distribution over vectors in $\{0,1\}^{[n]}$ that have exactly k nonzero coordinates. We interpret a sample from this distribution as a choice of which vertices belong to the planted clique. For any subset $S \subseteq [n]$, $\mathsf{Clique}(n,k,S)$ denotes the distribution $\mathsf{Clique}(n,k)$ conditioned on all nonzero coordinates being inside S.*

Now we define our planted distributions $G(n,1/2,k)$ and $G(n,k,M)$.

Definition 4 (Planted Clique distributions: $G(n,1/2,k)$ and $\mathsf{G}(n,k,M)$) *Let K be sampled from $\mathsf{Clique}(n,k)$. Then $\mathsf{G}(n,1/2,k)$ is the following distribution on $\{+1,-1\}^{\binom{[n]}{2}}$.*

1. *Coordinates corresponding to unordered pairs (i,j) with $K_i = K_j = 1$ are $+1$.*
2. *Every other coordinate is independent and uniform on $\{+1,-1\}$.*

We again interpret a sample from this distribution as a graph on vertex set $[n]$, where $+1$ denotes the presence of an edge. When restricted to the coordinates in the mask M, the distribution is denoted by $G(n,k,M)$.

Given a graph G restricted to the coordinates in the mask M, our task, then, is to determine whether the observations originated from the null distribution $G(n,M)$ or the planted distribution $G(n,k,M)$.

Our main result provides a tight condition on the size of the mask M which controls whether or not *separation*—defined presently—between the null distribution $G(n,M)$ and the planted distribution $G(n,k,M)$ is possible using a low-degree test. Our results are asymptotic in nature, and thus we will often refer to a *sequence* of problems (or algorithms, etc.), which are assumed to be indexed by the problem size n.

Definition 5 (Strong/weak separation) *Consider a sequence $N = N_n$ and two (sequences of) distributions $\mathbb{P}$ and $\mathbb{Q}$ on $\mathbb{R}^N$. A sequence of polynomials $f : \mathbb{R}^N \to \mathbb{R}$ separates $\mathbb{P}$ and $\mathbb{Q}$ weakly if as $n \to \infty$,*

$$\Big(\max\big\{\mathrm{Var}_{\mathbb{P}}(f), \mathrm{Var}_{\mathbb{Q}}(f)\big\}\Big)^{1/2} = O\big(\big|\mathbb{E}_{\mathbb{P}}[f] - \mathbb{E}_{\mathbb{Q}}[f]\big|\big), \qquad \text{(Weak separation)}$$

and strongly if as $n \to \infty$,

$$\Big(\max\big\{\mathrm{Var}_{\mathbb{P}}(f), \mathrm{Var}_{\mathbb{Q}}(f)\big\}\Big)^{1/2} = o\big(\big|\mathbb{E}_{\mathbb{P}}[f] - \mathbb{E}_{\mathbb{Q}}[f]\big|\big). \qquad \text{(Strong separation)}$$

As is standard in the low-degree testing literature, we take separation as the definition of "success" for low-degree tests. Separation is a natural sufficient condition that allows two distributions to be distinguished using the output of a polynomial. Specifically, strong separation implies (by Chebyshev's inequality) that $\mathbb{P}, \mathbb{Q}$ can be distinguished with probability $1 - o(1)$, and weak separation implies that $\mathbb{P}, \mathbb{Q}$ can be distinguished with nontrivial advantage over a random guess (see Bandeira et al., 2022). We are now in position to state our main result.

Theorem 6 *Fix constants $0 < \delta < 1/2$ and $0 < \gamma < 2$. Consider a sequence $k = \Theta(n^{1/2+\delta})$.*

(a) (Lower bound) If $\gamma < 3(1/2 - \delta)$ then for any sequence of masks with $|M| = O(n^\gamma)$, any sequence of degree-$o(\log^2 n)$ polynomials fails to weakly separate $G(n, M)$ and $G(n, k, M)$.

(b) (Upper bound) If $\gamma > 3(1/2 - \delta)$ then there exists a sequence of masks with $|M| = O(n^\gamma)$ and a sequence of polynomials with constant degree $C(\gamma, \delta)$ that strongly separates $G(n, M)$ and $G(n, k, M)$.

We provide the proof of the lower bound in Section 3 and the proof of the upper bound in the supplement. The upper bound essentially simulates the degree counting algorithm of Mardia et al. (2020) using polynomials. The lower bound is our main contribution, and we now turn to an overview of its proof.

2.1. Overview of the lower bound proof

The main challenge in proving our hardness result lies in establishing the failure of low-degree polynomials for an *arbitrary* mask with a small number of (at most $O(n^\gamma)$) edges.

Step 1: Reducing to masks with small maximum mask degree. We gain intuition about algorithmically useful masks by studying Mardia et al. (2020)'s sublinear-time algorithm. This algorithm is based on Kučera (1995)'s observation that for large planted clique sizes (e.g. $k = n^{1/2+\delta}$), with high probability the degree of all planted clique vertices is much larger than the degree of all non-clique vertices. Consider the first $(n/k) \cdot \mathsf{polylog}(n)$ vertices. With high probability, if the graph were drawn from the planted distribution $G(n, 1/2, k)$ (see Definition 4), then at least one of these vertices will belong to the planted clique and have large degree. By contrast, if the graph were drawn from the null distribution $G(n, 1/2)$, all of these vertices will have small degree.

Simply estimating the degree of each of these $(n/k) \cdot \mathsf{polylog}(n)$ vertices and checking if any of them is 'large enough' to be a planted clique vertex will let us distinguish between the null and planted cases. Mardia et al. (2020)'s observation was that even just an estimate of these degrees obtained by subsampling $O(n^2/k^2)$ potential neighbours (instead of computing the degree exactly by looking at all $n - 1$ potential neighbours) is good enough to distinguish between the planted and null cases.

Clearly any mask that allows for such an estimate for enough vertices will be algorithmically useful. Luckily, since $k = \Theta(n^{1/2+\delta})$ and $\gamma < 3(1/2 - \delta)$, our mask does not have enough mask edges to query $\Omega(n^3/k^3)$ entries of the adjacency matrix. In fact, this means that only $o(n/k)$ vertices can hope to have a 'large' mask degree $\Omega(n^2/k^2)$. But with probability $1 - o(1)$, because the planted clique vertices are chosen uniformly at random, none of these vertices with large mask degree will be planted vertices even in the planted case.

As a result, for masks with too few ($O(n^\gamma)$) mask edges, we can safely ignore vertices with 'large' mask degree, as those vertices will behave the same under both the null and planted distributions (with high probability), and intuitively this means they carry no useful information. Hence it suffices to show hardness just for masks with a small maximum mask degree. Formally, we implement this reduction via the conditional low-degree likelihood method (see, e.g., Bandeira et al.

(2022), Proposition 6.2): we condition on the high-probability event that no vertices of 'large' mask degree belong to the planted clique[1].

Step 2: Reducing to masks with few mask vertices. Consider two masks, both of size $\widetilde{\Theta}(n^3/k^3)$, the first mask being any mask of this size that lets us estimate enough degrees as well as needed for Mardia et al. (2020)'s algorithm, and the second being a 'square' mask consisting of all potential mask edges involving only a fixed set of $\widetilde{\Theta}(n^{3/2}/k^{3/2})$ vertices. We know that the former mask is algorithmically useful, but the latter mask is not, as we later discuss in Step 3.

To express our takeaway from this, we should consider the mask degree distribution. For two masks with the same number of mask edges, the average of this distribution will be the same. However, the example above indicates that for masks with the same average mask degree, having a 'more skewed' (lots of large mask degrees as well as small mask degrees) degree distribution is more useful than having a 'more uniform' one.

