arXiv:2212.05619v2 [cs.DS] 6 Jun 2023

Algorithms approaching the threshold for semi-random planted clique*

Rares-Darius Buhai
rares.buhai@inf.ethz.ch
ETH Zürich

Pravesh K. Kothari†
praveshk@cs.cmu.edu
CMU

David Steurer
dsteurer@inf.ethz.ch
ETH Zürich

June 7, 2023

Abstract

We design new polynomial-time algorithms for recovering planted cliques in the semi-random graph model introduced by Feige and Kilian [FK01]. The previous best algorithms for this model succeed if the planted clique has size at least $=^{2/3}$ in a graph with = vertices [MMT20, CSV17]. Our algorithms work for planted-clique sizes approaching $=^{1/2}$ — the information-theoretic threshold in the semi-random model [Ste17] and a conjectured computational threshold even in the easier fully-random model. This result comes close to resolving open questions by Feige [Fei19] and Steinhardt [Ste17].

To generate a graph in the semi-random planted-clique model, we first 1) plant a clique of size : in an =-vertex Erdős–Rényi graph with edge probability 1/2 and then adversarially add or delete an arbitrary number edges not touching the planted clique and delete any subset of edges going out of the planted clique. For every > 0, we give an $=^{\$(1/)}$-time algorithm that recovers a clique of size : in this model whenever $: \geq =^{1/2+}$. In fact, our algorithm computes, with high probability, a list of about $=/:$ cliques of size : that contains the planted clique. Our algorithms also extend to arbitrary edge probabilities ? and improve on the previous best guarantee whenever $? \leq 1 - =^{-0.001}$.

Our algorithms rely on a new conceptual connection that translates certificates of upper bounds on biclique numbers in *unbalanced* bipartite Erdős–Rényi random graphs into algorithms for semi-random planted clique. Analogous to the (conjecturally) optimal algorithms for the fully-random model, the previous best guarantees for semi-random planted clique correspond to spectral relaxations of biclique numbers based on eigenvalues of adjacency matrices. We construct an SDP lower bound that shows that the $=^{2/3}$ threshold in prior works is an inherent limitation of these spectral relaxations. We go beyond this limitation by using higher-order sum-of-squares relaxations for biclique numbers.

We also provide some evidence that the information-computation trade-off of our current algorithms may be inherent by proving an average-case lower bound for unbalanced bicliques in the low-degree polynomial model.

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 815464).

†Supported by NSF CAREER Award #2047933, NSF #2211971, an Alfred P. Sloan Fellowship, and a Google Research Scholar Award.

Contents

1 Introduction **1**

Results 3

2 Techniques **6**

2.1 Efficient algorithms and biclique certificates 8

2.1.1 Basic spectral certificate 8

2.1.2 Improved spectral certificates 10

2.2 Our certificate: bicliques imply sets of negatively correlated vectors 11

2.3 From biclique certificates to algorithms for semi-random planted clique 12

3 Preliminaries **14**

3.1 Sum-of-squares preliminaries 14

4 Certifying biclique bounds in unbalanced random bipartite graphs **17**

4.1 The case of ? = 1/2 18

4.2 The case of arbitrary ? 23

5 List-decoding semi-random planted cliques **29**

5.1 Proof of main results 31

6 Evidence of hardness for certifying blicliques **35**

6.1 Lower bounds against basic SDP 36

6.2 Low-degree lower bound for ? = 1/2 38

6.3 Low-degree lower bound for general densities 41

References **44**

1 Introduction

Clique is one of the most intensely studied combinatorial problems in theoretical computer science, both in terms of its worst-case and its average-case complexity. It was among the first graph problems shown to be NP-complete [Kar72]. In fact, it turns out that for every > 0, it is NP-hard to find cliques of size $=$ even in graphs that contain cliques of size $=^{1-}$ [Hås99, Zuc07, Kho14].

The most well-studied average-case counterpart is the *planted clique* problem [Jer92, Kuc95] where the goal is to recover a $:$-clique added to an Erdős–Rényi random graph $(=, 1/2)$. Such a clique is uniquely identifiable if $: \gtrsim 2\log_2 =$. There are polynomial time algorithms based on rounding the second eigenvector of the adjacency matrix [AKS98] as well as basic semidefinite programming relaxations (e.g., the Lovász theta function) [FK00, FK03] to recover the planted clique with high probability whenever $: \geq =^{1/2}$. Closing the exponential gap between the information-theoretic threshold value of $:$ and the threshold of the best known algorithms is a tantalizing open question that has inspired a large body of research, culminating in lower bounds against restricted classes of algorithms, such as statistical query algorithms [FGR+17] and sum-of-squares relaxations [BHK+16], that vastly generalize the current algorithms for this problem. These concrete lower bounds provide some rigorous evidence that current algorithms for planted clique are optimal.

Fragility of algorithms. Unfortunately, many algorithms for the planted clique problem are *fragile*: a small number of adversarial changes to the input can cause the natural algorithms to break down completely. This includes methods based on basic statistics such as degrees of vertices or eigenvalues of the adjacency matrix that provide the strongest possible guarantees for the problem. Such fragility can be viewed as known algorithms *overfitting* to the choice of the distributional model.

In response, a significant research effort has gone into finding algorithms resilient against even the most benign forms of adversarial modifications. This includes a long line of work on *monotone adversary* models introduced in [FK98] for average-case formulations of clique and coloring (i.e., community detection) [MS16, MPW16, LM22]. In the context of planted clique, such models correspond to starting from the standard planted clique input and allowing an adversary to *delete* any subset of edges not in the planted clique. Such deletions are, in principle, only helpful since the planted clique continues to be the true maximum clique in the resulting graph. And indeed, while basic statistics and spectral methods fail in the presence of monotone adversaries, natural analyses of more resilient algorithms based on semidefinite programming [FK00] succeed at the same $: = \$(\sqrt{=})$ threshold while tolerating *monotone adversaries*.

Semi-random model. A seminal work by Feige and Kilian [FK01] introduced the following semi-random planted clique model following the classical work of [BS95] on semi-random coloring. Such semi-random models combine a distributional input with a monotone adversary and an adversarial choice at the same time. After the introduction of this model, similar semi-random models have been studied for a wide range of combinatorial optimization problems, including

graph partitioning and constraint satisfaction problems. We refer the interested reader to the excellent survey [Fei19].

Definition 1.1 (Feige–Kilian semi-random planted clique model, FK(=, :, ?))**.** For =, : □ □ with : ≤ = and ? □ [0, 1], we let FK(=, :, ?) be the collection of distributions over graphs with vertex set + = [=] sampleable by a process of the following form:

1. *Random Generation Phase*: Choose a uniformly random subset (□ □ + of size : and add a clique on (□ to an Erdős–Rényi random graph (=, ?) (which includes each possible edge independently at random with probability ?),
2. *Adversarial Deletion Phase*: delete an arbitrary subset of edges going out of (□ adaptively (i.e., possibly depending on the previous random choices),
3. *Adversarial Addition Phase*: replace the subgraph induced on + \ (□ by an arbitrary one, again adaptively.

Unlike planted clique with monotone adversaries, semi-random models are far from "helpful". In particular, the planted clique isn't necessarily the maximum clique in the resulting graph. And the adversarial choices in the generation process are known to result in significantly altered information-theoretic thresholds at which efficient algorithms can succeed for related problems such as community detection in the stochastic block model [MPW16].

If ? = 1, the above model recovers the worst-case version of the clique problem. On the other hand, by omitting the last two steps, we recover the original planted clique model, and by omitting only the last step we recover the planted clique with "helpful" monotone adversaries. Importantly, the last two steps are *adaptive* and can be chosen adversarially in response to the first step. In absence of adaptivity (i.e., when the last two steps are oblivious to the distributional choices), the model becomes significantly easier algorithmically.

We write □ FK(=, :, ?) to denote a graph sampled according to one of the distributions in FK(=, :, ?). For particular choices of parameters : = :(=) and ? = ?(=), our goal is to develop an algorithm that succeeds with high probability for *every* distribution described by FK(=, :, ?).

What does it mean for the algorithm to succeed? Since the graph induced on + \ (□ could be a worst-case hard instance for the clique problem, it is NP-hard to find a maximum clique in . So the goal of the algorithm is to find a clique of size : in . For the original planted clique model (and the version with helpful monotone adversaries), we could with high probability recover the planted clique (□ in . In contrast, in the semi-random model, this task is impossible information-theoretically because the adversary could simulate multiple disjoint copies of the distributional process in + \ (□. Instead, we can ask the algorithm to compute a small list of (pairwise almost-disjoint) :-cliques in that contains the planted clique (□. Such a list also allows uniquely identifying (□ if, in addition, we are given a random vertex of (□ as advice.

In their work introducing this model, [FK01] gave an algorithm that uses a Gaussian rounding [GW95] of the vector solution for the Lovász theta SDP relaxation combined with a combinatorial cleanup step to produce a correct list. For any ? such that 1 − ? ≥ (1+) ln(=)/=, their algorithm

works if $k \geq \ell$ for some constant $\ell > 0$. Such a guarantee is essentially optimal if $1 - \eta = \tilde{\Omega}(1/n)$. The main focus of subsequent works has been in the case when $1 - \eta$ is larger. In particular, the case of $\eta = 1/2$ (and more generally, any constant < 1) is of special interest. In this case, one can ideally expect polynomial time algorithms that succeed for $k \geq \sqrt{n}$ as in the case of average-case planted clique. We will focus on the case of $\eta = 1/2$ in this introduction for the sake of clarity.

Prior work. Algorithms in prior works rely on rounding carefully designed semidefinite programming (SDP) relaxations. In the slightly easier setting that drops the adversarial deletion step from the model, Charikar, Steinhardt and Valiant [CSV17] gave an algorithm based on a semidefinite programming relaxation for *list-decodable mean estimation* that succeeds whenever $k \geq \Omega(n^{2/3} \log^{1/3}(n))$. Their guarantee was improved to $k \geq \Omega(n^{2/3})$ by Mehta, Mackenzie and Trevisan [MMT20]. The algorithm of [MMT20] is based on a variant of the Lovász theta SDP (that they call "crude" or C-SDP) with an objective function that incentivizes "spread-out" vector solutions and analyzed via the Grothendieck inequality. They suggest (though don't prove) that their SDP should fail if $k = o(n^{2/3})$. Further heightening the intrigue, Steinhardt [Ste17] proved that if $k = o(\sqrt{n})$, then it is information-theoretically impossible to identify an $\Omega(n/k)$-size list, indicating an *information-theoretic* (as opposed to computational) phase transition at $k \approx \sqrt{n}$.

Feige's open question. Given the apparent barrier for the basic semidefinite program at $k \approx n^{2/3}$, it is natural to ask: is the semi-random variant harder than the average-case planted clique problem or could there be algorithms that succeed for k approaching the $\Omega(\sqrt{n})$ threshold? In his survey on semi-random models [Fei19], Feige posed (see Section 9.3.4, Page 205) this as an outstanding open question and hoped for algorithms for semi-random planted clique matching the $k \approx \sqrt{n}$ threshold for the average-case variant.

Results

In this work, we nearly resolve Feige's question and give an algorithm for the semi-random planted clique problem that works for k approaching $\sqrt{n}$. Specifically, we give a scheme of algorithms that, for any $\ell > 0$, run in time $n^{\Omega(1/\ell)}$ and succeed in solving the semi-random planted clique problem whenever $k \geq n^{1/2+\ell}$:

Theorem 1.2 (Main result, see Theorem 5.6 for a detailed version)**.** *For every $\ell > 0$, there is an algorithm that, given a graph G on n vertices as input, computes a list of vertex subsets in time $n^{\Omega(1/\ell)}$ satisfying the following guarantee: If G is generated according to $\mathrm{FK}(n, k, 1/2)$ for $k \geq n^{1/2+\ell}$, then with probability at least 0.99 the algorithm outputs a list of at most $(1 + o(1))\frac{n}{k}$ cliques of size k such that one of them is the clique planted in G.*

In particular, our algorithm manages to recover planted cliques of size k approaching $n^{1/2}$ — the information-theoretic threshold [Ste17] and the conjectured computational threshold even for the easier fully-random planted clique problem. This improves on the best known prior algorithm [MMT20] that gives a polynomial time algorithm that succeeds whenever $k \geq \Omega(n^{2/3})$.

Our approach extends to edge probabilities ? beyond the choice ? = 1/2 and yields improved guarantees even when ? = 1 − >(1), though in that case we do not approach the information-theoretic threshold. Our hardness results (discussed below) show that such an outcome might be inevitable.

Higher degree sum-of-squares vs basic SDP. Our algorithm relies on rounding a high constant degree sum-of-squares relaxation (that maximizes a natural "entropy-like" objective function) of the natural integer program for finding :-cliques in graphs. As we discuss below, this is likely necessary as for the natural certification problem (discussed below) that arises in algorithms for semirandom planted clique, the basic SDP (Lemma 6.1) has a lower bound that precludes recovering : = $=^{2/3}$-size cliques. This is in sharp contrast to the average-case planted clique problem where no constant degree strengthening of the basic SDP allows recovering planted cliques of size : = >($\sqrt{=}$) [BHK+16] (i.e., asymptotically smaller than the threshold for recovery using the basic SDP). In fact, the best known analyses can only obtain an $=^{\$(C)}$ algorithm that succeeds whenever : ≥ $\$(\sqrt{=/2^C})$ [FK03].

Indeed, few combinatorial optimization problems are known to benefit from high constant degree sum-of-squares relaxations. Some notable exceptions include approximating constraint satisfaction problems on graphs with small threshold rank [ABS15, BRS11] (where the high degree corresponds to the threshold rank) and approximating the maximum bisection in a graph [RT12] (where the high degree helps deal with the cardinality constraint). Our work adds a new example to this list that appears to be more unstructured than earlier examples.

We note that, in contrast to combinatorial optimization, in statistical estimation higher degree sum-of-squares relaxations have recently been pivotal in algorithmic applications such as robust method of moments [KSS18], linear regression [KKM18, BP21], list-decodable learning [KSS18, KKK19, RY20a, RY20b, BK21, IK22], and settling the robust clusterability and learnability of high-dimensional Gaussian mixtures [KSS18, HL18, BK20, BDH+20, BDJ+20, LM21].

Rounding and connection to certifying bicliques. Our rounding algorithm is reminiscent of the "rounding by votes" strategy employed in several recent works on list-decodable learning [KKK19, BK21, IK22]. Our analysis relies on a new connection to efficient certificates: we can recover a small list which includes the planted clique if we can certify that the planted clique has small intersection with all other :-cliques. This reduces to certifying upper bounds on bipartite cliques in unbalanced bipartite *random* graphs: see Definition 2.4 for a standalone definition of the problem.

Given a bipartite graph = (*, +,) with |*| = :, |+| = =, and each bipartite edge included in with probability ? independently, it is easy to prove by a standard application of Chernoff and union bounds that there is no ℓ by : bipartite clique in with ℓ $\$(\log =/(1-?))$. The *bipartite clique certification* problem asks to find a polynomial time verifiable certificate that contains no ℓ by : biclique for ℓ as small as possible. This is a variant of the more standard biclique certificate problem (see, e.g., [FGR+13]) where both the graph and the cliques we are interested in are unbalanced.