Within the context of separation by low-degree polynomials, 'algorithmic utility' can be quantified by the norm of the low-degree likelihood ratio[2], which is a standard quantity (see e.g., Hopkins (2018)) defined in (1). We will use the intuition above to upper bound the 'algorithmic utility' (norm of the low-degree likelihood ratio) of a mask M by that of a closely related mask M'. We will obtain the mask M' through a small tweak to M that slightly skews its mask degree distribution while keeping the total number of mask edges the same.

If we repeat this process iteratively and use the fact that *we are only interested in masks with small maximum mask degree*, we can show that the 'algorithmic utility' of every such mask is upper bounded by the 'algorithmic utility' of a mask with only a few vertices. This reduction is the main technical contribution of our work.

Step 3: Analytically upper bounding the norm of the low-degree likelihood ratio for masks with few mask vertices. Once we know we only need to upper bound the 'algorithmic utility' (norm of the low-degree likelihood ratio) of masks with few vertices, we are ready to conclude. In particular, calculating such an upper bound analytically proves tractable using standard techniques, which gives our desired result.

Remark: Of the three steps above, Step 1 and Step 3 have natural analogues even when considering algorithmic classes other than low-degree polynomials. Only Step 2 seems to crucially rely on properties of low-degree polynomials. As alluded to in Section 1.2, it would be interesting to see if the above reduction program can be carried out for other algorithmic classes. In particular, it would be nice to show that—under the planted clique conjecture—the algorithm of Mardia et al. (2020) is essentially optimal among non-adaptive algorithms.

3. Proof of the lower bound: Theorem 6(a)

In this section, we provide a sequence of lemmas implementing the strategy outlined in Section 2.1. We defer the proofs of the lemmas to the supplement. This section culminates in the proof of the lower bound. We first require the following two definitions.

1. Conditioning is necessary here: Consider the mask which consists solely of all edges connected to the first vertex. This mask is clearly insufficient to detect the planted clique, yet the corresponding low-degree likelihood ratio blows up.
2. Technically in all our lemmas we will work with a natural upper bound to this quantity, but that detail is unimportant for this overview.

Definition 7 (Low-degree likelihood ratio upper bound: $\mathsf{LDUB}(n, M)$) *Given integers n, $k \leq n$, D and a mask M on ground set $[n]$, let X and X' be two independent draws from $\mathsf{Clique}(n, k)$ (as in Definition 3). For $i \in [n]$, let $Z_i := X_i \cdot X'_i$. The low-degree likelihood ratio upper bound at degree D is defined as*

$$\mathsf{LDUB}(n, M) := 1 + \sum_{d=1}^{D} \frac{1}{d!} \cdot \mathbb{E}\Big[\Big(\sum_{(i,j)\in M} Z_i \cdot Z_j\Big)^d\Big].$$

The values of k and D will always be clear from context, so we suppress them and denote the quantity as just $\mathsf{LDUB}(n, M)$ to simplify notation.

Definition 8 (Conditional low-degree likelihood ratio upper bound: $\mathsf{Cond}(n, M, S)$) *Given integers n, $k \leq n$, D, a mask M on ground set $[n]$, and a subset $S \subseteq [n]$, the conditional low-degree likelihood ratio upper bound $\mathsf{Cond}(n, M, S)$ is defined analogously to the low-degree likelihood ratio upper bound $\mathsf{LDUB}(n, M)$ with one crucial difference. The independent random vectors X and X' (as in Definition 7) are drawn from $\mathsf{Clique}(n, k, S)$ (rather than $\mathsf{Clique}(n, k)$). The rest of the definition proceeds as in Definition 7.*[3]

The motivation behind these definitions is the following. There is a standard quantity, the *norm of the (conditional) degree-D likelihood ratio*, to be defined in (1). It is well known that if this quantity is $1 + o(1)$ then this implies our goal: degree-D polynomials cannot achieve weak separation; see Bandeira et al. (2022, Proposition 6.2). Following Bandeira et al. (2021, Proposition B.1), $\mathsf{Cond}(n, M, S)$ provides a convenient upper bound on this quantity.

Algorithm 1: $\mathsf{Donate}(M, v \to u)$

```
// Give as many edges from v to u as possible
```
Input: Mask M, donating vertex $v \in V(M)$, receiving vertex $u \in V(M)$
Output: Mask $\mathsf{Donate}(M, v \to u)$
initialize $\mathsf{Donate}(M, v \to u) = M$
for *vertex* $s \in V(M) \setminus \{v, u\}$ **do**
  if *edge* $(s, v) \in M$ *and edge* $(s, u) \notin M$ **then**
    remove (s, v) from $\mathsf{Donate}(M, v \to u)$
    add (s, u) to $\mathsf{Donate}(M, v \to u)$
  end
end
output $\mathsf{Donate}(M, v \to u)$

Fact 3.1 (Donation leaves certain mask properties (almost) unchanged) *Let M be a mask with vertices $v, u \in V(M)$. It is an easy observation about Algorithm 1 that*

1. $\mathsf{Donate}(M, v \to u)$ *has the same number of edges as M.*
2. $V(\mathsf{Donate}(M, v \to u))$ *must be either $V(M)$ or $V(M) \setminus v$.*

3. Cleary, $\mathsf{Cond}(n, M, [n]) = \mathsf{LDUB}(n, M)$.

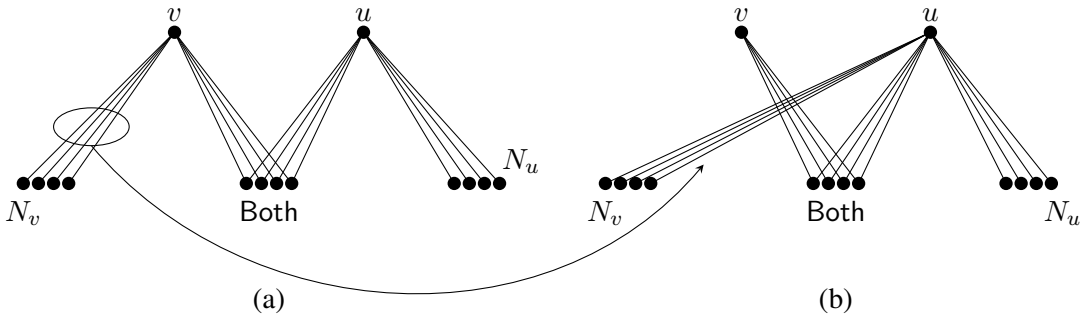

Figure 2: Illustration of the Donate process (Algorithm 1) using notation from the proof of Lemma 9. Panel (a) shows the neighborhoods of u (denoted as $N_u \cup \mathsf{Both}$) and v (denoted as $N_v \cup \mathsf{Both}$) before $\mathsf{Donate}(v \to u)$. Panel (b) shows the result: Each neighbor of v not originally connected to u (i.e. in N_v) is disconnected from v and connected to u instead.

Lemma 9 (Donation cannot hurt low-degree algorithms) *Let M be a mask with vertices $v, u \in V(M)$. Then, informally, donation from v to u (Algorithm 1) cannot decrease the low-degree likelihood ratio upper bound. Formally,*

$$\mathsf{LDUB}(n, M) \le \mathsf{LDUB}(n, \mathsf{Donate}(M, v \to u)).$$

Lemma 10 (Vertex Removal Lemma) *Let M be a mask with the following properties, where t is some positive integer.*

1. *The maximum M-degree of vertices in $V(M)$ is at most $2t$.*
2. *The number of vertices in $V(M)$ is large, with $|V(M)| \ge 2t + 2 + \frac{2|M|}{t}$.*

Then there exists a mask M' with the following properties.

1. *The maximum M'-degree of vertices in $V(M')$ is also at most $2t$.*
2. *The number of edges in M' is identical to the number of edges in M. That is, $|M'| = |M|$.*
3. *$\mathsf{LDUB}(n, M) \le \mathsf{LDUB}(n, M')$.*
4. *There are strictly fewer vertices in M' compared to M. That is, $|V(M')| < |V(M)|$.*

Lemma 11 (Converting masks with low maximum degree to masks with few vertices) *Let M be a mask where the maximum M-degree of any vertex in $V(M)$ is at most $2t$ for some positive integer t. Then there exists a mask M' with the following properties.*

1. *$\mathsf{LDUB}(n, M) \le \mathsf{LDUB}(n, M')$.*
2. *M' has very few vertices. That is, $|V(M')| \le 2t + 2 + \frac{2|M|}{t}$.*

Proof This follows in a straightforward manner from Lemma 10 which lets us remove vertices until the desired condition on the size of the vertex set is met. ■

Lemma 12 (Masks without enough vertices have small low-degree likelihood ratio) *For any mask M,*

$$\mathsf{LDUB}(n, M) \le 1 + \sum_{d=1}^{D} \frac{1}{d!} \cdot \left(\frac{2d \cdot \log e}{\log\left(\frac{2d \cdot n^2}{|V(M)| \cdot k^2} + 1\right)} \right)^{2d}.$$

Lemma 13 (Masks with small maximum degree have small low-degree likelihood ratio) *Let $n = \omega(1)$ and $k \le n$ be sequences of positive integers and M be a sequence of masks on the ground set $[n]$. Suppose that the maximum M-degree of any vertex in $V(M)$ is small. That is, there exists a sequence of positive integers t with the following properties.*

1. $\max_{v \in V(M)} \mathsf{deg}^M(v) \le 2t$.
2. $\left(2t + 2 + \frac{2|M|}{t}\right) \cdot \left(\frac{k^2}{n^2}\right) = O(n^{-\epsilon})$ *for some constant* $\epsilon > 0$.

Then for any sequence of degrees $D = o\left((\log n)^2\right)$, we have the low-degree likelihood ratio upper bound

$$\mathsf{LDUB}(n, M) = 1 + o(1).$$

The previous sequence of lemmas culminates in the following lemma, upon which our proof crucially relies.

Lemma 14 (Masks without enough edges have small conditional low-degree likelihood ratio) *Let $0 < \delta < 1/2$ be a constant. Let $n = \omega(1)$ and $k = \Theta(n^{1/2+\delta})$ be sequences of positive integers. Let M be a sequence of masks on the ground set $[n]$ without too many edges. That is,*

$$|M| \le O\left(n^{\gamma}\right) \text{ for some constant } \gamma < 3(1/2 - \delta).$$

Then there exists a sequence of subsets $S \subseteq [n]$ such that

1. $\mathbb{P}_{K \sim \mathsf{Clique}(n,k)}[\text{all nonzero coordinates of } K \text{ are in } S] = 1 - o(1)$.
2. *For any sequence of degrees $D = o(\log^2 n)$, the conditional low-degree likelihood ratio upper bound*[4] *(Definition 8) is small:*

$$\mathsf{Cond}(n, M, S) = 1 + o(1).$$

Equipped with Lemma 14, we provide the proof of our lower bound.

Proof of Theorem 6(a). Following Bandeira et al. (2022, Proposition 6.2), to rule out weak separation by degree-D polynomials, it suffices to show that the *norm of the conditional low-degree likelihood ratio*

$$\left\| \left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \right)^{\le D} \right\|_{\mathbb{Q}} := \sup_{\mathsf{deg}(f) \le D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \tag{1}$$

4. Recall that this quantity depends on k and D, but we do not denote this for notational simplicity.

is $1 + o(1)$, where $\mathbb{Q} = G(n, M)$, and $\mathbb{P}$ is $G(n, k, M)$ conditioned on some $(1 - o(1))$-probability event (which may depend on latent randomness such as the clique vertices). For our purposes, we choose to condition $G(n, k, M)$ on the event that all clique vertices lie in S, the set defined in Lemma 14. Note that Lemma 14 guarantees this to be a $(1 - o(1))$–probability event. We have the bounds

$$1 \le \left\| \left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \right)^{\le D} \right\|_{\mathbb{Q}}^2 \le \mathsf{Cond}(n, M, S),$$

where the first inequality comes from plugging in $f = 1$ to (1) and the second comes from Bandeira et al. (2021, Proposition B.1). Now the proof is complete, as Lemma 14 shows $\mathsf{Cond}(n, M, S) \le 1 + o(1)$. ■

Acknowledgments

We are thankful to the Simons Institute for the Theory of Computing for their hospitality during Fall 2021, where this work was initiated. K.A.V. was supported by European Research Council Advanced Grant 101019498. A.S.W. was partially supported by an Alfred P. Sloan Research Fellowship and NSF CAREER Award CCF-2338091.

References

Thomas D. Ahle. Sharp and simple bounds for the raw moments of the binomial and Poisson distributions. *Statist. Probab. Lett.*, 182, 2022. ISSN 0167-7152.

Miklós Ajtai. Generating hard instances of lattice problems. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96, page 99–108, New York, NY, USA, 1996. Association for Computing Machinery.

Matthew Aldridge, Oliver Johnson, and Jonathan Scarlett. Group testing: an information theory perspective. *Foundations and Trends in Communications and Information Theory*, 15(3-4):196–392, 2019.

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.

Ryan Alweiss, Chady Ben Hamida, Xiaoyu He, and Alexander Moreira. On the subgraph query problem. *Combinatorics, Probability and Computing*, 30(1):1–16, 2021.

Marshall Ball, Alon Rosen, Manuel Sabin, and Prashant Nalini Vasudevan. Average-case fine-grained hardness. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, page 483–496, New York, NY, USA, 2017. Association for Computing Machinery.

Afonso S. Bandeira, Jess Banks, Dmitriy Kunisky, Christopher 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, 2021.

Afonso S. Bandeira, Ahmed El Alaoui, Samuel Hopkins, Tselil Schramm, Alexander S. Wein, and Ilias Zadik. The Franz–Parisi criterion and computational trade-offs in high dimensional statistics. In *Advances in Neural Information Processing Systems*, pages 33831–33844, 2022.

Jess Banks, Sidhanth Mohanty, and Prasad Raghavendra. Local statistics, semidefinite programming, and community detection. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1298–1316. SIAM, 2021.

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

Enric Boix-Adserà, Matthew Brennan, and Guy Bresler. The average-case complexity of counting cliques in Erdős-Rényi hypergraphs. *SIAM Journal on Computing*, (0):FOCS19–39, 2021.

B. Bollobas and P. Erdős. Cliques in random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, 80(3):419–427, 1976.

Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 48–166. PMLR, 06–09 Jul 2018.

Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, and Philipp Loick. Optimal group testing. In *Conference on Learning Theory*, pages 1374–1388. PMLR, 2020.

Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, Alexander S Wein, and Ilias Zadik. Statistical and computational phase transitions in group testing. In *Conference on Learning Theory*, 2022.

David Conlon, Jacob Fox, Andrey Grinshpun, and Xiaoyu He. Online Ramsey numbers and the subgraph query problem. In *Building Bridges II: Mathematics of László Lovász*, pages 159–194. Springer, 2020.

Endre Csóka and András Pongrácz. Finding cliques and dense subgraphs using edge queries. *arXiv preprint arXiv:2310.06826*, 2023.

Mina Dalirrooyfard, Andrea Lincoln, and Virginia Vassilevska Williams. New techniques for proving fine-grained average-case hardness. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 774–785. IEEE, 2020.

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical review E*, 84(6):066106, 2011.

Yash Deshpande and Andrea Montanari. Finding hidden cliques of size n/e in nearly linear time. *Foundations of Computational Mathematics*, 15:1069–1128, 2015.

Abhishek Dhawan, Cheng Mao, and Alexander S Wein. Detection of dense subhypergraphs by low-degree polynomials. *arXiv preprint arXiv:2304.08135*, 2023.