Our main result is based on the following primitive that certifies in time $=^{\$(1/)}$ that a random

: by = bipartite graph does not contain ℓ by : blicliques for ℓ = = and any : ≥ $\tilde{\$}(\sqrt{=})$. Our certificates are based on $(1/)-degree sum-of-squares proofs and this is necessary – we prove that for any ℓ = >(=/:), there are no degree 2 sum-of-squares (i.e., basic SDP) certificates of absence of bicliques. In particular, unlike the case of balanced bipartite random graphs, the unbalanced setting seems to naturally benefit from large constant degree sum-of-squares certificates.

Theorem 1.3 (Informal, see Theorem 4.2 and Lemma 6.1)**.** *For every* > 0*, there is an* $=^{\$(1/)}$ *time algorithm that takes input a bipartite graph* = (*, +,) *with* |*| = :, |+| = = *and each bipartite edge included in with probability* ? = 1/2*, and with probability at least* 0.99 *over the draw of outputs an* $=^{\$(1/)}$*-time verifiable certificate that contains no* ℓ *by* : *biclique for* ℓ ≤ = *whenever* : ≥ $\tilde{\$}(\sqrt{=})$.

Further, 1) the certificate can be expressed as an $(1/) *degree sum-of-squares refutation of the biclique axioms (see* (3)*) and 2) there does* not *exist a degree* 2 *certificate (equivalently, based on the "basic SDP") to certify a bound of* ℓ = >(=/:).

The results of prior works can be obtained by a simple spectral certificate captured by the basic SDP relaxation for upper bounding such bicliques (see detailed discussion in Section 2.1.1 of the techniques). In contrast, in this work, we depart from the spectral certificates and rely on a certain simple *geometric* certificate based on upper bounds on the size of sets of *pairwise negatively correlated vectors*. If = $(log log =/log =) is chosen such that ℓ = $(log =) and : = $(= $\sqrt{\log =}$), the biclique refutation above translates into an algorithm for semi-random planted clique that works for : ≥ $(= $\sqrt{\log =}$) in time $=^{\$(\log =/\log\log =)}$, matching Steinhardt's information-theoretic lower bound [Ste17] and the threshold for the best known efficient algorithms for planted clique up to a $\sqrt{\log =}$ factor.

Hardness of refuting bicliques. We provide some evidence that improving on our biclique certification algorithms likely requires new techniques by proving a lower bound in the low-degree polynomial model. The low-degree polynomial model (see [KWB19] for a great exposition) is a restricted model for statistical *distinguishing* problems. More precisely, the model considers problems where we are given a single sample (an instance of an algorithmic problem, a graph in our case) with the promise that it is an independent sample from one of two possible distributions: $_{null}$ — a distribution that does not admit solutions, usually a natural random model, and $_{planted}$ — a closely related distribution that does admit solutions. Informally speaking, the low-degree model restricts distinguishers to thresholds of low-degree polynomials of the input. While low-degree polynomials might appear restricted, they capture several algorithms including power iteration, approximate message passing, and local algorithms on graphs (cf. [DMM09, GJW20]). Moreover, it turns out that they are enough to capture the best known spectral algorithms for several canonical problems such as planted clique, community detection, and sparse/tensor principal component analysis [BHK+19, HS17, DKWB19, HKP+17].

This model arose naturally from work on constructing sum-of-squares lower bounds for the planted clique problem [BHK+19]. It was formalized in [HKP+17] and conjectured to imply sum-of-squares lower bounds for certain average-case refutation problems. Subsequently, starting with [HS17] (see also [Hop18]), researchers have used the low-degree polynomial method as a technique to demarcate average-case algorithmic thresholds [HKP+17, GJW20, SW20, Wei20].

In our case, $_{\text{null}}$ will be the (:, =, ?) model: bipartite graphs with left vertex set of size :, right vertex set of size =, and each bipartite edge present in the graph with probability ?. Notice that if we had an algorithm that certifies the absence of ℓ by : − ℓ bicliques in such a graph then we can distinguish between $_{\text{null}}$ and any $_{\text{planted}}$ supported on bipartite graphs that admit ℓ by : − ℓ bicliques. Thus the distinguishing problem is formally easier than the task of certification (also known as refutation). Despite the restrictedness of the low-degree model, we observe that for average-case planted clique with : ≈ $\sqrt{=}$, *constant* degree polynomials suffice to distinguish between $_{\text{null}}$ = (=, 1/2) and $_{\text{planted}}$ = (=, 1/2) + :-clique.

Theorem 1.4 (Low-degree polynomial heuristic for biclique certification problem)**.** *Fix* > 0 *small enough and* = *large enough. Let* $_{\text{null}}$ = (:, =, 1/2) *be the distribution on* (:, =)*-bipartite graphs where every edge is included independently with probability* 1/2*. For* : = $=^{1/2+}$*, there is a distribution* $_{\text{planted}}$ *on* (:, =)*-bipartite graphs containing an* ℓ *by* : − ℓ *bipartite clique for* ℓ = $=^{0.1}$ *such that the norm of the degree-*Ω(1/) *truncated likelihood ratio between* $_{\text{planted}}$ *and* $_{\text{null}}$ *is* 1 + >(1)*.*

Informally speaking, the above theorem asserts that, for : = $=^{1/2+}$, statistical tests based on computing thresholds of Ω(1/)-degree polynomials fail to distinguish between $_{\text{null}}$ that does not admit $(log =) by : bicliques and $_{\text{planted}}$ that contains an $=^{0.1}$ by : biclique. It turns out that the most natural planted model (plant a random ℓ by : − ℓ clique and sample the rest of the graph independently) can be distinguished from $_{\text{null}}$ using just degree 1 polynomials and thus does not suffice to prove the above theorem. Instead, we use an edge-adjusted model where the probability of sampling edges outside the biclique is reduced in order to make the degree distribution of left vertices match that of (:, =, ?).

For ? = 1/2, the above theorem shows that we need polynomials of degree $(1/) in order to distinguish between $_{\text{null}}$ and bipartite graphs with $=^{0.1}$ by $=^{1/2+}$ bicliques. Given the contrast to the planted clique problem where the corresponding distinguishing problem can be solved by constant degree polynomials, we obtain some (weak) evidence that beating the guarantees of our current certificates may require new techniques for ? = 1/2. For general ?, a similar lower bound suggests that the degree of the polynomial required to distinguish between $_{\text{null}}$ and $_{\text{planted}}$ is larger than any function (independent of =) of 1/, or else that : needs to scale with 1/(1 − ?) instead of the information-theoretical optimal scaling of 1/ $\sqrt{1-?}$ — note that this discrepancy is poly(=) when 1 − ? = 1/poly(=).

2 Techniques

In this section, we provide a high-level overview of our algorithm for the semi-random planted clique problem. For simplicity of exposition, we will focus on the important case of ? = 1/2.

Given a graph generated according to FK(=, :, 1/2), our goal is to construct a small *list* of candidate :-cliques in such that the true planted clique (≈ is contained in the list (we will call such lists *correct*). Our construction will also ensure that a constant fraction of the vertices in (≈ do not appear in any other clique in the list. As a result, we can also uniquely recover (≈ with high probability when given, in addition, a uniformly random vertex in (≈.

Our algorithm and its analysis rely on the *proofs-to-algorithms* method (see [FKP19, BS16] for more on the usage of this method).

Inefficient algorithm. Let's first find an algorithm, even if inefficient, to generate a poly(=) size correct list, i.e., one that contains (. Notice that simply outputting all :-cliques in can lead to an exponentially large (i.e., =:) size list since we have no control over the subgraph induced on [=] \ ((e.g., consider a clique on [=] \ (). Instead, we will enumerate all :-cliques in that satisfy an additional property such that 1) the property is satisfied by the planted :-clique on (with high probability, and 2) every graph has at most (1 + >(1))=/: :-cliques satisfying the property. This property is quite natural and asks for the bipartite graph with the :-clique on the left and the rest of the vertices on the right to not contain a large *unbalanced* biclique with many vertices on the left side. Recall that an ℓ by Abiclique in a bipartite graph is a set of vertices that consists of ℓ left vertices and A right vertices such that contains all possible bipartite edges between the two sides.

Definition 2.1 (Good :-cliques)**.** Let be a graph on = vertices. We say that a :-clique (in is ℓ*-good* if every biclique (!, ') in the bipartite graph with left vertex set (, right vertex set [=] \ (, and edge set cut(() satisfies |!| ≤ ℓ whenever |'| ≥ 1 and |!| + |'| = :.

The planted :-clique on (is $(log =)-good with high probability over the draw of cut((().

Proposition 2.2 (Bipartite clique number of cut((()). *Let* :, = and FK(=, :, 1/2)*. Then, for large enough* = *and a constant* 2 > 0*, with probability at least* 0.99 *over the draw of edges in* cut((), *for any* ! (, ' [=] \ (*such that* (!, ') *is a biclique in* cut(() *satisfying* |'| ≥ 1 *and* |!| + |'| = :*, we have* |!| ≤ $2\log_2 =$.

Proof. The proof is a simple application of the first moment method. Note that it is enough to argue the proposition in the absence of the monotone adversary as deleting any subset of edges in cut(() maintains the goodness of (.

The probability that cut(() contains all the edges between ! (and ' [=] \ (is at most $2^{-(:-|!|)|!|}$. Thus, the expected number of bicliques (!, ') such that |!| ≥ $2\log_2 =$ is at most $\sum_{\ell+A=:,\ell\geq 2\log_2 =} \binom{=-:}{:-\ell}\binom{:}{\ell} 2^{-(:-\ell)\ell} \to 0$ as $= \to \infty$ if 2 is a large enough constant. The proposition then follows by an application of Markov's inequality.

A simple greedy argument upper bounds the number of ℓ-good :-cliques if : ≥ $\$(\sqrt{= \log =})$.

Proposition 2.3 (Number of good :-cliques)**.** *Let be a graph on = vertices. Then, for any* ℓ*, if* : > 2 $\sqrt{=\ell}$/ *for some* < 1*, then the number of* ℓ*-good* :*-cliques in is at most* (1 +)=/:.

Proof. Suppose not and take any < = (1 +)=/: such good :-cliques. Observe that any pair of ℓ-good :-cliques (, (′ can only intersect in at most ℓ vertices, as otherwise cut(() would contain a biclique with more than ℓ left vertices. Thus, the < good :-cliques must cover at least $<: - <^2\ell = = + = - (4=^2/:^2)\ell$ vertices, a number that exceeds the total number of vertices = if : > 2 $\sqrt{=\ell}$/.

Propositions 2.2 and 2.3 immediately yield an $=^{\$(:)}$ time algorithm to generate a correct list of :-cliques of size (1 +)(=/:). In fact, this algorithm can be made to run in time $=^{\$(\log =)}$ by enumerating all $2\log_2 =$ size cliques & in and adding a :-clique to the list if the common neighborhood of & is of size ≥ : − |&| and forms a clique with &.

2.1 Efficient algorithms and biclique certificates

In the inefficient algorithm above a key idea is the claim that cut((□) does not have an ℓ by : − ℓ bipartite clique for ℓ > $(log =). Note that cut((□) is an unbalanced (left side is much smaller than the right) : by = − : ≈ = bipartite graph and we proved that it does not have an unbalanced (□ $(log =) vertices from the left) biclique in it.

Key to our efficient algorithm for semi-random planted clique is an efficiently computable *certificate* of non-existence of unbalanced bicliques in as above (i.e., a *refutation*).

Let ($_1, $_2, ?) denote the distribution on bipartite graphs with $_1 left and $_2 right vertices and every bipartite edge included with probability ? independently. Let us phrase the version relevant to us formally before continuing:

Definition 2.4 (Refuting unbalanced bicliques)**.** An algorithm that takes as input a bipartite graph = (*, +,) with |*| = :, |+| = = − : refutes ℓ by : − ℓ bicliques in random : by = − : bipartite graphs if it has the following two properties:

1. **Correctness:** If the algorithm outputs B, then there is no B by : − B biclique in .
2. **Utility:** If □(:, = − :, 1/2), then the algorithm outputs B ≤ ℓ with probability at least 0.99 over the draw of .

Remark 2.5 (From certificates to algorithms: a heuristic). In Section 2.3, we overview the translation of a (constant degree sum-of-squares) certificate that the left side of any size-: biclique in □(:, = − :, 1/2) has at most ℓ vertices into an algorithm for the semi-random planted :-clique problem that succeeds whenever : ≥ $(√(=ℓ)). This matches the simple bound in Proposition 2.3 for the "brute-force" algorithm above. We postpone the discussion of sum-of-squares proofs for now while noting that all certificates discussed in this section are in fact constant degree sum-of-squares certificates.

Observe that our simple analysis of the inefficient algorithm gives an $=^{\$(\log =)}$ algorithm that refutes the existence of ℓ by : − ℓ bicliques in (:, = − :, 1/2) with probability at least 0.99 for ℓ = $(log =). Our goal is to find a polynomial time algorithm that succeeds for ℓ as close to $(log =) as possible.

The biclique refutation problem appears to be an interesting analog of refuting cliques in random (non-bipartite) graphs □(=, 1/2) (that underlies algorithms for the fully-random planted clique problem) or bicliques in (=, =, 1/2) (i.e., the balanced bipartite graph). It can be thought of as *certifying* the correctness of the candidates in the list that is purportedly a solution to the semi-random planted clique problem. Finding solutions together with a certificate of correctness is an important goal by itself. For example, this is a key advantage (in addition to tolerating a monotone adversary) of the method of Feige and Krauthgamer [FK00] over the spectral algorithm [AKS98] for the planted clique problem.

2.1.1 Basic spectral certificate

Let us start by recalling the basic spectral certificate that underlies the algorithms for the average-case planted clique problem. This certificate implicitly underlies the algorithms of [MMT20,

CSV17]. Our framework translates it into an algorithm for semi-random planted clique whenever : ⍰ $(=$^{2/3}$).

Proposition 2.6 (Basic spectral certificate for clique number). *In any graph , the clique number* $() ≤ 1 + kk$_2$ *where is the* {±1} *adjacency matrix of .*

Proof. If G is a {0, 1}-indicator of a :-clique in , then, note that :(: − 1) = G⍰G ≤ kGk$_2$ kk$_2^2$ = : kk$_2$. Thus, : ≤ 1 + kk$_2$ for any graph .

Thus, simply outputting the (polynomial time computable) largest singular value of gives a certificate of an upper bound on $(). Further, if ⍰ (=, 1/2), then standard spectral norm bounds on random symmetric {±1} matrices imply that the algorithm outputs with high probabil-ity a bound of $(√ =).

Let's now see an analog of this method for bicliques.

Proposition 2.7 (Basic spectral certificate for bicliques, see Lemma 4.13 for a general version). *Let be the* {±1} *adjacency matrix of a : by = − : bipartite graph . For any :-clique in , the number of left vertices* ℓ *satisfies* ℓ(: − ℓ) ≤ kk$_2$.

Proof. Let G, H be the {0, 1} indicators of the left and right sides of a biclique in . Then, kGk$_2^2$ kHk$_2^2$ = G⍰H ≤ kGk$_2$ kHk$_2$ kk$_2$. Or, (8G8)(8H8) = kGk^2 kHk^2 ≤ kk^2.