Jian Ding, Hang Du, and Zhangsong Li. Low-degree hardness of detection for correlated Erdős-Rényi graphs. *arXiv preprint arXiv:2311.15931*, 2023a.

Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Subexponential-time algorithms for sparse PCA. *Foundations of Computational Mathematics*, pages 1–50, 2023b.

Uriel Feige and Tom Ferster. A tight bound for the clique query problem in two rounds. *arXiv preprint arXiv:2112.06072*, 2021.

Uriel Feige, David Gamarnik, Joe Neeman, Miklós Z Rácz, and Prasad Tetali. Finding cliques using few probes. *Random Structures & Algorithms*, 56(1):142–153, 2020.

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.

Asaf Ferber, Michael Krivelevich, Benny Sudakov, and Pedro Vieira. Finding Hamilton cycles in random graphs with few queries. *Random Structures & Algorithms*, 49(4):635–668, 2016.

David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Low-degree hardness of random optimization problems. In *61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 131–140. IEEE, 2020.

Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017.

Oded Goldreich and Guy Rothblum. Counting t-cliques: Worst-case to average-case reductions and direct interactive proof systems. In *Foundations of Computer Science (FOCS)*, pages 77–88, 2018.

Shuichi Hirahara and Nobutaka Shimizu. Nearly optimal average-case complexity of counting bicliques under SETH. In *Symposium on Discrete Algorithms (SODA)*, pages 2346–2365, 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.

Wasim Huleihel, Arya Mazumdar, and Soumyabrata Pal. Random subgraph detection using queries. *arXiv preprint arXiv:2110.00744*, 2021.

Mark Jerrum. Large cliques elude the Metropolis process. *Random Structures & Algorithms*, 3(4): 347–359, 1992.

Kumar Joag-Dev and Frank Proschan. Negative association of random variables, with applications. *Ann. Statist.*, 11(1):286–295, 1983. ISSN 0090-5364.

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 *Mathematical Analysis, its Applications and Computation: ISAAC*. 2022.

Jay Mardia, Hilal Asi, and Kabir Aladin Chandrasekher. Finding planted cliques in sublinear time. *arXiv preprint arXiv:2004.12002*, 2020.

Miklós Z Rácz and Benjamin Schiffer. Finding a planted clique by adaptive probing. *arXiv preprint arXiv:1903.12050*, 2019.

Cyrus Rashtchian, David Woodruff, Peng Ye, and Hanlin Zhu. Average-case communication complexity of statistical problems. In *Conference on Learning Theory*, pages 3859–3886. PMLR, 2021.

Oded Regev. The learning with errors problem. In *Conference on Computational Complexity*, pages 191–204, 2010.

Emile Richard and Andrea Montanari. A statistical model for tensor PCA. *Advances in neural information processing systems*, 27, 2014.

Ronitt Rubinfeld and Asaf Shapira. Sublinear time algorithms. *SIAM Journal on Discrete Mathematics*, 25(4):1562–1588, 2011.

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

Qi-Man Shao. A comparison theorem on moment inequalities between negatively associated and independent random variables. *J. Theoret. Probab.*, 13(2):343–356, 2000. ISSN 0894-9840.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

Appendix A. Deferred proofs for the lower bound

In this section, we provide the proofs of each of the stated lemmas in Section 3 in sequence, starting with Lemma 9.

Proof of Lemma 9. Let $M' = \mathsf{Donate}(M, v \to u)$. Denote $\Phi(M) := \sum_{(i,j)\in M} Z_i \cdot Z_j$ and $\Phi(M') := \sum_{(i,j)\in M'} Z_i \cdot Z_j$, where the Z_i's are as in Definition 7. We will show that for any positive integer d,

$$\mathbb{E}\left[\left(\Phi(M)\right)^d\right] \le \mathbb{E}\left[\left(\Phi(M')\right)^d\right],$$

since the desired conclusion follows easily from this.

In fact, by the law of total expectation, it suffices to simply show the following inequality for any set $\{z_i \in \{0,1\} : i \in [n] \setminus \{v,u\}\}$ in the suppport of $\{Z_i : i \in [n] \setminus \{v,u\}\}$:

$$\mathbb{E}\left[(\Phi(M))^d \,|Z_i = z_i, \forall i \in [n] \setminus \{v,u\}\right] \leq \mathbb{E}\left[\left(\Phi(M')\right)^d \,|Z_i = z_i, \forall i \in [n] \setminus \{v,u\}\right]. \quad (2)$$

Hence, for the rest of the proof, we condition on $Z_i = z_i$ for $i \in [n] \setminus \{v,u\}$ and the only randomness is in the random variables Z_v, Z_u. For the rest of the proof when we take expectations, even though the aforementioned conditioning exists, we will not indicate it notationally.

1. Let N_u be the vertices in $V(M) \setminus \{v,u\}$ connected (by edges in M) to u but not to v.
2. Let N_v be the vertices in $V(M) \setminus \{v,u\}$ connected (by edges in M) to v but not to u.
3. Let Both be the vertices in $V(M) \setminus \{v,u\}$ connected (by edges in M) to both u and v.
4. Let Other be the set of edges in M involving neither u nor v.

With this notation and the fact that $\mathbb{1}_{(u,v)\in M'} = \mathbb{1}_{(u,v)\in M}$, we can rewrite

$$\Phi(M) = \sum_{s\in N_v} z_s \cdot Z_v + \sum_{s\in N_u} z_s \cdot Z_u + \sum_{s\in \mathsf{Both}} z_s \cdot (Z_u + Z_v) + \sum_{(i,j)\in \mathsf{Other}} z_i \cdot z_j + \mathbb{1}_{(u,v)\in M} \cdot Z_u Z_v$$

and

$$\Phi(M') = \sum_{s\in N_u \cup N_v} z_s \cdot Z_u + \sum_{s\in \mathsf{Both}} z_s \cdot (Z_u + Z_v) + \sum_{(i,j)\in \mathsf{Other}} z_i \cdot z_j + \mathbb{1}_{(u,v)\in M} \cdot Z_u Z_v.$$

It is clear from the expressions above that if $Z_u = Z_v$, we have $\Phi(M) = \Phi(M')$, and hence

$$\mathbb{E}\left[(\Phi(M))^d \,|Z_u = Z_v\right] = \mathbb{E}\left[\left(\Phi(M')\right)^d \,|Z_u = Z_v\right].$$

This fact combined with the law of total expectation means we now only need to prove

$$\mathbb{E}\left[(\Phi(M))^d \,|Z_u \neq Z_v\right] \leq \mathbb{E}\left[\left(\Phi(M')\right)^d \,|Z_u \neq Z_v\right],$$

assuming the event $Z_u \neq Z_v$ has positive conditional probability (if it does not, we are already done).

For the rest of the proof, assume $Z_u \neq Z_v$. We must have $Z_u + Z_v = 1$ and $Z_u Z_v = 0$. As a result, there exist non-negative constants c, c_u, and c_v (which depend on the values of z_i for $i \in [n] \setminus \{v,u\}$) such that

$$\Phi(M) = c_u \cdot Z_u + c_v \cdot Z_v + c$$

and

$$\Phi(M') = (c_u + c_v) \cdot Z_u + c.$$

By symmetry, the events $Z_u = 1, Z_v = 0$ and $Z_u = 0, Z_v = 1$ both have probability $1/2$ (as long as the events we have conditioned on so far occur with positive probability). Hence,

$$\mathbb{E}\left[(\Phi(M))^d \,|Z_u \neq Z_v\right] = \frac{(c_u + c)^d + (c_v + c)^d}{2}$$

and

$$\mathbb{E}\left[\left(\Phi(M')\right)^d \middle| Z_u \neq Z_v\right] = \frac{(c_u + c_v + c)^d + c^d}{2}.$$

For positive integers d, the function $f(x) := (c_u + x)^d - x^d$ is non-decreasing for $x \geq 0$. This follows by using the binomial expansion and elementary calculus, along with the fact that $c_u \geq 0$.