For a random bipartite graph from (:, = − :, 1/2), the is a : by = − : matrix with independent random {±1} entries. For such matrices, standard results (see Fact 3.14) show that kk$_2$ ≤ $(√: + √=) = $(√=). Further, by a union bound, the degrees of all right vertices are at most :/2 + $(p: log =) with high probability, so : − ℓ ≥ :/4 if : ⍰ log =. In that case, the above proposition shows that the spectral certificate refutes the existence of an ℓ by : − ℓ clique for ℓ ≤ $(=/:).

By applying the heuristic from Remark 2.5, we obtain an algorithm for semi-random planted clique if : ≥ $(√=ℓ) with ℓ = $(=/:), that is, if : ≥ $(=$^{2/3}$), matching the guarantees of [MMT20].

It turns out that the bound of ℓ = $(=/:) based on the basic SDP/spectral relaxations is essentially tight. In Lemma 6.1, we show that the basic SDP provably fails to certify that ℓ = >(=/:). This shows an inherent limitation of certificates based on the basic SDP/spectral relaxations.

The Charikar-Steinhardt-Valiant approach. In their work on algorithms for list-decodable mean estimation [CSV17], the authors devised a method for the analog of the semi-random planted clique problem without the monotone adversary step. When viewed from our vantage point of biclique refutation, their idea can be thought of as taking the ±1-neighborhood indicators of the *right* hand side of the graph and treating them as = − : samples of a :-dimensional distribution. An ℓ by : −ℓ biclique translates 1 into the distribution having a non-zero mean. Thus, one can apply (analogs of)

1We note that the CSV approach directly applies to the semi-random planted clique model and does not actually yield a biclique certificate. The reason is that an ℓ by : − ℓ biclique does not translate into non-zero mean for arbitrary bipartite graphs. We ignore this distinction in order to allow an intuitive comparison of their technique in the context of our work.

list-decodable mean estimation algorithms [CSV17, KS17] to refute the existence of bicliques. The guarantees of the algorithm depend on higher directional moments of the input distribution. The "base case" corresponds to using just the second moments of the distribution — and this roughly relates to the use of the basic spectral certificate above. The higher moment variants can indeed yield improvements but this does not apply to our setting, because when seen from the vantage point of list-decodable mean estimation we have = □ :^{2} samples of a :-dimensional distribution — a bound not sufficient for the 4th moments to converge! Indeed, this is the key bottleneck that leads to a barrier at : = $\tilde{\Omega}(=^{2/3})$ for the CSV approach (and led to Steinhardt's open question for semi-random planted clique [Ste17]).

2.1.2 Improved spectral certificates

Can we improve on the basic spectral certificate? We note that for related problems (e.g., densest :-subgraph, random constraint satisfaction, coloring random graphs) we usually get no asymptotic improvement by considering spectral certificates with larger (but polynomial size) matrices built from the instance. Indeed, one can prove strong lower bounds [KMOW17, JPR+22] that rule out such larger polynomial size certificates captured by constant degree sum-of-squares proofs.

Neighborhood reduction. A natural way to improve the spectral certificate for the clique number of □ (=, 1/2) from Proposition 2.6 is to cycle through all possible subsets of C vertices, move to the common neighborhood of the C vertices and then apply Proposition 2.6 to the induced graph on this common neighborhood. This strategy yields an upper bound of $\Omega() \le C + 1 + \max_{(\subseteq[=],|(|=C} k_(k_2$ where $_($ is the adjacency matrix of the induced subgraph on the common neighborhood of (. One can prove that $k_(k_2 \le \tilde{\Omega}(\sqrt{=/2^C})$ with high probability simultaneously for all (of size C, certifying an upper bound of $\tilde{\Omega}(\sqrt{=/2^C})$ on the clique number $\Omega()$. Since the resulting certificate has polynomial size only when C = $\Omega(1)$, the improvement makes no asymptotic difference in the threshold : at which polynomial time algorithms work. As an aside, this simple certificate happens to be optimal for the degree C Lovász-Schrijver SDP hierarchy [FK03] applied to □ (=, 1/2). Repeating an analogous argument in our case also yields no asymptotic improvement unless C = $\Omega(1)$ (though it does allow us to get arbitrary constant factor improvements).

Tensoring. We consider next a natural class of "tensoring" schemes for producing improved spectral certificates. Consider a bipartite graph with {±1} adjacency matrix $'$ with the same right side but the left side containing all pairs of left vertices from . The ((8, 9), :)-th entry of $'$ equals $_(8,:)_(9,:)$ – the "parity" or product of the {±1} indicators of edges (8, :) and (9, :) in . $'$ is a :^{2} by = matrix, and further, an ℓ by : − ℓ biclique in translates into an ℓ^2 by : − ℓ biclique in $'$.
The basic spectral certificate from Proposition 4.13 applied to $'$ yields that ℓ$^2 \le \tilde{\Omega}(k' k_2 /:)^2$
If $'$ were a matrix of independent random {±1} entries, $k' k_2 = \tilde{\Omega}(\sqrt{:^2 =}) = \tilde{\Omega}(:\sqrt{=})$ yielding ℓ ≤ $\tilde{\Omega}(\sqrt{=}:)$. Despite $'$ having correlations in its entries, this optimistic[2] bound is essentially correct (we will omit the proof here). Plugging this back into our heuristic, we get an algorithm

[2] Every rectangular matrix of larger dimension :^{2} and Frobenius norm :$\sqrt{=}$ has a spectral norm ≥ :.

even for semi-random planted :-clique if : ≥ $(= :) or : = =2/3, the same as before! That is, though the tensoring trick gives a different asymptotic estimate, it does not lead to any improvement in the threshold for : in our semi-random planted clique application.

What happens if we "tensor the left side" Ctimes for C> 2? An optimistic estimate such as the above yields a bound of ℓC ≤ $(:C−1) or ℓ ≤ :1−1/C – a bound that appears to *degrade* as we increase C! We will omit the details here but a similarly worse bound results if we tensor the right side of instead.

Two-sided tensoring beats the =2/3 barrier but fails a long way off . It turns out simultaneously tensoring both sides unequally helps beat the ℓ ≤ max{ :, =/:} bound obtained via one-sided tensoring above. Intuitively speaking, the "optimal" two-sided tensoring attempts to make the resulting adjacency matrix as "square" in dimensions as possible. Formal proofs require analyzing matrices of correlated random entries using the graph matrix method devised in the context of proving sum-of-squares lower bounds in [BHK+16] and follow-ups. We note without further details that two-sided tensoring appears to break down at : = 0.61.

2.2 Our certificate: bicliques imply sets of negatively correlated vectors

Our key idea to circumvent the bottlenecks in the natural spectral certificates is to abandon the idea of spectral certificates altogether. Instead, we will show that a simple family of "geometric" certificates for biclique numbers allows us to show ℓ ≤ = for any fixed > 0. Specifically, we will show that if there is an ℓ by : − ℓ biclique in , then one can extract 2ℓ − 1 *pairwise negatively correlated vectors* in = dimensions.

In order to explain this connection, let us note a property of a random bipartite graph = (*, +,) ∼ (:, = − :, 1/2). For any subset (⊆ * of |(| ≤ Cvertices from the left vertex set of , let #((9) = ∏8∈((8, 9) where is the {±1}-adjacency matrix of . Then #(is an = dimensional vector of "parities" of {±1} indicators of all edges from (to {:}. Further, in a random , every #(is nearly *balanced*. That is, by a simple Chernoff and union bound argument (see Lemma 4.5), |∑8≤=−: #((8)| ≤ $(=Clog =) for every (of size C.

Let's call a : by = − : bipartite graph C-fold balanced if the above property holds: that is, every #(is approximately balanced for |(| ≤ C. We will now show that given an ℓ by : − ℓ biclique in a C-fold balanced graph, we can produce a set of $(C^{ℓ/2}) pairwise negatively correlated vectors in = dimensions.

Proposition 2.8 (Bicliques and negatively correlated vectors)**.** *Suppose is a : by = − : bipartite graph that is C-fold balanced for some C ≥ 2. Suppose that contains an ℓ by : − ℓ biclique* (!, ') *for : − ℓ ≥ :/4. Then, if : ≥ $(√(=Clog =)), there exist* $(C^{ℓ/2}) *different* (= − 2: + ℓ)*-dimensional vectors #*¯(*(one for each* (⊆ ! *of size* C/2*) such that* h#¯(, #¯)i < 0 *whenever* (≠) *.*

Proof. First observe that for any (,) ⊆ ! of size C/2, h#(, #)i = ∑9≤=−: #(Δ)(9) = $(√(=C log =)) where we invoked the C-fold balancedness of . Now, without loss of generality, assume that ' is the set of the first : − ℓ vertices on the right. Consider the vectors #¯ in = − 2: + ℓ dimensions

obtained by stripping the first $: - \ell$ coordinates off of $\#_{(}$ for every $(\in !$ of size $C/2$. Since $(,) \in !$, the first $: - \ell$ coordinates contribute $+(: - \ell)$ to $h\#_{(}, \#_{)}i$. Thus, $h\#_{(}, \#_{)}i \leq \$(=C\log =) - :/4 < 0$ if $: - \ell \geq :/4$ and $: \geq \$(=C\log =)$.

It is a standard fact that there can only be $3 + 1$ pairwise negatively correlated vectors in 3 dimensions. A weaker version can be proved via a simple argument involving quadratic polynomials over the vectors:

Proposition 2.9 (Bound on negatively correlated vectors)**.** *Let* $E_1, E_2, \ldots, E_\#$ *be* $=$*-dimensional vectors of length* $\sqrt{=}$ *each satisfying* $hE_8, E_9i \leq -A$*. Then* $\# \leq 1 + =/A$*.*

Proof. We know that $\sum_{8 \leq \#} E_8^2 \geq 0$. On the other hand, $\sum_{8=1}^{\#} E_8^2 = \sum_{8=1}^{=} kE_8k_2 + \sum_{8 \neq 9} hE_8, E_9i \leq \# = - \#(\# - 1)A$. Putting the lower and upper bound together yields that $\# - 1 \leq =/A$ or $\# \leq 1 + =/A$.

Now, Proposition 2.8 yields $\ell^{C/2}$ vectors with pairwise correlations at most $-2:$ for some constant $2 > 0$ if $: \geq p =C\log =$. On the other hand, Proposition 2.9 yields that the number of such vectors can only be $1 + \$(=/:)$. Putting these two bounds together yields that $\ell \cdot (=/:)^{2/C}$. Choosing $C = 1/$ gives us an $=^{\$(1/)}$ size certificate that ℓ is at most $=$.

The above argument can be converted into a sum-of-squares refutation of bicliques in (see Theorem 4.2). The main observation is that the step where we strip the first $: - \ell$ coordinates off of $\#_{(}$ can be done "within sum-of-squares" while the remaining argument is a sum-of-squares proof by virtue of the above simple proposition. It turns out that we need some additional careful arguments to place the certificate in a usable form, which we will omit for the purpose of this overview (see Remark 4.7).

2.3 From biclique certificates to algorithms for semi-random planted clique

Our algorithms use the biclique certificates discussed previously to analyze a rounding algorithm for SDP relaxations of the standard $:$-clique axioms. Specifically, consider the standard integer programming formulation of the $:$-clique problem written as the quadratic polynomial system $A = A()$ below. Note that the solutions to $A()$ are $:$-cliques in the graph on vertex set $[=]$.

$$A(): \quad \begin{array}{ll} \forall 8 \in [=] & F_8^2 = F_8 \\ \forall 8 \in [=] & \sum_{8=1}^{=} F_8 = : \\ \forall 8, 9 \text{ s.t. } \{8, 9\} \notin & F_8 F_9 = 0 \end{array} \tag{1}$$

Finding a solution to this quadratic program is clearly NP-hard. So we will instead work with "sum-of-squares" SDP relaxations of the quadratic program, whose solutions can be interpreted as a generalization of probability distributions over solutions to the quadratic program. Specifically, a degree 3 pseudo-distribution is a relaxation of a probability distribution on $\{0, 1\}^=$ in that the associated "mass" function can take negative values while still inheriting a non-trivial subset of the properties of probability distributions. We will postpone the formal definition of pseudo-distributions to Section 3 and for now note the following relevant bits: 1) Unlike an

actual probability distribution, we only get access to low-degree moments (i.e., expectations of monomials) of and thus can only compute expectations of degree ≤ 3 polynomials, 2) pseudo-distributions can assign "negative probabilities" and thus may not assign non-negative expectations to pointwise non-negative degree 3 polynomials 5, but 3) degree 3 pseudo-distributions do assign non-negative expectations to any 5 that is a sum of squares of degree ≤ 3/2 polynomials, and 4) a pseudo-distribution of degree 3 satisfying A satisfies all "low-degree inferrable" properties of :-cliques but need not be supported on F that indicate :-cliques at all. Here, low-degree inferrable property means that for any degree ≤ 3 − 2 polynomial 5 and any {8, 9} ⍰ , ..[5̃F₈F₉] = 0.

A degree 3 pseudo-distribution minimizing any convex objective in the pseudomoments ..̃[Î ₈⍰(F₈] for |(| ≤ 3 and approximately satisfying A at degree 3 can be computed in time =$(3) (see Section 3).

Though a pseudo-distribution is not a probability distribution over solutions to A, it is still helpful for the reader to imagine it to be as such.

How do our biclique certificates help us? It turns out that while degree 3 pseudo-distributions are far from actual probability distributions for 3 ⍰ =, they behave so for the purpose of polynomial inequalities that can be derived from A using degree 3 sum-of-squares proofs. The conclusion of our biclique certificate from Proposition 2.2 can be written (see Theorem 4.2) as a degree $(C) consequence of the quadratic system ⍰ (see (3)) that identifies bicliques in bipartite graphs of total size :. Consider the bipartite graph cut((⍰). Let F₁ be the restriction of F to coordinates in (⍰ and F' be the restriction of F to coordinates outside of (⍰. Then, A implies that (F₁, F') satisfy ⍰ for the bipartite graph cut((⍰). Since the pseudo-distribution satisfies A, we can conclude that

$$..\tilde{:}\left[\left(\sum_{8 \in (^{⍰}} F_8 \right)^{C} \left(\sum_{8 \in (^{⍰}} F_8 \right)^{!} \right] \leq \$(=^5/:^4) \qquad (2)$$

whenever the pseudo-distribution has degree at least $(C). Note that (⍰ is not known to us but the above inequality forces the pseudo-distribution computed by the SDP to capture some non-trivial information about it.

The need for coverage constraints. Roughly speaking, (2) can be interpreted as saying that the pseudo-distribution is "supported" only on those F that cannot simultaneously appreciably intersect (⍰ and [=] \ (⍰. Such a fact by itself seems unhelpful. After all, the pseudo-distribution could completely ignore (⍰ and focus on the "worst-case" graph on [=] \ (⍰. Given the worst-case hardness of clique, the pseudo-distribution may not have any information about :-cliques in [=] \ (⍰ and consequently the input graph.