Because $0 \leq c \leq c_v + c$, this means

$$(c_u + c)^d - c^d \leq (c_u + c_v + c)^d - (c_v + c)^d.$$

Rearranging this inequality yields

$$\mathbb{E}\left[(\Phi(M))^d \middle| Z_u \neq Z_v\right] \leq \mathbb{E}\left[\left(\Phi(M')\right)^d \middle| Z_u \neq Z_v\right],$$

which completes the proof. ■

Proof of Lemma 10. Let Low be the subset of vertices in $V(M)$ with M-degree at most t. That is, $\mathsf{Low} := \left\{v \in V(M) : \deg^M(v) \leq t\right\}$. Consider the set $V(M) \setminus \mathsf{Low}$. Since every vertex in this set has M-degree greater than t, there must be at least $\frac{(|V(M)| - |\mathsf{Low}|) \cdot t}{2}$ edges in M. Consequently,

$$\frac{\left(2t + 2 + \frac{2|M|}{t} - |\mathsf{Low}|\right) \cdot t}{2} \leq \frac{(|V(M)| - |\mathsf{Low}|) \cdot t}{2} \leq |M|.$$

Rearranging this inequality yields the useful conclusion $|\mathsf{Low}| \geq 2t + 2$.

Arbitrarily order the vertices in Low, naming them $\{v_1, v_2, ..., v_{|\mathsf{Low}|}\}$ and run the following algorithm to obtain the mask M'. Since $|\mathsf{Low}| \geq 2$, the algorithm is not vacuous.

---

initialize $M_2 = M$
for $i \in \{2, ..., |\mathsf{Low}|\}$ **do**
  if $v_1 \in V(M_i)$ **then**
    $M_{i+1} = \mathsf{Donate}(M_i, v_1 \to v_i)$ (Algorithm 1)
  end
  else
    $M_{i+1} = M_i$
  end
end
output $M' = M_{|\mathsf{Low}|+1}$

For the above process to be well defined, we need every call to Algorithm 1 to be well defined. Since we always check whether $v_1 \in V(M_i)$ before invoking Algorithm 1, and $\{v_2, ..., v_{|\mathsf{Low}|}\}$ must be in $V(M_i)$ at every iteration by Fact 3.1, every such call is well defined.

By repeated applications of Fact 3.1, it is clear that $V(M')$ must be either $V(M)$ or $V(M) \setminus v_1$.

1. By construction, only the vertices in $\mathsf{Low} \setminus v_1$ can have greater M'-degree than M-degree. However, since v_1, which is in Low, is the only vertex that donates edges in our construction, the degree of any other vertex can increase by at most t. Since every vertex in Low has M-degree at most t, they can have M'-degree at most $2t$. Hence, the maximum M'-degree of any vertex in $V(M')$ is at most $2t$.

2. $|M'| = |M|$ by repeated applications of Fact 3.1.

3. $\mathsf{LDUB}(n, M) \leq \mathsf{LDUB}(n, M')$ by repeated applications of Lemma 9.

4. To show $|V(M')| < |V(M)|$, we just need to show that $v_1 \notin V(M')$.

   Suppose this is false, and $v_1 \in V(M')$. Then there exists a $u \in V(M') \setminus v_1$ such that $(v_1, u) \in M'$. Further, we must have taken the "if" branch in every iteration of our construction of M', and v_1 must have donated edges (as specified by Algorithm 1) to every vertex in $\{2, ..., |\mathsf{Low}|\}$. Then the only way for (v_1, u) to exist in M' is if (v_i, u) exists in M' for all $v_i \in \mathsf{Low} \setminus u$. This means u must have M'-degree at least $|\mathsf{Low}| - 1 \geq 2t + 1$. This contradicts the fact that every vertex in $V(M')$ has M'-degree at most $2t$. Our assumption must be false, and we must have $v_1 \notin V(M')$.

This completes the proof of Lemma 10. ■

Proof of Lemma 12. For any mask M, it is an easy observation that adding edges to M cannot decrease $\mathsf{LDUB}(n, M)$. Hence we have

$$\mathsf{LDUB}(n, M) \leq 1 + \sum_{d=1}^{D} \frac{1}{d!} \cdot \mathbb{E}\left[\left(\sum_{(i,j) \in \binom{V(M)}{2}} Z_i \cdot Z_j\right)^d\right]$$

where Z_i, Z_j are $\{0, 1\}$-valued random variables as in Definition 7.

Let $H := \sum_{i \in V(M)} Z_i$. Then we have $\sum_{(i,j) \in \binom{V(M)}{2}} Z_i \cdot Z_j \leq \sum_{i,j \in V(M)} Z_i \cdot Z_j = H^2$, and this gives

$$\mathsf{LDUB}(n, M) \leq 1 + \sum_{d=1}^{D} \frac{1}{d!} \cdot \mathbb{E}\left[H^{2d}\right].$$

Unfortunately, H is the sum of dependent random variables. Since it is often easier to analyze the moments of sums of independent random variables, we use the following approach.

1. The $\{Z_i : i \in V(M)\}$ form a set of negatively associated (henceforth NA) random variables. This can be proved as follows.

   (a) The $\{X_i : i \in [n]\}$ (as in Definition 7) are NA because they can be viewed as a permutation distribution (Joag-Dev and Proschan, 1983, Definition 2.10 and Theorem 2.11). Similarly, the $\{X'_i : i \in [n]\}$ are NA and independent of the $\{X_i : i \in [n]\}$.

   (b) $\{X_i : i \in [n]\} \cup \{X'_i : i \in [n]\}$ are jointly NA because they are the union of independent sets of NA random variables (Joag-Dev and Proschan, 1983, Property P7).

   (c) $\{Z_i = X_i \cdot X'_i : i \in [n]\}$ are NA because they are non-decreasing functions of disjoint subsets of NA random variables (Joag-Dev and Proschan, 1983, Propety P6).

   (d) $\{Z_i : i \in V(M)\}$ are NA because they are a subset of NA random variables (Joag-Dev and Proschan, 1983, Property P4).

2. Negatively associated random variables can be coupled to independent random variables.

   Let $\{Z_i^* : i \in V(M)\}$ be a collection of independent random variables where each Z_i^* has the same marginal distribution as Z_i. Let $H^* := \sum_{i \in V(M)} Z_i^*$. For positive integers d, the function $f(x) = x^{2d}$ is convex when $x \geq 0$. Hence we can use (Shao, 2000, Theorem 1) to conclude

   $$\mathbb{E}\left[H^{2d}\right] \leq \mathbb{E}\left[(H^*)^{2d}\right].$$

3. Each Z_i (and hence Z_i^*) is 1 with probability k^2/n^2 and 0 otherwise. This means H^* is a Binomial random variable with $|V(M)|$ trials each having success probability k^2/n^2. We can now use known bounds on the moments of Binomial random variables (e.g., (Ahle, 2022, Corollary 1)) and obtain

   $$\mathbb{E}\left[H^{*2d}\right] \leq \left(\frac{2d \cdot \log e}{\log\left(\frac{2d \cdot n^2}{|V(M)| \cdot k^2} + 1\right)}\right)^{2d}.$$

Putting this all together completes the proof. ∎

Proof of Lemma 13. Our first hypothesis that $\max_{v \in V(M)} \deg^M(v) \leq 2t$ immediately lets us combine the following:

- the simplifcation to masks without too many vertices from Lemma 11,
- the calculation of the low-degree likelihood ratio upper bound based on the number of vertices in the mask from Lemma 12.