In order to make (2) useful, we must somehow "force" the pseudo-distribution to have a non-trivial mass on vertices in (⍰. Of course, we do not know (⍰, so how can we do it? It turns out that this can be accomplished by certain "max coverage" constraints. Specifically, instead of finding any pseudo-distribution consistent with A, we find one that minimizes ..̃[F]²₂. This is a convex function of the pseudo-distribution and thus can be minimized efficiently using the ellipsoid method. This objective forces the pseudo-distribution to be "spread-out". Indeed, in a

different language, such an objective is used also in [MMT20], though arguably our treatment of such an objective as a max coverage constraint on sum-of-squares relaxations of A appears to demystify the use of crude-SDP in [MMT20]. We note that such a max coverage constraint is at the heart of the rounding algorithms for several problems in list-decodable learning starting with [KKK19].

A key consequence of the max coverage constraint is that, by an elementary convexity argument, it implies the following proposition:

Proposition 2.10 (Max coverage pseudo-distributions)**.** *For any pseudo-distribution on* F *satisfying* A *of degree at least* 2 *and minimizing* ..[F]$_2^2$*, we have* $\sum_{8\in(} ..[F_8] \geq =\cdot \frac{:^2}{=}$

A rounding algorithm now falls naturally out of the above two discussions. We look at an $=^C$ by = matrix indexed by subsets of size C = $(1/) on the rows and singleton vertices on the columns, whose value at index ((, 8) is $\frac{..[F_(F_8]}{..[F_( ]}$. Proposition 2.10 implies that the rows of this (huge) matrix corresponding to the unknown planted clique must have a large total sum. On the other hand, as a consequence of the biclique certificate, we learn that for such rows the columns corresponding to [=] \ (must have a low total contribution. Together these two statements allow us to use a simple greedy algorithm that selects a uniformly random row of the above matrix and takes the largest : entries to recover a list containing a set of : vertices that has a large constant fraction intersection with (. Such a set can then be refined using a simple combinatorial "cleanup" step.

3 Preliminaries

We will use letters , to denote graphs and also their {±1}-entry adjacency matrices. We adopt the convention that (8, 9) = 1 if edge {8, 9} is present in the graph . For any G ∈ $^=$ and (⊆ [=], we use G$_($ to denote the monomial $\prod_{8\in(} G_8$. For any G ∈ {0, 1}$^=$, we use |G| to denote $\sum_{8=1}^{=} G_8$. We use the notation $(=) and Ω(=) to mean an absolute constant multiplied by = (in the former case, a "large enough" constant, and in the latter case, a "small enough" constant).

The bit complexity of a rational number ?/@ is ⌈log$_2$?⌉ + ⌈log$_2$ @⌉.

3.1 Sum-of-squares preliminaries

We refer the reader to the monograph [FKP19] and the lecture notes [BS16] for a detailed exposition of the sum-of-squares method and its usage in average-case algorithm design. A *degree-ℓ pseudo-distribution* is a finitely-supported function : $^=$ → ℝ such that $\sum_G (G) = 1$ and $\sum_G (G)5(G)^2 \geq 0$ for every polynomial 5 of degree at most ℓ/2. We define the *pseudo-expectation* of a function 5 on ^{3} with respect to a pseudo-distribution , denoted ..$_{(G)}$5(G), as ..$_{(G)}$5(G) = $\sum_G (G)5(G)$. The degree-ℓ pseudo-moment tensor of a pseudo-distribution is the tensor ..$_{(G)}$(1, G$_1$, G$_2$, . . . , G$_=$)$^{\otimes \ell}$ with entries corresponding to pseudo-expectations of monomials of degree at most ℓ in G. The set of all degree-ℓ moment tensors of degree 3 pseudo-distributions is also closed and convex.

Definition 3.1 (Constrained pseudo-distributions)**.** Let be a degree-ℓ pseudo-distribution over $^=$. Let A = {5$_1$ ≥ 0, 5$_2$ ≥ 0, . . . , 5$_<$ ≥ 0} be a system of < polynomial inequality constraints. We

say that *satisfies the system of constraints* A *at degree* A (satisfies it -approximately, respectively), if for every (⊆ [<] and every sum-of-squares polynomial ⊢ with deg ⊢ + ∏_{8∈(} max{deg 5_8, A} ≤ ℓ, ⊢ · ∏_{8∈(} 5_8 ≥ 0 (⊢ · ∏_{8∈(} 5_8 ≥ · k⊢k_2 ∏_{8∈(} k 5_8 k_2 where k⊢k_2 for any polynomial ⊢ is the Euclidean norm of its coefficient vector, respectively). We say that satisfies (similarly for approximately satisfying) A (without mentioning degree) if satisfies A at degree 0.

Basic facts about pseudo-distributions.

Fact 3.2 (Hölder's inequality for pseudo-distributions)**.** *Let* 5, , *be polynomials of degree at most* 3 *in indeterminate* G ⊆ ⊆^3. *Fix* C ⊆ ⊆. *Then, for any degree* 3C *pseudo-distribution* , ..[5^{C−1},] ≤ (..[5^C])^{C−1} (..[, ^C])^{1/C}.

Observe that the special case of C = 2 corresponds to the Cauchy-Schwarz inequality. The following idea of *reweighted* pseudo-distributions follows immediately from definitions and was first formalized and used in [BKS17]).

Fact 3.3 (Reweightings [BKS17])**.** *Let be a pseudo-distribution of degree* : *satisfying a set of polynomial constraints* A *in variable* G. *Let* ? *be a sum-of-squares polynomial of degree* C *such that* ..[?(G)] ≠ 0. *Let* ′ *be the pseudo-distribution defined so that for any polynomial* 5, ..[5(G)] = ..[?(G)5(G)]/..[?(G)]. *Then,* ′ *is a pseudo-distribution of degree* : − C *satisfying* A.

Sum-of-squares proofs. A *sum-of-squares proof* that the constraints {5_1 ≥ 0, . . . , 5_< ≥ 0} imply the constraint { , ≥ 0} consists of sum-of-squares polynomials (?_( )_{(⊆[<]} such that , = ∏_{(⊆[<]} ?_(· Π_{8∈(} 5_8.

We say that this proof has *degree* ℓ if for every set (⊆ [<], the polynomial ?_(Π_{8∈(} 5_8 has degree at most ℓ and write:

$$\{5_8 \geq 0 \mid 8 \leq A\} \vdash_{\ell} \{ , \geq 0\}.$$

Fact 3.4 (Soundness)**.** *If satisfies* A *for a degree-*ℓ *pseudo-distribution and there exists a sum-of-squares proof* A ⊢_{A′} ⊢, *then satisfies* ⊢ *at degree* AA′ + A′.

Definition 3.5 (Total bit complexity of sum-of-squares proofs)**.** Let 5_1, 5_2, . . . , 5_< be polynomials in indeterminate G with rational coefficients. For a polynomial , with rational coefficients, we say that {5_1 ≥ 0, . . . , 5_< ≥ 0} derives { , ≥ 0} in degree : and total bit complexity if , = ∏_{(⊆[<]} ?_(· Π_{8∈(} 5_8 where each ?_(is a sum-of-squares polynomial of degree at most : − ∏_{8∈(} deg(5_8) for every (, and the total number number of bits required to describe all the coefficients of all the polynomials 5_8, , , ?_(is at most .

There's an efficient separation oracle for moment tensors of pseudo-distributions that allows approximate optimization of linear functions of pseudo-moment tensors approximately satisfying constraints. The *degree-*ℓ *sum-of-squares algorithm* optimizes over the space of all degree-ℓ pseudo-distributions that approximately satisfy a given set of polynomial constraints:

Fact 3.6 (Efficient optimization over pseudo-distributions [Sho87, Par00, Nes00, Las01])**.** *Let* > 0. *There exist an algorithm that for* =, < ⊆ ⊆ *runs in time* (= + <)^{$(ℓ)} poly log 1/, *takes input an explicitly bounded and satisfiable system of* < *polynomial constraints* A *in* = *variables with rational coefficients and outputs a level-*ℓ *pseudo-distribution that satisfies* A *-approximately.*

Basic sum-of-squares proofs.

Fact 3.7 (Operator norm bound). *Let be a symmetric* 3 × 3 *matrix with rational entries with numerators and denominators upper-bounded by* 2 *and* E *be a vector in* ▯^{3}. *Then, for every* ≥ 0*,*

$$\vdash_{5} \; E^{▯}E \leq kk_2 kEk_2 + \quad {}_2$$

Further, the total bit complexity of the sum-of-squares proof is poly(, 3, log 1/).

Fact 3.8 (SoS Hölder's inequality). *Let* $5_{8},_{,8}$ *for* 1 ≤ 8 ≤ B *be indeterminates. Let* ? *be an even positive integer. Then,*

$$\vdash_{?^2}^{5,,} \; ▯ \frac{1}{B} \tilde{O}^{B}_{8=1} \, 5_8,_{8}^{?-1} \; !^{?} \leq \; \frac{1}{B} \tilde{O}^{B}_{8=1} 5_8^{?} \; ! \; \frac{1}{B} \tilde{O}^{B}_{8=1} ,_8^{?} \; !^{?-1} ▯ \, .$$

Further, the total bit complexity of the sum-of-squares proof is B$^{\$(?)}$.

Observe that using ? = 2 yields the SoS Cauchy-Schwarz inequality.

Fact 3.9 (SoS almost triangle inequality). *Let* $5_1, 5_2, \ldots, 5_A$ *be indeterminates. Then,*

$$\vdash_{2C}^{5_1,5_2,\ldots,5_A} \; ▯ \tilde{O}_{8 \leq A} 5_8 \; !^{2C} \leq A^{2C-1} \; \tilde{O}^{A}_{8=1} 5_8^{2C} \; ! ▯ \, .$$

Further, the total bit complexity of the sum-of-squares proof is A$^{\$(C)}$.

Fact 3.10 (SoS AM-GM inequality, see Appendix A of [BKS15]). *Let* $5_1, 5_2, \ldots, 5_<$ *be indeterminates. Then,*

$$\{ 5_8 \geq 0 \mid 8 \leq < \} \vdash_{<}^{5_1, 5_2, \ldots, 5_<} \left(\; \frac{1}{<} \tilde{O}^{<}_{8=1} 5_8 \; !^{<} \geq \Pi_{8 \leq <} 5_8 \right) .$$

Further, the total bit complexity of the sum-of-squares proof is exp($(<)).

Fact 3.11 (Cancellation within sum-of-squares, Lemma 9.3 in [BKar]). *Let* 0, *be indeterminates. Then,*

$$\{0 \geq 0\} ▯ \{0^C \leq 0^{C-1}\} \vdash_{2C}^{0,} 0^{2C} \leq {}^{2C} \qquad .$$

Further, the total bit complexity of the sum-of-squares proof is exp($(C)).

Fact 3.12 (Univariate sum-of-squares proofs). *Let* ? *be a degree-*3 *univariate polynomial with rational coefficients of bit complexity such that* ?(G) ≥ 0 *for every* G ▯ ▯*. Then, for every* > 0*, there is a degree-*3 *sum-of-squares polynomial* @(G) *with coefficients of bit complexity* $(poly(, log 1/)) *such that* + ?(G) = @(G)*.*

Lemma 3.13 (Simple cancellation within sum-of-squares). *Let* 0 *be an indeterminate and be some positive constant. Then,*

1.

$$\{0^2 \le 0\} \;{}_2\big|^{0}\; 0^2 \le {}^2 \quad .$$

2.

$$\{0^2 \le \;\} \;{}_2\big|^{0}\; 0 \le \sqrt{\;} \quad .$$

The total bit complexity of the sum-of-squares proofs is poly().

Proof. For the first claim, we have:

$$\{0^2 \le 0\} \;{}_2\big|^{0}\; 0^2 \le 0^2 + (0 - \;)^2 = {}^2 + 20^2 - 20 \le {}^2 \quad .$$

For the second claim, note that it is enough to prove the claim for $= 1$ (and apply this special case to $0/$). Using the fact that ${}_2\big|^{0}$ $(1 + 0)^2 \le 20^2 + 2$, we have:

$$\{0^2 \le 1\} \;\big|^{0}_{2}\; 0 = \frac{1}{4}(0 + 1)^2 - \frac{1}{4}(1 - 0)^2 \le \frac{1}{2}(0^2 + 1) \le 1 \quad .$$

We also need the following fact about random matrices:

Fact 3.14 (Singular values of random matrices, consequence of Theorem 2.3.21 [Tao12])**.** *Fix any* > 0*. Let be a* $: \times =$ *matrix for* $: \le =$ *with independent entries with magnitude at most* $=^{0.5-}$*, mean* 0 *and variance* 1*. Then, for large enough* $=$*, with probability at least* 0.99, $\|\;\|_2 \le \$(\sqrt{\;=})$*.*

Fact 3.15 (Singular values of rectangular random matrices, consequence of Theorem 4.5.1 [Ver18])**.** *Let be a* $: \times =$ *matrix for* $: \le =$ *with independent entries chosen uniformly from* $\{-1, 1\}$*. Then, with probability at least* 0.99 *over the draw of entries of , the largest singular value of is at most* $\$(\sqrt{\;=})$ *and the* $:$*-th smallest singular value of is at least* $\Omega(\sqrt{\;=} - \sqrt{: - 1})$*.*

4 Certifying biclique bounds in unbalanced random bipartite graphs

In this section, we develop low-degree sum-of-squares certificates of upper bounds on biclique sizes in unbalanced random bipartite graphs. We use $= (*, +,)$ to denote a bipartite graph with left vertex set $*$, right vertex set $+$, and edge set .

For a bipartite graph $= (*, +,)$, let ⍰ = ⍰() be the following system of polynomial constraints, which has as solution every biclique $((,))$ in of total size $:$ with $(= \{D ⍰ * \mid G_D = 1\}$ and $) = \{E ⍰ + \mid H_E = 1\}$:

$$⍰() : \quad \begin{cases} ⍰ D ⍰ * & G_D^2 = G_D \\ ⍰ E ⍰ + & H_E^2 = H_E \\ & |G| + |H| = : \\ ⍰ D ⍰ *, E ⍰ + \text{ s.t. } \{D, E\} ⍰ & G_D H_E = 0 \end{cases} \quad . \tag{3}$$

Remark 4.1. Notice that the biclique formulation above places a constraint on the total size of the clique. This is the natural formulation that arises in our reduction from the semi-random planted clique problem. Intuitively, given a graph ⍰ FK(=, :, 1/2), the bipartite graph we care about is = cut((⍰) where (⍰ is the planted clique of size :. The bicliques we want to refute are obtained by taking an arbitrary :-clique (in and looking at the induced biclique in with left vertices (∩ (⍰ and right vertices (\ (⍰. In particular, notice that the total size of the biclique is : and as long as (≠ (⍰ the right hand side of the clique contains at least one vertex. For more a detailed commentary, we direct the reader to Section 2.

For ease of exposition, we will present our certificates and analysis for the most important case of ? = 1/2 first and then follow it up with a generalization to arbitrary ? in the following subsection.