Defining $v_{\mathsf{max}} := \left(2t + 2 + \frac{2|M|}{t}\right)$ for notational convenience, this yields

$$\begin{aligned}
\mathsf{LDUB}(n, M) &\leq 1 + \sum_{d=1}^{D} \frac{1}{d!} \cdot \left(\frac{2d \cdot \log e}{\log\left(\frac{2d \cdot n^2}{v_{\mathsf{max}} \cdot k^2} + 1\right)}\right)^{2d} \leq 1 + \sum_{d=1}^{D} \left(\frac{2d \cdot \log e}{(d!)^{1/2d} \cdot \log\left(\frac{n^2}{v_{\mathsf{max}} \cdot k^2}\right)}\right)^{2d} \\
&\leq 1 + \sum_{d=1}^{D} \left(\frac{2e \cdot \sqrt{d} \cdot \log e}{\log\left(\frac{n^2}{v_{\mathsf{max}} \cdot k^2}\right)}\right)^{2d} \leq 1 + \sum_{d=1}^{D} \left(\frac{2e \cdot \sqrt{D} \cdot \log e}{\log\left(\frac{n^2}{v_{\mathsf{max}} \cdot k^2}\right)}\right)^{2d} \\
&\leq 1 + \sum_{d=1}^{\infty} \left(\frac{2e \cdot \sqrt{D} \cdot \log e}{\log\left(\frac{n^2}{v_{\mathsf{max}} \cdot k^2}\right)}\right)^{2d} \leq \left(1 - \left(\frac{2e \cdot \sqrt{D} \cdot \log e}{\log\left(\frac{n^2}{v_{\mathsf{max}} \cdot k^2}\right)}\right)^{2}\right)^{-1} \\
&= 1 + o(1).
\end{aligned}$$

Above, we have used the inequality $\frac{1}{d!} \leq \left(\frac{e}{d}\right)^d$, the formula for the sum of a geometric series, the fact that $d \leq D = o\left((\log n)^2\right)$, and our second hypothesis $v_{\mathsf{max}} \cdot \left(\frac{k^2}{n^2}\right) = \left(2t + 2 + \frac{2|M|}{t}\right) \cdot \left(\frac{k^2}{n^2}\right) = O(n^{-\epsilon})$. ∎

Proof of Lemma 14. Let $\mathsf{grow} = \log n = \omega(1)$[5].

Let $t := |M| \cdot \frac{k}{n} \cdot \mathsf{grow}$ and define $S \subseteq [n]$ as the subset of vertices in M whose M-degree (Definition 1) is at most $2t$. Because there are only $|M|$ edges in M, we must have $|[n] \setminus S| \cdot \frac{2t}{2} \leq |M|$ by the pigeonhole principle. This gives $|[n] \setminus S| \leq \frac{|M|}{t}$.

1. By a union bound, the probability that $K \sim \mathsf{Clique}(n,k)$ has a nonzero coordinate in $[n] \setminus S$ is at most
$$|[n] \setminus S| \cdot \frac{k}{n} \leq \frac{|M|}{t} \cdot \frac{k}{n} = \frac{1}{\mathsf{grow}} \leq o(1).$$
Thus we have
$$\mathop{\mathbb{P}}_{K \sim \mathsf{Clique}(n,k)} [\text{all nonzero coordinates of } K \text{ are in } S] \geq 1 - o(1).$$

2. Let $n' := |S|$ and fix any bijection $\phi : [n'] \to S$. Define the mask M_S on ground set $[n']$ as
$$M_S := \left\{ (i,j) \in \binom{[n']}{2} : (\phi(i), \phi(j)) \in M \right\}.$$
This is the natural restriction of the mask M onto a ground set of size n' corresponding to S. It is straightforward to observe that for any n, k, D, M we have the following equality between a conditional low-degree likelihood ratio upper bound and a low-degree likelihood ratio upper bound:
$$\mathsf{Cond}(n, M, S) = \mathsf{LDUB}(n', M_S).$$
Further,
   (a) By construction, $n' = \Theta(n) = \omega(1)$ and $k = \Theta\left((n')^{1/2+\delta}\right)$.
   (b) $\max_{i \in V(M_S)} \mathsf{deg}^{M_S}(i) \leq 2t$. That is, the maximum mask degree of any vertex in $V(M_S)$ is at most $2t$. This is because the M_S-degree of any vertex $i \in [n']$ is at most the M-degree of the vertex $\phi(i) \in S$, and the latter is at most $2t$ by construction of S.
   (c) By construction, we also have $|M_S| \leq |M| \leq O\left(n^{\gamma}\right)$. Using this with the definition of grow and t and the facts $0 < \delta < 1/2$, $\gamma < 3(1/2 - \delta)$ gives
$$\left(2t + 2 + \frac{2|M_S|}{t}\right) \cdot \left(\frac{k^2}{(n')^2}\right) \leq O\left((n')^{-\epsilon'}\right).$$
   for some constant $\epsilon' > 0$.

This lets us invoke Lemma 13 to conclude $\mathsf{LDUB}(n', M_S) \leq 1 + o(1)$ and complete the proof.

■

5. Any sequence that grows to infinity slower than a polynomial would work.

Appendix B. Proof of the upper bound: Theorem 6(b)

We begin with a few preliminaries before turning to the proof of the theorem. Our strategy is to implement the degree counting algorithms of Kučera (1995); Mardia et al. (2020) via low-degree polynomials. To this end, we note that it suffices to furnish a mask M with $|M| = O(n^\gamma)$, where $\gamma > 3(1/2-\delta)$ and a polynomial f which, when evaluated on the masked observations corresponding to M, strongly separates (in the sense of Definition 5) the distributions $\mathbb{P} = G(n,k,M)$ (see Definition 4) and $\mathbb{Q} = G(n,M)$ (see Definition 2).

Towards constructing this mask, we define the gap $\epsilon = \gamma - 3(1/2-\delta) > 0$ and the pair R and L as

$$R = \min\{\lceil (n/k)^2 \cdot n^{2\epsilon/3} \rceil, \lceil n/2 \rceil\} \qquad \text{and} \qquad L = \lceil \sqrt{R} \rceil. \tag{3}$$

Note that $R \le \lceil n/2 \rceil$ and $L \le \sqrt{n} = o(n)$. This thus ensures that the vertex sets $V_L := \{1,2,\dots,L\}$ and $V_R := \{n-R+1,\dots,n\}$ are disjoint. We will consider the 'rectangular' mask $M = V_L \times V_R$, observing that it satisfies $|M| = R \cdot L = O(n^\gamma)$ by construction.

We turn now to the construction of our distinguishing polynomial $f : \{-1,1\}^{V_L \times V_R} \to \mathbb{R}$. In order to build intuition, let $K_1, K_2, \dots, K_n$ denote binary indicators of whether or not vertex $i \in [n]$ belongs to the planted clique. That is, under the null distribution $\mathbb{Q}$, each of the $\{K_i\}_{i\in[n]}$ are identically zero, whereas under the planted distribution $\mathbb{P}$, $\{K_i\}_{i\in[n]} \sim \mathsf{Clique}(n,k)$ (as in Definition 3). In the sequel, we show that the polynomial $(K_1, K_2, \dots, K_L) \mapsto \sum_{i\in V_L} K_i$ strongly separates $\mathbb{P}$ and $\mathbb{Q}$. Our distinguishing polynomial emulates this oracle polynomial by thresholding estimates related to the degree counts for each vertex in V_L. To do so, we require the following lemma from Schramm and Wein (2022), which provides a polynomial approximation to the binary threshold function.