4.1 The case of ? = 1/2

The goal of the following theorem is to show that with high probability over the draw ⍰(:, = − :, 1/2) of a bipartite Erdős-Rényi random graph with edge density ? = 1/2 and : ≤ =, there is a degree A (and thus verifiable in time $=^{\$(A)}$) sum-of-squares certificate of (informally speaking) the fact that any ℓ by : − ℓ biclique with : − ℓ ≥ 1 satisfies ℓ ≤ poly(A) · $(=/:)^{\$(1/A)}$. In particular, for any > 0, by choosing A = $(1/), we get an $=^{\$(1/)}$-time verifiable certificate of the absence of = × (: − =)-bicliques in . Formally, we will prove:

Theorem 4.2 (Sum-of-squares certificates for unbalanced bicliques in random bipartite graphs)**.** *Let* ⍰(:, = − :, 1/2) *with* : ≤ = *be a bipartite Erdős-Rényi graph with edge probability* 1/2*. Then, for any* A ≤ $\$\left(\frac{:^2}{= \log =}\right)$*, with probability at least* 0.99 *over the draw of , the sizes of the sets indicated by* G *and* H *respectively satisfy*

$$⍰() \;{}_4{}^{A}\big|^{G,H}_{2} \quad |G|^{4A}|H| \le (1000A)^{10A} = \left(\frac{=}{:}\right)^4 .$$

Further, the total bit complexity of the sum-of-squares proof is $=^{\$(A)}$.

As a corollary, we obtain that with high probability over the choice of ⍰(:, = − :, 1/2), for every pseudo-distribution of degree at least 4A + 2 satisfying ⍰(), we must have $\tilde{..}[|G|^{4A}|H|] \le (1000A)^{10A} =(=/:)^4$.

Remark 4.3. Observe that if contains an ℓ by : − ℓ biclique for : − ℓ ≥ 1 then the above theorem yields that $ℓ^{4A}(: - ℓ) \le |G|^{4A}|H| \le (1000A)^{10A} = (=/:)^4$ and thus, ℓ ≤ poly(A) · $=^{\$(1/A)}$. That is, there exist degree $(A) certificates of absence of ℓ by : − ℓ bicliques in for ℓ ⍰ $=^{\$(1/A)}$.

Our proof of Theorem 4.2 uses two simple pseudorandom properties of the graph and thus works for all graphs that satisfy these properties. For every (⍰ *, let $D_($ be a vector in $\{-1, 1\}^{|+|}$ so that $D_((9) = 1_{8⍰(}(8, 9)$. Then, we will need the following A-fold balancedness property that informally asks that the vectors $D_($ be nearly balanced for all subsets (⍰ * of size at most A. Additionally, we will need that every vertex on the *right* side of has degree no larger than $:/2 + \$(\sqrt{: \log =})$.

Definition 4.4 (Balancedness). Let $G = (A, B, E)$ be a bipartite graph. For every $S \subseteq A$, let v_S be the $|B|$-dimensional vector defined by setting $v_S(j) = \prod_{i\in S} G(i, j)$. Then, we say that G has k-*fold balancedness* Δ_k if, for all $S \subseteq A$ of size $|S| \le k$, it holds that $\left|\sum_{j\in B} v_S(j)\right| \le \Delta_k$.

The following lemma verifies that the two pseudorandom properties hold for random bipartite graphs by a simple application of Hoeffding's inequality and union bounds.

Lemma 4.5 (Balancedness of random bipartite graphs). *Let $G = (A, B, E) \sim \mathcal{G}(n, m - n, 1/2)$. Then, for any $k \le |A|$, with probability at least 0.99 over the draw of G, 1) G has k-fold balancedness $O(\sqrt{km \log n})$, and 2) the maximum degree of a vertex in B is at most $n/2 + O(\sqrt{n \log m})$.*

Proof. For $S \subseteq A$ such that $|S| \le k$, we have that $v_S(j)$ has mean 0 and is bounded between -1 and 1. Then, by Hoeffding's inequality,

$$\Pr\left[\left|\sum_{j\in B} v_{?,S}(j)v_{?,T}(j)\right| \ge t\sqrt{|B|}\right] \le 2e^{-t^2/2},$$

so, by a union bound over all choices of S,

$$\Pr\left[\exists S \subseteq A \text{ s.t. } |S| \le k, \left|\sum_{j\in B} v_{?,S}(j)v_{?,T}(j)\right| \ge t\sqrt{|B|}\right] \le |A|^k \cdot 2e^{-t^2/2}.$$

Choosing $t = O(\sqrt{k \log |A|})$ makes the right-hand side a small constant, so we have k-fold balancedness $O(\sqrt{k|B| \log |A|})$.

The degree of a vertex in B is a binomial random variable $\mathrm{Bin}(n, 1/2)$, which by standard bounds is larger than $n/2 + t$ with probability at most $e^{-t^2/n}$. By a union bound over all vertices in B, the maximum degree is larger than $n/2 + t$ with probability at most $|B|e^{-t^2/n}$, so choosing $t = O(\sqrt{n \log |B|})$ makes the probability a small constant. Hence, the maximum degree is at most $n/2 + O(\sqrt{n \log |B|})$.

The key component of the proof of Theorem 4.2 is the following lemma that gives a sum-of-squares certificate of an upper bound on a quantity closed related to $|x|^4_k|y|$.

Lemma 4.6. *Let $G = (A, B, E)$ be a bipartite graph with $|A| = n$ and $|B| = m - n$ and $2k$-fold balancedness Δ_{2k}. Then,*

$$\mathcal{A}() \vdash^{x,y}_{4k} \left\{ \left(\sum_{|S|=k} x_S\right)^2 |y| \le m \left(\sum_{|S|\le k} x_S\right) + \Delta_{2k} \left(\sum_{|S|\le k} x_S\right)^2 \right\}. \tag{4}$$

Further, the total bit complexity of the sum-of-squares proof is $m^{O(k)}$.

Remark 4.7 (Proof plan). In order to interpret this lemma, we suggest the readers to think of $\sum_{|S|=k} x_S \approx |x|^k$ (this is formally shown to be fine in Lemma 4.10). Then, rearranging the conclusion of the lemma yields a statement of the form $|x|^{2k}(|y| - \Delta_{2k}) \le |x|^k m$. At this point, "in real life" (as

opposed to within the sum-of-squares proof system), we could reason as follows: if |H| > Δ2A, then "canceling" |G|^A from both sides and "dividing through" by (|H| − Δ2A) yields that |G|^A ≤ =, giving us a bound on the left hand side of biclique as desired. On the other hand, if |H| ≤ Δ2A, then for : ⍰ Δ2A we have |G| ⍰ :/2, which can be ruled out by the upper bound on the maximum degree of a vertex on the right side.

This argument, however, is not easy to implement within the low-degree sum-of-squares proof system because of the case analysis involved. Indeed, a similar issue arises in the context of list-decodable learning and robust clustering algorithms that rely on certifiable anticoncentration (see overview of [BK20] for a discussion and a general resolution, and also the discussion on the need for *a priori* bounds in [DHKK20]). In our situation, we can resolve this need using a more straightforward observation (see Lemma 4.8). The rest of the steps above can indeed by implemented within low-degree sum-of-squares via cancellation inequalities (e.g., see Fact 3.11).

We postpone the proof of this key lemma and first show how to use it. The following simple lemma uses a bound on the degree of the right vertices in in order to lower bound the LHS of the conclusion of Lemma 4.6. This will allow us to eliminate the term $\Delta_{2A} \sum_{|(|=A} G_(^2$ from the RHS of the conclusion of Lemma 4.6.

Lemma 4.8 (Lower bounding the LHS of (4)). *Let = (* , + ,) be a bipartite graph with |*| = : and |+| = = − : and maximum degree of a vertex in + at most :/2 + Δℓ. Then, for any 9⍰ +, we have:*

$$⍰() \;\; {}_{4A}\vdash^{G,H}_{2} \;\; H_9 \left(\sum_{(:|(|=A} G_(|H| \right)^2 \geq \left(\frac{:}{2} - \Delta_\ell \right) H_9 \left(\sum_{(:|(|=A} G_(\right)^2 .$$

Further, the total bit complexity of the sum-of-squares proof is $=^{\$(A)}$.

Proof. Every 9⍰ + has degree at most $\frac{:}{2} + \Delta_\ell$. Thus, we have using the constraint system ⍰()

$$⍰() \;\; \vdash^{G,H}_{2} \;\; H_9 |G| = \sum_{8⍰*:\{8,9\}⍰} G_8 H_9 \leq \left(\frac{:}{2} + \Delta_\ell \right) H_9 .$$

Thus, $⍰() \vdash^{G,H}_{2} \{|G| = : - |H|\}$ allows us to conclude:

$$⍰() \;\; \vdash^{G,H}_{2} \;\; H_9 |H| \geq \left(\frac{:}{2} - \Delta_\ell \right) H_9$$

Multiplying both sides by the sum-of-squares polynomial $\sum_{(:|(|=A} G_(^2$ completes the proof.

As a direct consequence of Lemma 4.6 and Lemma 4.8, we obtain:

Lemma 4.9. *Let = (* , + ,) be a bipartite graph with |*| = : and |+| = = − : and 2A-fold balancedness Δ2A and maximum degree of a vertex in + at most :/2 + Δℓ. Suppose further that* $\frac{:}{2} - \Delta_\ell - \Delta_{2A} \geq \frac{:}{4}$. *Then, we have:*

$$?() \; {}_{4A+2}\Big|^{G,H} \left(© \tilde{O}_{«(:|(|=A} G_{(}^{ª} \; |H|^{®} ¬ \leq = 4= \frac{\quad}{:} \right)^{4} . $$

Further, the total bit complexity of the sum-of-squares proof is $=^{\$(A)}$.

Proof. We first multiply both sides of the conclusion of Lemma 4.6 with the sum-of-squares polynomial H_9^2 for an arbitrary $9 ? +$:

$$?() \; {}_{4A|2}\Big|^{G,H} H_9 \left(© \tilde{O} \; G_{(}^{®} \; ª|H| ¬ \right)^{2} \leq =H_9 \; © \tilde{O}_{«|(|=A} G_{(}^{®} ª ¬ + \Delta_{2A} H_9 \; © \tilde{O}_{«|(|=A} G_{(}^{®} ª ¬ \; . «|(|=A $$

Next, we use Lemma 4.8 to replace the left-hand side by a useful lower bound:

$$?() \; {}_{4A|2}\Big|^{G,H} \left(\frac{:}{2} \Delta_{\ell} \; H_9 \; © \tilde{O}_{«(:|(|=A} G_{(}^{®} ª ¬ \right)^{2} \leq =H_9 \; © \tilde{O}_{«|(|=A} G_{(}^{®} ª ¬ + \Delta_{2A} H_9 \; © \tilde{O}_{«|(|=A} G_{(}^{®} ª ¬ \; . $$

We then move the second term on the right-hand side to the left-hand side and use that $\frac{:}{2} - \Delta_{\ell} - \Delta^2{}_A \geq \frac{:}{4}$ to conclude:

$$?() \; {}_{4A+2}\Big|^{G,H} \left(H_9 \; © \tilde{O}_{«(:|(|=A} G_{(}^{®} ª ¬ \right)^{2} \leq \frac{4=}{:} H_9 \; © \tilde{O}_{«|(|=A} G_{(}^{®} ª ¬ . $$

We finally apply Lemma 3.11 with $0 = H_9(\sum_{(:|(|=A} G_{(})$, $= \frac{4=}{:}$ and $C = 2$ to obtain:

$$?() \; {}_{4A+2}\Big|^{G,H} \left(H_9 \; © \tilde{O}_{«(:|(|=A} G_{(}^{®} ª ¬ \right)^{4} \leq H_9^4 \; © \tilde{O}_{«(:|(|=A} G_{(}^{®} ª ¬ \leq \left(\frac{4=}{:} \right)^{4} . $$

Summing up as 9 varies over + completes the proof.

Finally, we invoke the following simple observation that allows us to replace $\sum_{|(|=A} G_{(}$ by $|G|^{A}$:

Lemma 4.10. *For every* > 0*, there is a sum-of-squares proof with coefficients of bit complexity* $\$(\text{poly}(|*|, \log 1/))$

$$ \left(G^{2}_{8} = G_8 \; ?8 ? * \right) \; \Big|^{G}_{A} \; \left(\frac{1}{2^A A!} \left(\tilde{O}_{8} G_8 \right)^{A} - \frac{2A^A}{A!} - \leq \tilde{O}_{|(|=A} G_{(} \leq \frac{1}{A!} \left(\tilde{O}_{8} G_8 \right)^{A} + \right) . \tag{5} $$

Proof. The following polynomial identity holds: $\sum_{|(|=A} G_{(} = \frac{1}{A!} (\sum_8 G_8)(\sum_8 G_8 - 1) \cdots (\sum_8 G_8 - (A-1))$. Then $\sum_{|(|=A} G_{(} \leq \frac{1}{A!} (\sum_8 G_8)^{A}$ and

$$ \tilde{O}_{|(|=A} G_{(} \geq \frac{1}{A!} \left(\tilde{O}_{8} G_8 - A \right)^{A} - \frac{A^A}{A!} \geq \frac{1}{2^A A!} \left(\tilde{O}_{8} G_8 \right)^{A} - \frac{2A^A}{A!} , $$

where in the last inequality the subtracted term makes the inequality trivial unless Í $_8$ G$_8$ ≥ 2A, case in which we use that Í $_8$ G$_8$ − A ≥ Í $_8$ G$_8$/2.

Notice that Í $_{|(|=A}$ G$_($ is a univariate degree-A polynomial in Í $_8$ G$_8$. Then the inequalities Í $_{|(|=A}$ G$_($ ≤ $\frac{1}{A!}$(Í $_8$ G$_8$)^A and Í $_{|(|=A}$ G$_($ ≥ $\frac{1}{A!}$(Í $_8$ G$_8$)^A − $\frac{2A^A}{!}$ can be written as univariate polynomial inequalities ?$_*$(Í $_8$ G$_8$) ≥ 0 and ?$_!$(Í $_8$ G$_8$) ≥ 0, respectively, with ?$_*$ and ?$_!$ of degree at most A. It is easy to check that the coefficients of ?$_*$ and ?$_!$ have bit complexity $(poly(|*|)), so by Fact 3.12 the conclusion follows.

We can finish the proof of Theorem 4.2 from here:

Proof of Theorem 4.2. From Lemma 4.10, we have:

$$G^8 = G_8 \text{ 8 * } \Big|\frac{G}{A} \left(\frac{1}{2^A A!}|G|_A - \frac{2A^A}{A!} \right) - \left(\tilde{O} \le \sum_{|(|=A} G_(\right).$$

Setting = 1 and using that {0 ≤ 0 ≤ } ^{0,} |{0^4/4 ≤ ^{4}}, we have:

$$(\,)\quad G^2 = G_8 \text{ 8 * } \quad (\;^4)\; {}_{4A}\; |G|^4_A \le \tilde{O}(100A)^{10A} + (100A)^{10A} \copyright\; G_(\Big|\frac{G}{a}.$$

« (|$_{=A}$ ¬

Now we want to combine this with the conclusion of Lemma 4.9. We briefly verify that we satisfy the condition $\frac{:}{2}$ − Δ$_ℓ$ − Δ2$_A$ ≥ $\frac{:}{4}$. We have by Lemma 4.5 that Δ$_ℓ$ = $($\sqrt{: \log |+|}$) = $($\sqrt{: \log =}$) and Δ2$_A$ = $($\sqrt{A|+|\log|*|}$) = $($\sqrt{A=\log =}$). Observe that for : ≥ $($\sqrt{A=\log =}$) large enough the condition is satisfied. Then we have:

$$\text{()}\ {}^{4A}\Big|\frac{G,H}{2}\ \left(|G|^{4A}|H| \le (100A)^{10A}|H| + (100A)^{10A} = \frac{4=}{:} \right)^4.$$

Observing that () $_2$ |{|H| ≤ =} completes the proof.