Lemma 15 (Schramm and Wein (2022), Prop. 4.1) *For any integer $\ell \in \mathbb{Z}_{\ge 0}$, consider the degree-$(2\ell+1)$ polynomial $\tau = \tau_\ell : \mathbb{R} \to \mathbb{R}$, $\tau(y) = (2\ell+1)\binom{2\ell}{\ell}\int_0^y t^\ell (1-t)^\ell \mathrm{d}t$. For any $b \in \{0,1\}$ and any $0 \le \Delta \le \frac{1}{2}$, the following holds*

$$|\tau(y) - b| \le \left(\ell + \frac{1}{2}\right)(6\Delta)^\ell, \qquad \textit{for any } y \in \mathbb{R} \textit{ such that} \qquad |y-b| \le \Delta.$$

Let $Y = \{Y_{ij}\}_{i\in[V_L], j\in[V_R]} \in \{-1,1\}^{V_L \times V_R}$ denote our observations. Using $\tau = \tau_\ell$ as defined in Lemma 15, with $\ell = \lceil 3/\epsilon + 1/\delta \rceil$, we define our separating polynomial $f : \{-1,1\}^{L\times R} \to \mathbb{R}$ as

$$f(Y) = \sum_{i\in V_L} \tau\Big(\frac{n}{k} \cdot \frac{1}{R} \sum_{j\in V_R} Y_{ij}\Big). \tag{4}$$

For convenience, we will use the shorthand

$$g_i(Y) = \frac{n}{k} \cdot \frac{1}{R} \sum_{j\in V_R} Y_{ij}.$$

The key property of our polynomial f is that $\tau(g_i(Y))$ emulates the clique indicators K_i up to a small error, as summarized by the following lemma.

Lemma 16 *Under the conditions of Theorem 6, let L and R be as in (3) and consider the mask $M = V_L \times V_R$. Let $Y \in \{-1,1\}^M$ denote the observations. Suppose that $\ell = \lceil 3/\epsilon + 1/\delta \rceil$ and let $\tau = \tau_\ell$ be as defined in Lemma 15. Then, if either $Y \sim \mathbb{P} = G(n,k,M)$ or $Y \sim \mathbb{Q} = G(n,M)$, both*

$$\mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^2\big] = o\Big(\Big(\frac{k}{n}\Big)^2\Big) \qquad \text{and} \qquad \mathbb{E}_{\mathbb{Q}}\big[\{\tau(g_1(Y))\}^2\big] = o\Big(\Big(\frac{k}{n}\Big)^2\Big). \tag{5}$$

We defer the proof of this lemma to the end of the section. Equipped with this lemma, we turn to the proof of Theorem 6(b).

Proof of Theorem 6(b). We turn to lower bounding the expectation gap and upper bounding the variance induced by the polynomial f in (4).
Lower bounding the expectation gap: Expanding yields

$$\begin{aligned}\big|\mathbb{E}_{\mathbb{P}}[f(Y)] - \mathbb{E}_{\mathbb{Q}}[f(Y)]\big| &= L\big|\mathbb{E}_{\mathbb{P}}[K_1] + \mathbb{E}_{\mathbb{P}}[\tau(g_1(Y)) - K_1] - \mathbb{E}_{\mathbb{Q}}[\tau(g_1(Y)) - K_1]\big| \\ &= L\Big|\frac{k}{n} + \mathbb{E}_{\mathbb{P}}[\tau(g_1(Y)) - K_1] - \mathbb{E}_{\mathbb{Q}}[\tau(g_1(Y))]\Big|,\end{aligned}$$

where we have used the fact that $K_1 = 0$ under the null distribution $\mathbb{Q}$. Then, applying the triangle inequality in conjunction with Jensen's inequality yields

$$\begin{aligned}\big|\mathbb{E}_{\mathbb{P}}[f(Y)] - \mathbb{E}_{\mathbb{Q}}[f(Y)]\big| &\geq \frac{Lk}{n} - L\big|\mathbb{E}_{\mathbb{P}}[g_1(Y) - K_1] - \mathbb{E}_{\mathbb{Q}}[g_1(Y)]\big| \\ &\geq \frac{Lk}{n} - L\sqrt{\mathbb{E}_{\mathbb{P}}\big[\{g_1(Y) - K_1\}^2\big]} - L\sqrt{\mathbb{E}_{\mathbb{Q}}\big[\{g_1(Y)\}^2\big]}.\end{aligned}$$

We conclude by applying Lemma 16 to obtain the bound

$$\big|\mathbb{E}_{\mathbb{P}}[f(Y)] - \mathbb{E}_{\mathbb{Q}}[f(Y)]\big| = \frac{Lk}{n} - o\Big(\frac{Lk}{n}\Big) = \Omega\big(\min\{n^\delta, n^{\epsilon/3}\}\big).$$

Upper bounding the variance:
Planted distribution:

$$\mathrm{Var}_{\mathbb{P}}(f) = \mathrm{Var}_{\mathbb{P}}\Big(\sum_{i\in V_L}[\tau(g_i(Y)) - K_i] + K_i\Big) \leq 2\,\mathrm{Var}_{\mathbb{P}}\Big(\sum_{i\in V_L}[\tau(g_i(Y)) - K_i]\Big) + 2\,\mathrm{Var}_{\mathbb{P}}\Big(\sum_{i\in V_L} K_i\Big).$$

To bound the first term, we apply Lemma 16 to obtain

$$2\,\mathrm{Var}_{\mathbb{P}}\Big(\sum_{i\in V_L}[\tau(g_i(Y)) - K_i]\Big) \leq 2L^2\mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^2\big] = o\big(\big[\min\{n^\delta, n^{\epsilon/3}\}\big]^2\big).$$

Moreover, we compute

$$\begin{aligned}2\,\mathrm{Var}_{\mathbb{P}}\Big(\sum_{i\in V_L} K_i\Big) = 2L\,\mathrm{Var}_{\mathbb{P}}(K_1) + 2L(L-1)\mathsf{Cov}_{\mathbb{P}}(K_1,K_2) &\leq 2L\Big(\frac{k}{n}\Big) + 2L(L-1)\Big(\frac{k\cdot(k-1)}{n\cdot(n-1)} - \frac{k^2}{n^2}\Big) \\ &= O\big(\min\{n^\delta, n^{\epsilon/3}\}\big).\end{aligned}$$

Null distribution: Proceeding similarly yields

$$\mathrm{Var}_{\mathbb{Q}}(f) \le 2L^2\mathbb{E}_{\mathbb{Q}}[\tau(g_1(Y))^2] = o\big(\big[\min\{n^{\delta}, n^{\epsilon/3}\}\big]^2\big).$$

Putting the pieces together then yields

$$\Big(\max\big\{\mathrm{Var}_{\mathbb{P}}(f), \mathrm{Var}_{\mathbb{Q}}(f)\big\}\Big)^{1/2} = o\big(\min\{n^{\delta}, n^{\epsilon/3}\}\big) = o\big(\big|\mathbb{E}_{\mathbb{P}}[f(Y)] - \mathbb{E}_{\mathbb{Q}}[f(Y)]\big|\big),$$

which confirms that f strongly separates $\mathbb{P}$ and $\mathbb{Q}$. ■

Proof of Lemma 16. Note that, conditioned on the event that vertex i is not contained in the clique, $\mathbb{E}_{\mathbb{P}}[g_i(Y) \mid K_i = 0] = 0$, whereas conditioned on the event that vertex i is contained in the clique, $\mathbb{E}_{\mathbb{P}}[g_i(Y) \mid K_i = 1] = 1$. Then, conditioned on $K_i = 0$, by Bernstein's inequality (e.g., (Vershynin, 2018, Theorem 2.8.4)),

$$\mathbb{P}\bigg(\Big|\sum_{j\in V_R} Y_{ij}\Big| \ge t \;\Big|\; K_i = 0\bigg) \le 2\exp\Big(-\frac{t^2/2}{R + t/3}\Big).$$

We next express Y_{ij} as $Y_{ij} = (1 - K_iK_j)A_{ij} + K_iK_j$, where $A_{ij} \overset{\text{i.i.d.}}{\sim} \mathsf{Rademacher}(1/2)$ for $1 \le i < j \le n$. Then, conditionally on the event $\{K_i = 1\}$, the collection $\{Y_{ij}\}_{j\in V_R}$ are monotone functions of the negatively associated random variables $\{K_j\}_{j\in V_R}$ and the independent random variables $\{A_{ij}\}_{j\in V_R}$. Consequently, by Joag-Dev and Proschan (1983, Property P6) we find that $\{Y_{ij}\}_{j\in V_R}$ forms a negatively associated collection, conditionally on the event $\{K_i = 1\}$. Next, let $\{Y^*_{ij}\}_{j\in V_R}$ denote a collection of independent random variables with the same marginal distributions as Y_{ij}. Further let $S = \sum_{j\in V_R} Y_{ij} - \mathbb{E}[Y_{ij} \mid K_i = 1]$ and define S^* similarly. Then, applying Shao (2000, Theorem 1), we obtain the MGF bound $\mathbb{E}[\exp\{\lambda S\} \mid K_i = 1] \le \mathbb{E}[\exp\{\lambda S^*\}]$, for all $\lambda \in \mathbb{R}$ such that the RHS exists. The discussion above thus shows that Bernstein's inequality for bounded random variables (Vershynin, 2018, Theorem 2.8.4) continues to hold for the collection $\{Y_{ij}\}_{j\in V_R}$, whence we obtain the inequality

$$\mathbb{P}\bigg(\Big|\sum_{j\in V_R} Y_{ij} - R\cdot\frac{k}{n}\Big| \ge t \;\Big|\; K_i = 1\bigg) \le 2\exp\bigg(-\frac{t^2/2}{R + 2t/3}\bigg).$$

Combining the previous two displays yields the inequality

$$\mathbb{P}\bigg(\big|g_i(Y) - K_i\big| \ge t\cdot\frac{n}{kR}\bigg) \le 2\exp\bigg(-\frac{t^2/2}{R + 2t/3}\bigg). \tag{6}$$

Equipped with this concentration inequality, we turn to bounding the second moment $\mathbb{E}_{\mathbb{P}}[\{\tau(g_1(Y)) - K_1\}^2]$, which we decompose as

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^2\big] &= \mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^2\mathbb{1}\{|g_1(Y) - K_1| \le 1/2\}\big] \\ &\quad + \mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^2\mathbb{1}\{|g_1(Y) - K_1| > 1/2\}\big]. \end{aligned}\tag{7}$$

We claim the following two upper bounds, deferring their proofs to the end

$$\mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^2\mathbb{1}\{|g_1(Y) - K_1| \le 1/2\}\big] \le C_\ell\cdot\max\Big\{n^{-\epsilon\ell/3}, n^{-\delta\ell}\Big\} \tag{8a}$$

$$\mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^2\mathbb{1}\{|g_1(Y) - K_1| > 1/2\}\big] \le C_\ell\cdot n^{4\ell+2}\max\Big\{e^{-cn^{2\epsilon/3}}, e^{-cn^{2\delta}}\Big\}, \tag{8b}$$

where C_ℓ denotes a constant which depends only on ℓ and may change line by line. Then, taking n large enough to ensure that the RHS of inquality (8b) is upper bounded by $C_\ell \cdot \max\{n^{-\epsilon\ell/3}, n^{-\delta\ell}\}$ and repeating similar steps under $\mathbb{Q}$, we obtain the pair of bounds

$$\mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^2\big] \le C_\ell \cdot \max\Big\{n^{-\epsilon\ell/3}, n^{-\delta\ell}\Big\} \qquad \text{and}$$
$$\mathbb{E}_{\mathbb{Q}}\big[\{\tau(g_1(Y))\}^2\big] \le C_\ell \cdot \max\Big\{n^{-\epsilon\ell/3}, n^{-\delta\ell}\Big\},$$

as desired. The desired result follows by noting that $\ell \ge \max\{3/\epsilon, 1/\delta\}$ and $1/n = o((k/n)^2)$. It remains to establish the pair of inequalities (8a) and (8b).

Proof of the inequality (8a)**.** Applying Lemma 15 yields

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^2 \mathbb{1}\{|g_1(Y) - K_1| \le 1/2\}\big] &\le 6^\ell(\ell + 1/2)\mathbb{E}_{\mathbb{P}}\big[|g_1(Y) - K_1|^\ell \mathbb{1}\{|g_1(Y) - K_1| \le 1/2\}\big] \\ &\le 6^\ell(\ell + 1/2)\mathbb{E}_{\mathbb{P}}\big[|g_1(Y) - K_1|^\ell\big].\end{aligned}$$

We then integrate and apply the tail bound (6) to obtain the inequality

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\big[|g_1(Y) - K_1|^\ell\big] &= \int_0^\infty \ell t^{\ell-1}\mathbb{P}\{|g_1(Y) - K_1| \ge t\}\mathrm{d}t \le 4\int_0^\infty \ell t^{\ell-1}\exp\Big\{-\frac{ct^2k^2R}{n^2 + tkn}\Big\}\mathrm{d}t \\ &\le \int_0^{Cn/k} \ell t^{\ell-1}\exp\Big\{-\frac{ct^2k^2R}{n^2}\Big\}\mathrm{d}t + \int_0^\infty \ell t^{\ell-1}\exp\Big\{-\frac{ctkR}{n}\Big\}\mathrm{d}t \\ &= C_\ell\Big(\frac{n}{k\sqrt{R}}\Big)^\ell \cdot \int_0^{C\sqrt{R}} \ell u^{\ell-1}e^{-u^2}\mathrm{d}u + C_\ell\Big(\frac{n}{kR}\Big)^\ell \cdot \int_0^\infty \ell u^{\ell-1}e^{-u}\mathrm{d}u \\ &\le C_\ell\Big(\frac{n}{k\sqrt{R}}\Big)^\ell \cdot \{\Gamma(\ell/2) + \Gamma(\ell)\} \le C_\ell\Big(\frac{n}{k\sqrt{R}}\Big)^\ell \le C_\ell \cdot \max\Big\{n^{-\epsilon\ell/3}, n^{-\delta\ell}\Big\},\end{aligned}$$

where $\Gamma(\cdot)$ denotes the Gamma function, the penultimate inequality follows from the numeric inequality $\Gamma(\ell) \le 3\ell^\ell$ for all $\ell \ge 1/2$ and the final inequality follows from substituting $R = \min\{\lceil (n/k)^2 \cdot n^{2\epsilon/3}\rceil, n/2\}$.

Proof of the inequality (8b)**.** Applying the Cauchy–Schwarz inequality yields

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^2 \mathbb{1}\{|g_1(Y) - K_1| > 1/2\}\big] &\le \mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^4\big]^{1/2}\mathbb{P}\{|g_1(Y) - K_1| > 1/2\}^{1/2} \\ &\overset{\text{(i)}}{\le} \sqrt{2}\mathbb{E}_{\mathbb{P}}\big[\{\tau(g_1(Y)) - K_1\}^4\big]^{1/2} \max\Big\{e^{-cn^{2\epsilon/3}}, e^{-cn^{2\delta}}\Big\} \\ &\overset{\text{(ii)}}{\le} 4\big(\mathbb{E}_{\mathbb{P}}[\tau(g_1(Y))^4] + 1\big)^{1/2} \max\Big\{e^{-cn^{2\epsilon/3}}, e^{-cn^{2\delta}}\Big\} \\ &\le C_\ell n^{4\ell+2} \max\Big\{e^{-cn^{2\epsilon/3}}, e^{-cn^{2\delta}}\Big\},\end{aligned}$$

where step (i) follows from the inequality (6) and step (ii) follows from the numeric inequality $(a-b)^4 \le 8a^4 + 8b^4$. ∎