Proof of Lemma 4.6. We now return to the proof of Lemma 4.6.

Proof of Lemma 4.6. Let us write D$'_($ for the vector-valued linear function in indeterminate H defined by D$'_($(8) = D$_($(8)(1 − H$_8$). Then, observe that () ^{2A}|$\frac{G,H}{2}$ {G$_($D$_($(8)H$_8$ = G$_($H$_8$}. In particular, () ^{G,H} |$_{4A}$ n G$_($D$'^{\,2}_($ ≤ $_2$ G$^($(= − |H|) o. Further, for any ^{A} and any (* such that |(| = A, we have:

$$\text{()}\ \Big|^{G,H}_{4A} G_(\left(\tilde{O} D_((8) = G_(\tilde{O}_8 D_((8)(1 - H_8) \le \Delta_{2A} G_(- G_(|H| \right)^8.$$

Next, let (,) * such that (≠) and |(|, |)| ≤ A. Then, by noting that D$^{('}$∘ D$_)$$'$ = D$_(^{Δ}_)$, we have:

$$\text{()}\ \Big|^{G,H}_{4A}\ G_(G_)\langle D'_(, D_)\rangle = G_{()} \left(\tilde{O}_8 D_{(\Delta)}(8)(1 - H_8) \le \Delta 2_A G_{()} - G_{()}|H| \right).$$

Next, we have: [3]

$$?() \overset{G,H}{\underset{4A}{\vdash}} 0 \leq \sum_{|(|=A} G_{(} D' {}^{2} = \sum_{|(|=A} G_{(} D'^{2} + \sum_{\substack{(\ ) \\ |(|=A \\ (\neq)}} G_{(} D', G_{)} D$$

$$\leq \sum_{|(|=A} G_{(} (= - |H|) + \Delta_{2A} \sum_{(\neq)} (G_{(?)} - G_{(?)}|H|)$$

$$\leq = \sum_{|(|=A} G_{(} + \Delta_{2A} \sum_{(,)?*,|(|,|)|=A} G_{(?)} - \sum_{(,)?*,|(|,|)|=A} G_{(?)}|H|$$

$$= \sum_{|(|=A} G_{(} = + \Delta_{2A} \left(\sum_{|(|=A} G_{(} \right)^{2} - \left(\sum_{|(|=A} G_{(} \right)^{2} |H| .$$

Rearranging gives:

$$?() \overset{G,H}{\underset{4A}{\vdash}} \left(\sum_{|(|=A} G_{(} \right)^{2} |H| \leq = \sum_{|(|=A} G_{(} + \Delta_{2A} \left(\sum_{|(|=A} G_{(} \right)^{2}$$

4.2 The case of arbitrary ?

In this section, we generalize the certificates of Section 4 to general edge densities. The certificates use the same system of polynomial constraints ?() as in the previous section.

Theorem 4.11 (Sum-of-squares certificates for unbalanced bicliques in random bipartite graphs for general densities)**.** *Let* ? (:, = − :, ?) *be a bipartite Erdős-Rényi with edge probability* ?. *Then with probability* 0.99 *we obtain the following two bounds:*

1. *Fix any* > 0 *independently of the other parameters. For* ?, 1 − ? ≥ $=^{-(1-)}$,

$$?() \overset{G,H}{\underset{4}{\vdash}} |G||H| \leq \$ \frac{=?}{1-?} .$$

2. *For any* A *such that* : ≥ max{$($\sqrt{A = \log = ?^{2A}/(1-?)^{2A+1}}$), $((log =)?/(1 − ?))}, ?()

$$\overset{G,H}{\underset{4A+2}{\vdash}} |G|^{4A}|H| \leq (1000A)^{10A} = \frac{= \max\{?/(1-?), (1-?)/?\}^{A} ?^{A} {}^{4}}{:(1-?)^{A+1}} .$$

[3] This is a sum-of-squares proof of the classical fact upper bounding the number of negatively correlated vectors in = dimensions.

Further, the total bit complexity of the sum-of-squares proofs is $=^{\$(1)}$ *and* $=^{\$(A)}$*, respectively.*

For the proof of Theorem 4.11, we will work with matrices with ?-biased characters as entries. We first define these well-studied objects.

Definition 4.12 (?-biased characters and normalized adjacency matrix)**.** Let = (*, +,) be a bipartite graph. We define the ?-biased character corresponding to an edge (8, 9) to be

$$
?(8,9) = \begin{cases} \sqrt{\frac{1-?}{?}} & \text{if } (8,9) = 1\,, \\ -\sqrt{\frac{?}{1-?}} & \text{if } (8,9) = -1\,. \end{cases}
$$

The normalized adjacency matrix $_?$ of the graph is matrix with the (8, 9)-th entry equal to $_?(8, 9)$.

Let us first analyze a simple spectral certificate (that confirms that our algorithm recovers the bounds of [MMT20, CSV17] from a basic relaxation in our scheme) in order to recover the first bound above.

Lemma 4.13 (Simple spectral certificate)**.** *Let* ? (:, = − :, ?) *be a bipartite Erdős-Rényi with edge probability* ?*. Fix any* > 0*. Then, for any* ?, 1 − ? ≥ $=^{-(1-)}$*, we have:*

$$
?() \;\Big|\!\!\frac{G,H}{4}\; |G||H| = {}_?^2 \le \$ \;\frac{1-?}{2} \cdot \frac{=?}{}
$$

Further, the total bit complexity of the sum-of-squares proof is $=^{\$(1)}$*.*

Proof. We have:

$$
?() \;\Big|\!\!\frac{G,H}{4}\; \frac{1-?}{?}|G|^2|H|^2 = G^{?}{}_?H^2 \le \|G\|_2^2\,{}_?H^2 \le \|G\|_2^2\,{}_?^2\,\|H\|_2^2\,.
$$

In the inequality above, we used the sum-of-squares Cauchy-Schwarz inequality.

Applying the first part of Lemma 3.13 with $0 = \|G\|_2^2\|H\|_2^2$:

$$
?() \;\Big|\!\!\frac{G,H}{4}\; |G|^2|H|^2 \le \frac{?^2{}_?^2}{(1-?)^2} \cdot {}^4
$$

Applying the second part of Lemma 3.13 with $0 = \|G\|_2^2\|H\|_2^2$ we obtain:

$$
?() \;\Big|\!\!\frac{G,H}{4}\; |G||H| \le \frac{?}{1-?}\,{}_?^2 \cdot {}^2
$$

Finally, notice that the entries of $_?$ are mean 0, variance 1 and are bounded above by $\max\{\sqrt{?/1-?}, \sqrt{1-?/?}\}$ in magnitude. For ?, 1 − ? ≥ $=^{-(1-)}$, the bound on the entries evaluates to $=^{0.5-/2}$. So we can apply Fact 3.14 to conclude that $\|{}_?\|_2 = \$(\sqrt{=})$ with probability at least 0.99.

The proof of second bound in Theorem 4.11 uses a generalization of A-fold balancedness defined in terms of ?-biased characters. We call this new property A-fold ?-balancedness.

Definition 4.14 (Balancednes for general densities)**.** Let = (* , + ,) be a bipartite graph. For every (⊆ * , let D?,(be the |+|-dimensional vector defined by setting D?,((9) = ∏ 8∈(?(8, 9). Then, we say that has 2A-*fold* ?-*balancedness* Δ if, for all (,) ⊆ * of size |((|, |)| ≤ A, it holds that | ∑ 9∈+ D?,((9)D?,)(9)| ≤ Δ.

The following lemma verifies A-fold ?-balancedness of random bipartite graphs, as well as an upper bound on the maximum degree of the vertices on the righ-hand side.

Lemma 4.15 (Balancedness of random bipartite graphs for general densities)**.** *Let* = (* , + ,) ∼ (:, = − :, ?)*. Then, for any* A ≤ :*, with probability at least* 0.99 *over the draw of , 1) has* 2A*-fold* ?*-balancedness* $ (√(= log : ?^A/(1 − ?)^A)*, and 2) the maximum degree of a vertex in + is at most* : ? + $ (√(: ?(1 − ?) log =)).

Proof. For (,) ⊆ * such that |((|, |)| ≤ A, we have that D?,((9)D?,)(9) has mean 0 and is bounded between − √(?/(1−?))^{2A} = −?^A/(1 − ?)^A and √(?/(1−?))^{2A} = ?^A/(1 − ?)^A. Then, by Hoeffding's inequality,

$$\Pr\left[\left|\sum_{9 \in +} D_{?,(}(9)D_{?,)}(9)\right| \geq C\sqrt{|+|}\, ?^A/(1-?)^A\right] \leq 24^{-C^2/2},$$

so, by a union bound over all choices of (and) ,

$$\Pr\left[\exists (,) \subseteq * \text{ s.t. } |(|,|)| \leq A, \left|\sum_{9\in +} D_{?,(}(9)D_{?,)}(9)\right| \geq C\sqrt{|+|}\,?^A/(1-?)^A\right] \leq |*|^{2A} \cdot 24^{-C^2/2}.$$

Choosing C = $ (√(A log *)) makes the right-hand side a small constant, so we have 2A-fold ?-balancedness $ (√(A|+| log |*|) ?^A/(1 − ?)^A).

The degree of a vertex in + is a binomial random variable Bin(:, ?), which by standard bounds is larger than : ? + C with probability at most min{$4^{-C^2/(2:(1-?))}$, $4^{-C^2/(2:?+2C/3)}$}. By a union bound over all vertices in + , the maximum degree is larger than : ? + C with probability at most |+| min{$4^{-C^2/(2:(1-?))}$, $4^{-C^2/(2:?+2C/3)}$}, so choosing C = $ (√(: ?(1 − ?) log |+|)) makes the probability a small constant. Hence, the maximum degree is : ? + $ (√(: ?(1 − ?) log |+|)).

The following lemma is the key component of the proof of Theorem 4.11, and is analogous to Lemma 4.6 in Section 4.

Lemma 4.16. *Let* = (* , + ,) *be a bipartite graph with* |*| = : *and* |+| = = − : *and* 2A*-fold* ?*-balancedness* Δ2A*. Then,*

$$\vdash^{G,H}_{4A} \left\{ (1-?)^A/?^A \left(\sum_{|(|=A} G_(\right)^2 \sum_8 H_8 \leq = \max\{?/(1-?), (1-?)/?\}^A \left(\sum_{|(|=A} G_(\right) + \Delta_{2A} \left(\sum_{|(|=A} G_(\right)^2 \right\}. \quad (6)$$

Further, the total bit complexity of the sum-of-squares proof is =$^{\$(A)}$*.*

We postpone the proof of the lemma, and combine the result with an observation analogous to that in Lemma 4.8.

Lemma 4.17 (Lower bounding the LHS of (6))**.** *Let* $= (*, +,)$ *be a bipartite graph with* $|*| = :$ *and* $|+| = = - :$ *and maximum degree of a vertex in* $+$ *at most* $: ? + \Delta_{\ell}$. *Then, for any* $9 ∈ +$, *we have:*

$$ⓐ() \Big|\frac{G,H}{4A} H_9 \left(\sum_{(:|(|=A} G_(\right)^2 |H| \geq (:(1-?) - \Delta_{\ell}) H_9 \left(\sum_{(:|(|=A} G_(\right)^2 .$$

Further, the total bit complexity of the sum-of-squares proof is $=^{\$(A)}$.

Proof. Every $9 ∈ +$ has degree at most $: ? + \Delta_{\ell}$. Thus, we have using the constraint system ⓐ():

$$ⓐ() \Big|\frac{G,H}{2} \left\{ H_9 |G| = \sum_{D ∈ * : D ∼ 9} G_D H_9 \leq (: ? + \Delta_{\ell}) H_9 \right\}$$

Thus, $ⓐ() \Big|\frac{G,H}{2} \{|G| = : - |H|\}$ allows us to conclude:

$$ⓐ() \Big|\frac{G,H}{2} H_9 |H| \geq (:(1-?) - \Delta_{\ell}) H_9$$

Multiplying both sides by the sum-of-squares polynomial $\left(\sum_{(:|(|=A} G_(\right)^2$ completes the proof.

As a direct consequence of Lemma 4.16 and Lemma 4.17, we obtain:

Lemma 4.18. *Let* $= (*, +,)$ *be a bipartite graph with* $|*| = :$ *and* $|+| = = - :$ *and* $2A$*-fold* $?$*-balancedness* Δ_{2A} *and maximum degree of a vertex in* $+$ *at most* $: ? + \Delta_{\ell}$. *Suppose further that*

$$:(1-?)^{A+1}/?^A - \Delta_{\ell}(1-?)^A/?_A - \Delta_{2A} \geq \frac{:}{2}(1-?)^{A+1}/?^A .$$

Then, we have

$$ⓐ() \Big|\frac{G,H}{4A+2} \left\{ \left(\sum_{(:|(|=A} G_(\right)^4 |H| \leq = \left(\frac{2 = \max\{?/(1-?), (1-?)/?\}^A ?_A}{:(1-?)^{A+1}} \right)^4 \right\} .$$

Further, the total bit complexity of the sum-of-squares proof is $=^{\$(A)}$.

Proof. We first multiply both sides of the conclusion of Lemma 4.16 with the sum-of-squares polynomial H_9^2 for an arbitrary $9 ∈ +$:

$$ⓐ() \Big|\frac{G,H}{4A+2} (1-?)^A/?^A H_9 \left(\sum_{(|=A} G_(\right)^2 |H|$$

$$\leq \; = \max\{?/(1-?), (1-?)/?\}^{A} H_9 \; © \; \tilde{O} \; G_{(} ^{®} ª + \Delta_{2A} H_9 \; © \; \tilde{O} \; G_{(}^{®} ª^{2}) .$$

« (|_= A ¬ « (|_= A ¬

Next, we use Lemma 4.17 to replace the left-hand side in the above by a useful lower bound:

$$?() \; {}_{4A+2}|^{G,H} (: (1-?) - \Delta_{\ell}) (1-?)^{A}/?^{A} H_9 \; © \; \tilde{O} \; G_{(} ª^{2} ®$$

« (: | (| = A ¬

$$\leq \; = \max\{?/(1-?), (1-?)/?\}^{A} H_9 \; © \; \tilde{O} \; G_{(} ª ® + \Delta_{2A} H_9 \; © \; \tilde{O} \; G_{(} ª ®^{2}) .$$

« (|_= A ¬ « (|_= A ¬

We then move the second term on the right-hand side to the left-hand side and use that $:(1-?)^{A+1}/?^{A} - \Delta_{\ell}(1-?)^{A}/?^{A} - \Delta_{2A} \geq \frac{:}{2}(1-?)^{A+1}/?^{A}$ to conclude:

$$?() \; |^{G,H}_{4A+2} \; H_9 \; © \; \tilde{O} \; G \; ª ®^{2} \leq \frac{2 = \max\{?/(1-?), (1-?)/?\}^{A} ?_{A}}{:(1-?)^{A+1}} H_9 \; © \; \tilde{O} \; G_{(} ª ® .$$

« (: | (| = A ¬ « | (| = A ¬

We finally apply Lemma 3.11 with $0 = H_9(\sum_{(:|(|=A} G_{(})$, $= \frac{2 = \max\{?/(1-?),(1-?)/?\}^{A} ?^{A}}{:(1-?)^{A+1}}$, and $C = 2$ to obtain:

$$?() \; |^{G,H}_{4A+2} \; H_9 \; © \; \tilde{O} \; G_{(} ª ®^{4} \leq \frac{2 = \max\{?/(1-?), (1-?)/?\}^{A} ?_{A}}{:(1-?)^{A+1}}^{4} .$$

« (: | (| = A ¬

Summing up as 9 varies over + completes the proof.

We now finish the proof of Theorem 4.11:

Proof of Theorem 4.11. The first bound follow by Lemma 4.13. In the rest of the proof we focus on the second bound.

From Lemma 4.10, we have:

$$G_{8}{}_{2} = G_{8} ?8 ? * \; |^{G}_{A} \; \frac{1}{2^{A} A!} |G|^{A} - \frac{2A^{A}}{A!} - \leq \tilde{O}_{|(|=A} G_{(} .$$

Setting = 1 and using that $\{0 \leq 0 \leq \}^{0,} |\{0^{4}_{4} \leq {}^{4}\}$, we have:

$$G_{8}{}^{2} = G_{8} ?8 ? * \; |^{G}_{4A} |G|^{4A} \leq (100A)^{1 \; A}_{0} + (100A)^{1 \; A}_{0} \; © \; \tilde{O} \; G_{(} ª ®^{4} .$$

« | (|_= A ¬

Now we want to combine this with the conclusion of Lemma 4.18. We briefly verify that we satisfy the condition $:(1-?)^{A+1}/?^{A} - \Delta_{\ell}(1-?)^{A}/?^{A} - \Delta_{2A} \geq \frac{:}{2}(1-?)^{A+1}/?^{A}$. We have by Lemma 4.15

that $\Delta_{\ell} = \$(\sqrt{:?(1-?)\log|+|}) = \$(\sqrt{:(1-?)\log =})$ and $\Delta_{2A} = \$(A\sqrt{|+|\log *}|?(1-?)^{A})^{A/} = \$(A=\log = ?^{A}/(1-?)^{A})$. Observe that for $: \geq \max\{\$((\log =)?/(1-?)), \$(A=\log = ?^{2A}/(1-?)^{2A+1})\}$ large enough the condition is satisfied. Then we have:

$$\mathbb{?}()_{\$(A)} \left| {}^{G,H} |G|^{4A}|H| \leq (100A)^{10A}|H| + (100A)^{10A} = \frac{2=\max\{?/(1-?),(1-?)/?\}^{A} ?^{A}}{:(1-?)^{A+1}}\right)^{4}.$$

Observing that $\mathbb{?}()_{2}\left|{}^{H}\{|H| \leq =\}\right.$ completes the proof.

Finally, we complete the proof of Lemma 4.16.

Proof of Lemma 4.16. Let us write $D'_{?,(}$ for the vector-valued linear function in indeterminate H defined by $D'_{?,(}(8) = D_{?,(}(8)(1-H_8)$. Then, observe that $\mathbb{?}()\left|{}^{G,H}_{4A}\right. G_{(?)}D_{?,(}(8)D_{?,)}(8)H_8 = G_{(?)}H_8(1-?)^{A}/?^{A}$ and

$$\mathbb{?}()\left|{}^{G,H}_{4A}\right. G_{(}\left(D'_{?,(}\right)^{2} = \sum_{8} G_{(}D_{?,(}(8)^{2}(1-H_8)$$

$$\leq =\max\{?/(1-?),(1-?)/?\}^{A} G_{(} - (1-?)_{A}/?_{A} G_{(}|H| .$$

Let $(,) \mathbb{?}^{*}$ such that $(\neq)$ and $|(|,|)| \leq A$. We have:

$$\mathbb{?}()\left|{}^{G,H}_{4A}\right. G_{(}G_{)}\langle D'_{(}, D'_{)}\rangle = G_{(?)}\sum_{8} D_{(}(8)D_{)}(8)(1-H_8) \leq \Delta^{2}_{A} G_{(?)} - (1-?)_{A}/?^{A}G_{(?)}|H| .$$

Then, we have:

$$\mathbb{?}()\left|{}^{G,H}_{4A}\right. 0 \leq \sum G_{(}D' \,{}^{2} = \sum_{|(|=A} G_{(}D'^{2} + \sum_{\substack{2\\ |(|=A\\ (\neq)}} G_{(}D', G_{)}D'$$

$$\leq \sum_{|(|=A} G_{(} = \max\{?/(1-?),(1-?)/?\}^{A} - (1-?)^{A}/?_{A}|H|$$

$$+ \Delta^{2}_{A} \sum_{(\neq)} G_{(?)} - (1-?)_{A}/?_{A}\; G_{(?)}|H|$$

$$\leq =\max\{?/(1-?),(1-?)/?\}^{A} \sum_{|(|=A} G_{(} + \sum_{(,)\mathbb{?}^{*},|(|,|)|=A} \Delta^{2}_{A} G_{(?)}$$

$$- (1-?)^{A}/?^{A} \sum_{(,)\mathbb{?}^{*},|(|,|)|=A} G_{(?)}|H|$$

$$= = \max\{?/(1-?), (1-?)/?\}^{A} ⓒ \tilde{O} G_{(} ® ª + \Delta_{2A} ⓒ \tilde{O} G_{(} ® ª^{2} - (1-?) /? {}^{A} ⓒ \tilde{O} G_{(} ® |H| ª^{2}) .$$

Rearranging gives:

$$□() \; {}^{G,H}_{4A} (1-?)^{A} /?^{A} \tilde{O} ⓒ G_{(} ª^{2} |H| ® \le = \max\{?/(1-?), (1-?)/?\}^{A} ⓒ \tilde{O} G_{(} ª + \Delta_{2A} ® ⓒ \tilde{O} G_{(} ª^{2}) .®$$

5 List-decoding semi-random planted cliques

In this section, we describe our algorithm for list-decoding semi-random planted cliques using high-constant degree sum-of-squares relaxations. We will abstract out our requirement of sum-of-squares refutation of biclique numbers in random bipartite graphs in order to transparently show that the explicitness of the certificate is irrelvant to our algorithm. In Section 5.1, we will immediately obtain our algorithmic results as a direct consequence of our certificates from the previous section and an elementary cleanup step that takes a list with an approximately correct candidate and fixes it up to a list containing the planted clique (□.

Theorem 5.1. *Fix any* C□ □*. There is an* $=^{\$(C)}$ *time algorithm that takes as input a graph on = vertices with the following guarantees. Suppose has a clique* (□ *of size : in it. Suppose that the bipartite graph defined by keeping only the edges from* cut((□) *in admits an* $(C)*-th order sum-of-squares certificate of unbalanced biclique number as below for some function* $ = $(=, :, ?)*:*

$$□() \; {}_{\$(C)} {}^{G,H} |G|^{C} |H| \le \$.$$

Then, if $\$ \cdot (=/:^{2})^{C} \le :$*, the algorithm outputs a list of* $\$((=/:)^{C})$ *subsets, each of size at most* $:/(1-2)$ *such that with probability at least* 0.99 *over the randomness of the algorithm there is an element* (*of the list that satisfies* $|(\cap (□| \ge (1-2):$.

In the main results, the list that comes from Theorem 5.1 will be pruned (using that (□ has small intersection with other :-cliques, see Lemma 5.7) and refined to consist of (1 + >(1))=/: cliques of size :.

We will prove Theorem 5.1 using the following natural algorithm. Recall the standard :-clique constraint system A defined earlier. Our rounding scheme is reminiscent of those used in rounding algorithms for list-decodable learning [KKK19, BK21, IK22].

Algorithm 5.2 (List-decoding semi-random planted cliques)**.**

Given: A graph on = vertices with a clique (□ of size :.

Output: A list ! □ □^{3} of size $\$((=/:)^{C})$ that contains an (such that $|(\cap (□| \ge (1-):$.

Operation:

1. Find a degree-$(C) pseudo-distribution on F satisfying the :-clique axioms on A( ) and minimizing k…[F̃]k$_2$.
2. For every & ⊆ [=]^C, an ordered C-tuple on [=] such that …[F$_&$] > 0, let $_&$ = $\frac{\tilde{\ldots}[F_& F]}{\tilde{\ldots}[F_&]}$.
3. For # = $((=/:)^C)$ repetitions, choose an ordered C-tuple & ⊆ [=]^C with probability proportional to …[F$_&$] and add $_&$ to the list ⊆′.
4. For each element $_&$ ∈ ⊆′, construct the set ($_&$ = {8 | $_&$(8) ≥ 1 − 2} and add it to ⊆.
5. Output ⊆.

To analyze this algorithm, we first observe that the maximal coverage property (i.e., minimizing …[F̃]) implies that …[F] has a non-trivial weight on the true (but unknown) :-clique (⊆. This lemma is by now standard with analogous usages in the context of list-decodable learning [KKK19, BK21, IK22]. It can be proven by showing that if $\sum_{8 \notin (⊆} \tilde{\ldots}[F_8] < :^2/=$ then one can take a "mix" of and the distribution that places all its mass on (⊆ (which does satisfy A) and produce another pseudo-distribution ′ with smaller …[F$_8$]$_2$.

Lemma 5.3 (Maximal coverage implies non-trivial weight on (⊆, see Lemma 4.3 in [KKK19])**.** *Let be a pseudo-distribution of degree* ≥ 4 *satisfying* A( ) *that minimizes* …[F] *. Then,* $\sum_{8 \in (⊆} \tilde{\ldots}[F_8] \geq :^2/=$.

As an immediate corollary, we observe the following consequence of our rounding scheme:

Lemma 5.4. *Let be the pseudo-distribution constructed in Step 1 of the algorithm. Then, in Step 3 of the algorithm, each of the chosen* C*-tuples* & *satisfies* & ⊆ ((⊆)^C *with probability at least* (:/=)^C.

Proof. The probability that in Step 3 of the algorithm an ordered C-tuple & is in $(( ) ^C$ is $\tilde{\ldots}[(\sum_{8 \in (⊆} F_8)^C]/:^C$, where we used that $\tilde{\ldots}[\sum_{8=1}^{=} F_8{}^C] = :^C$. The proof now follows by applying Hölder's inequality for pseudo-distributions to conclude that $\tilde{\ldots}[(\sum_{8 \in (⊆} F_8)^C] \geq \tilde{\ldots}[\sum_{8 \in (⊆} F_8]^C \geq (:^2/=)^C$ (from Lemma 5.3).

Next, we argue that for & ⊆ ((⊆)^C chosen with probability proportional to …[F$_&$], with probability at least 0.5, the corresponding (= $_&$ has a non-trivial intersection with (⊆.

Lemma 5.5. *Assume the hypothesis of Theorem 5.1. Then, in Step 4 of the algorithm, conditioned on* & ⊆ ((⊆)^C*, with probability at least* 0.5, $\sum_{8 \in (⊆} $_&$(8) \geq (1 -):$.

Proof. Consider the bipartite graph formed by keeping only the edges that lie in cut((⊆) in with left vertex set equal to (⊆ and right vertex set equal to [=] \ (⊆. Then, observe that A( ) ⊢ ⊆() via the polynomial map G$_D$ = F$_D$ for every D ⊆ (⊆ and H$_E$ = F$_E$ for every E ⊆ [=] \ (⊆. Since satisfies A( ), the polynomial transformation above applied to gives a pseudo-distribution on (G, H) that satisfies ⊆(). From the biclique certificate, we have: …[|G|^C|H|] ≲ $.

This yields that

$$\sum_{8_1,8_2,\dots,8_C} \tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C} |H|] \le \$.$$

Rescaling and rewriting yields

$$\frac{1}{\tilde{..}[|G|^C]} \sum_{8_1,8_2,\dots,8_C : \tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C}]>0} \tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C}] \frac{\tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C} |H|]}{\tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C}]} \le \frac{\$}{\tilde{..}[|G|^C]} .$$

Using Hölder's inequality for pseudo-distributions and Lemma 5.3, we know that $\tilde{..}[|G|^C] = \tilde{..}[(\sum_{8 \in (} F_8)^C] \ge \tilde{..}[\sum_{8 \in (} F_8]^C \ge (:^2/=)^C$. Thus the right-hand side of the above is at most $\$(=/:^2)^C$.

Observe that the left-hand side can be interpreted as the expected value of the random variable $\frac{\tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C} |H|]}{\tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C}]}$ where each $8_1, 8_2, \dots, 8_C$ is chosen with probability equal to $\frac{\tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C}]}{\sum_{8_1,8_2,\dots,8_C} \tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C}]}$.

For an ordered tuple $\& \in (()^C$ chosen with probability proportional to $\tilde{..}[F_\&]$, consider the $(= - :)$-dimensional vector $\frac{\tilde{..}[F_\& H]}{\tilde{..}[F_\&]}$. Its ℓ_1-norm is equal to $\frac{\tilde{..}[F_\& |H|]}{\tilde{..}[F_\&]}$, which is also equal to the random variable $\frac{\tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C} |H|]}{\tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C}]}$ where each $8_1, 8_2, \dots, 8_C$ is chosen with probability equal to $\frac{\tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C}]}{\sum_{8_1,8_2,\dots,8_C} \tilde{..}[G_{8_1} G_{8_2} \cdots G_{8_C}]}$.

Thus, we have concluded that the expected value of the ℓ_1-norm of $\frac{\tilde{..}[F_\& H]}{\tilde{..}[F_\&]}$ is at most $\$(=/:^2)^C$. By Markov's inequality, with probability at least 0.5 over the choice of $\&$, thus, the ℓ_1-norm of $\frac{\tilde{..}[F_\& H]}{\tilde{..}[F_\&]}$ is at most $2\$(=/:^2)^C$. Also note that, by Fact 3.3, $\sum_{8=1}^{=} {}_\&(8) \ge :$. Then, with probability at least 0.5 over the choice of $\&$, $\sum_{8 \in (} {}_\&(8) \ge : - \frac{\tilde{..}[F_\& H]}{\tilde{..}[F_\&]} \ge : - \$(=/:^2)^C \ge (1 - \,):$ using that $\$(=/:^2)^C \le :$.

Proof of Theorem 5.1. From Lemma 5.4, in Step 4, we choose a $\& \in (^C$ with probability at least $(:/=)^C$. Conditioned on this event happening, Lemma 5.5 shows that $\sum_{8 \in (} {}_\&(8) \ge (1 - \,):$ with probability at least 0.5. We call such $\&$ good.

By averaging, for a good $\&$, we must have that for a $(1 - 2)$-fraction of $8 \in ($, ${}_\&(8) \ge 1 - 2$. Further, the total number of coordinates of ${}_\&$ larger than $1 - 2$ cannot be more than $:/(1 - 2)$. Thus, $(_\&$ is a set of size at most $:/(1 - 2)$ such that $|(_\& \cap (| \ge (1 - 2):$.

The $\$((=/:)^C)$ repetitions in Step 3 ensure that with probability at least 0.99 we choose at least one good $\&$.

5.1 Proof of main results

We combine our biclique certificates, the rounding algorithm, and a simple cleanup step to obtain the main results of our work. We start by stating the main result.

Theorem 5.6 (Main result). *Consider a graph on = vertices such that is generated according to* FK$(=, :, ?)$. *Then the following two results hold:*

1. ($=^{2/3}$ guarantee) *For any* > 0 *and* $?, 1 - ? \geq =^{-(1-)}$, *there exists an algorithm that takes input , runs in polynomial time, and for* $: \geq \max\{\$(=^{2/3} ?^{1/3}/(1-?)^{2/3}), \$(=^{1/2})\}$, *with probability* 0.99 *outputs a list of at most* $(1+>(1))=/:$ $:$*-cliques such that one of them is the planted clique in .*

2. ($=^{1/2+}$ guarantee) *For any* > 0 *small enough, there exists an algorithm that takes input , runs in time* $=^{\$(1/)}$, *and for* $: \geq =^{1/2+}/(1-?)^{1/}$, *with probability* 0.99 *outputs a list of at most* $(1+>(1))=/:$ $:$*-cliques such that one of them is the planted clique in .*

Before we prove this, we state and prove two auxiliary lemmas that help us prune the list of subsets returned by the list-decoding algorithm in Theorem 5.1.

Lemma 5.7 (Intersection of cliques with the planted clique)**.** *Let* $\in$ FK$(=, :, ?)$. *Let* $(^{*}$ *be the planted clique in . Then, with probability at least* $1 - \frac{1}{=^2}$, *any other clique* $($ *of size at least* $:$ *satisfies* $|(\cap (^{*}| \leq 3 \frac{\log =}{\log 1/?}$.

Proof. The proof is analogous to that of Proposition 2.2 and is an easy consequence of a Chernoff bound and a union bound.

Lemma 5.8 (Subsets with small intersection)**.** *Let* $(_1, \ldots, (_< \subseteq [=]$ *with* $|(_8| = :$ *and* $|(_8 \cap (_9| \leq \Delta$. *Then, if* $: \geq \sqrt{2=\Delta}$, *we have* $< \leq \frac{=}{:}\left(1 + \frac{2=\Delta}{:^2}\right)$.

Proof. By the inclusion-exclusion principle, we need

$$< : - \frac{<^2}{2}\Delta \leq = .$$

By inspecting the above as a quadratic equation in $<$, we get that for $: \geq \sqrt{2=\Delta}$ the equation is violated when $< > \frac{: - \sqrt{:^2 - 2=\Delta}}{\Delta}$. We note that

$$\frac{: - \sqrt{:^2 - 2=\Delta}}{\Delta} = \frac{:}{\Delta}\left(1 - \sqrt{1 - \frac{2=\Delta}{:^2}}\right) \leq \frac{:}{\Delta}\frac{=\Delta}{:^2}\left(1 + \frac{2=\Delta}{:^2}\right) = \frac{=}{:}\left(1 + \frac{2=\Delta}{:^2}\right)$$

and therefore obtain that $< \leq \frac{=}{:}\left(1 + \frac{2=\Delta}{:^2}\right)$.

We are now ready to prove the main result. The $=^{2/3}$ guarantee uses the first certificate in Theorem 4.11 and produces a result similar to that of [MMT20], and the $=^{1/2+}$ guarantee uses the second certificate in Theorem 4.11 and is the main contribution of our work.

We start by proving the $=^{2/3}$ guarantee.

Proof of $=^{2/3}$ *guarantee in Theorem 5.6.* First, we note that A( ) implies A($'$) for any $'$ that is obtained by adding edges to . Therefore, in our sum-of-squares programs we can ignore the adversarial deletion phase and assume that we work with a graph in which the edges going out from $(^{*}$ are random.

By the first certificate in Theorem 4.11, we have

$$\vdash \sum_{GH \in '} 4_G |H| \leq \$\left(\frac{=?}{1-?}\right).$$

Next, we want to apply Theorem 5.1 with $ = $(=?/(1 − ?)) and = (1 − ?)/24. To apply the theorem, we need $ · (=/:²) ≤ :, which we rewrite as

$$: \geq \$ \frac{= \quad ?^{2/3} \quad ?^{1/3}}{(1 \quad ?)_{2/3}} .$$

Theorem 5.1 yields a list of $(=/:) subsets, each of size at most :/(1 − (1 − ?)/12) ≤ (1 − (1 − ?)/6):, such that with probability at least 0.99 one one them interesects the true clique (□ in at least (1 − (1 − ?)/12): ≥ (1 − (1 − ?)/6): vertices.

To obtain a list that contains (□ exactly, we will remove from each (in the list all vertices that are connected to few vertices in (, and we will add to (all vertices that are connected to many vertices in (. Formally, we will make use of the following claim:

Claim 5.9. With probability at least 0.99, for all subsets (□ [=] with |(∩ (□| ≥ (1 −): and |(| ≤ (1 +):, every vertex E □ (□ is connected to at least (1 −): − 1 vertices in (and every vertex E □ (□ is connected to at most :? + $($\sqrt{:?(1 - ?)\log =)}$ + 2: vertices in (.

Proof of claim. The first claim is trivial: every vertex E □ (□ has at least |(∩ (□| − 1 ≥ (1 −): − 1 edges to (.

For the second claim, we begin by noting that, for a vertex E □ (□, the number of edges to (□ is at most a binomial random variable Bin(:, ?). By standard bounds, this is larger than :? + C with probability at most $\min\{4^{-C^2/(2:(1-?))}, 4^{-C^2/(2:?+2C/3)}\}$. Then, by a union bound, the probability that the number of edges is larger than :? + C for any E □ (□ is at most = $\min\{4^{-C^2/(2:(1-?))}, 4^{-C^2/(2:?+2C/3)}\}$. Choosing C = $ $\sqrt{:?(1 - ?)\log =}$ makes this probability a small constant. Then, with probability a ea e e l a t l st 0.99, no v rt x E □ (□ has more than :? + $ $\sqrt{:?(1 - ?)\log =}$ dg s to (□. n ddition, a vertex E □ (□ has at most $\sqrt{|(\setminus (□|}$ ≤ (1 +): − (1 −): = 2: edges to (\ (□. Therefore, overall, it has at most :? + $($\sqrt{:?(1 - ?)\log =)}$ + 2: edges to (□.

Consider the subset (in the list for which |(∩ (□| ≥ (1 − (1 − ?)/6). We can apply the claim to this subset with = (1 − ?)/6. Then, every vertex E □ (□ is connected to at least $(1 - (1 - ?)/6): - 1 - \sqrt{:?(1 - ?)\log =)}$ vertices in (, and every vertex E □ (□ is connected to at most (? + (1 − ?)/3): + $($\sqrt{:?(1 - ?)\log =)}$ < (1 − (1 − ?)/6): − 1 vertices in (, where we used that : > $ (log =)?/(1 − ?) .

Therefore, we do the following: for each subset (in the list, we remove from (all vertices that are connected to less than (1 − (1 − ?)/6): − 1 of the vertices in (, and we add to (all vertices that are connected to at least (1 − (1 − ?)/6): − 1 of the vertices in (. This ensures that the subset (for which |(∩ (□| ≥ (1 − (1 − ?)/6) is transformed by this procedure into (□.

After that, we remove from the list the subsets with size different than : and the subsets that are not cliques. Then we iterate the following procedure: find (, (′ in the list such that |(∩ (′| ≥ $(log =/log 1/?) and remove one of them from the list. By Lemma 5.8, the resulting list has size at most (1 + >(1))=/:, where we use that our choice of : satisfies : ≥ $\sqrt{2=\$(\log =/\log 1/?)}$. Furthermore, by Lemma 5.7, this procedure cannot remove (□ from the list, because it intersects other cliques in at most $(log =/log 1/?) vertices.

We note that : ≥ max{$(=^{2/3}?^{1/3}/(1 − ?)^{2/3}), $′(=^{1/2})} satisfies the lower bounds on : that we require. The time complexity of the algorithm is polynomial in =.

Finally, we prove the $=^{1/2+}$guarantee, which we split into the cases ? ≤ 1/2 and ? ≥ 1/2.

Lemma 5.10 ($=^{1/2+}$ guarantee of Theorem 5.6, ? ≤ 1/2)**.** *Fix any > 0 small enough. There is an algorithm that takes input a graph on = vertices, runs in time* $=^{\$(1/)}$*, and provides the following guarantee: If is generated according to* FK(=, :, ?) *with* ? ≤ 1/2*, for* : ≥ $=^{1/2+}$*, with probability* 0.99 *the algorithm outputs a list of at most* (1 + >(1))=/: :*-cliques such that one of them is the planted clique in .*

Proof. First, we note that A() implies A($'$) for any $'$ that is obtained by adding edges to . Therefore, in our sum-of-squares programs we can ignore the adversarial deletion phase and assume that we work with a graph in which the edges going out from (⋆ are random.

For ? < 1/2, the second certificate in Theorem 4.11 is the same up to constant factors as the one in Theorem 4.2 for ? = 1/2. Furthermore, for ? < 1/2, the range of : for which the second certificate in Theorem 4.11 holds is a superset of the range of : under which the one in Theorem 4.2 holds. Therefore, in this proof, we assume without loss of generality that ? = 1/2, noting that all the steps in the proof continue to be valid even if ? < 1/2.

By Theorem 4.2, for : ≥ \$($\sqrt{\text{C=log =}}$), we have

$$⋆() \;{}_{4C+2}\Big|^{G,H} |G|^{4C}|H| \leq (1000C)^{10C} = \left(\frac{=}{:}\right)^4 .$$

Next, we want to apply Theorem 5.1 with \$ = $(1000C)^{10C} = (=/:)^4$ and = 1/(4:). The choice of ensures that each subset of the returned list has size at most :. To apply the theorem, we need \$ · $(=/:^2)^C$ ≤ :, which we rewrite as

$$: \geq \text{poly}(C) \cdot \sqrt{=} \cdot =^{\frac{3}{8C+4}} .$$

Theorem 5.1 yields a list of \$($(=/:)^{4C}$) subsets, each of size at most :, such that with probability at least 0.99 the true clique (⋆ is in the list.

Next, we remove from the list the subsets with size different than : and the subsets that are not cliques. Then we iterate the following procedure: find (, ($'$ in the list such that |(∩ ($'$| ≥ \$(log =) and remove one of them from the list. By Lemma 5.8, the resulting list has size at most (1+>(1))=/:, where we use that our choice of : satisfies : ≥ $\sqrt{2=\$(\log =)}$. Furthermore, by Lemma 5.7, this procedure cannot remove (⋆ from the list, because it intersects other cliques in at most \$(log =) vertices.

We choose the smallest C such that ≥ $\frac{3+0.1}{8C+4}$, which is C= ⌈$\frac{3}{8}$ − $\frac{1}{2}$⌉ = **\$**(1/). Then : ≥ $=^{1/2+}$ satisfies the lower bounds on : that we required in the proof. Finally, the time complexity of the algorithm is $=^{\$(C)} = =^{\$(1/)}$.

Lemma 5.11 ($=^{1/2+}$ guarantee of Theorem 5.6, ? ≥ 1/2)**.** *Fix any > 0 small enough. There is an algorithm that takes input a graph on = vertices, runs in time* $=^{\$(1/)}$*, and provides the following guarantee: If is generated according to* FK(=, :, ?) *with* ? ≥ 1/2*, for* : ≥ $=^{1/2+}/(1-?)^{1/}$*, with probability* 0.99 *the algorithm outputs a list of at most* (1 + >(1))=/: :*-cliques such that one of them is the planted clique in .*

Proof. First, we note that A( ) implies A($'$) for any $'$ that is obtained by adding edges to . Therefore, in our sum-of-squares programs we can ignore the adversarial deletion phase and assume that we work with a graph in which the edges going out from (are random.

By the second certificate in Theorem 4.11, for : ≥ \$ ($\sqrt{C= \log =}$?^{2C}/(1 − ?)$_2^{C+1}$), we have

$$() \; 4C \frac{G,H}{+2} |G|^{4C}|H| \leq (1000C)^{10C} = \left(\frac{= ?^{2C}}{:(1-?)^{2C+1}} \right)^4 .$$

Next, we want to apply Theorem 5.1 with \$ = $(1000C)^{10C} = \left(\frac{=?^{2C}}{:(1-?)^{2C+1}}\right)^4$ and = 1/(4:). The choice of ensures that each subset of the returned list has size at most :. To apply the theorem, we need \$ · $(=/:^2)^C$ ≤ :, which we rewrite as

$$: \geq \text{poly}(C) \cdot \sqrt{=} \cdot =^{\frac{3}{8C+4}} ?^{1-\frac{1}{8C+4}} /(1-?) .$$

Theorem 5.1 yields a list of \$($(=/:)^{4C}$) subsets, each of size at most :, such that with probability at least 0.99 the true clique (is in the list.

Next, we remove from the list the subsets with size different than : and the subsets that are not cliques. Then we iterate the following procedure: find (, ($'$ in the list such that |(∩ ($'$| ≥ \$(log =/(1 − ?)) and remove one of them from the list. By Lemma 5.8, the resulting list has size at most (1 + >(1))=/:, where we use that our choice of : satisfies : ≥ $\sqrt{2=\$(\log =/(1-?))}$. Furthermore, by Lemma 5.7, this procedure cannot remove (from the list, because it intersects other cliques in at most \$(log =/log 1/?) = \$(log =/(1 − ?)) vertices, where we used that log 1/? = Ω(1 − ?) for ? ≥ 1/2.

We choose the smallest C such that ≥ $\frac{3+0.1}{8C+4}$ which is C = $\lceil \frac{31}{80} - \frac{1}{2} \rceil$ = \$(1/). For this choice of C, we actually have $(1-?)^{2C+1} \geq (1-?)^{1/}$ for ≤ 0.1. Then : ≥ $=^{1/2+}/(1-?)^{1/}$ satisfies the lower bounds on : that we require. Finally, the time complexity of the algorithm is $=^{\$(C)} = =^{\$(1/)}$.

Proof of second guarantee in Theorem 5.6. By Lemma 5.10 we obtain the desired result for ? ≤ 1/2 when : ≥ $=^{1/2+}$, and by Lemma 5.11 we obtain the desired result for ? ≥ 1/2 when : ≥ $=^{1/2+}/(1-?)^{1/}$. Then both results hold when : ≥ $=^{1/2+}/(1-?)^{1/}$.

6 Evidence of hardness for certifying blicliques

In this section, we collect some evidence that suggests that improving on our guarantees for the unbalanced bipartite clique certification problem is hard. Our hardness results are in two settings: in the first we will prove a lower bound on the basic SDP relaxation for the problem of finding large bicliques in random graphs that gives a concrete reason for the $=^{2/3}$ barrier (for ? = 1/2) in prior works, and in the second we will prove lower bounds in the low-degree polynomial model for hypothesis testing problems.

6.1 Lower bounds against basic SDP

We consider the following SDP relaxation for finding large bicliques in a given bipartite graph = (*, +,) where |*| = : and |+| = =. It is equivalent to the degree 2 sum-of-squares relaxation of the biclique constraint system (3).

$$
\begin{cases}
\forall 8, 9 & 0 \leq -(8, 9) \leq 1 \\
& \mathrm{tr}(-) = : \\
& \sum_{D \in *} -(D, D) = \ell \\
& \sum_{E \in +} -(E, E) = : - \ell \\
& \sum_{D \in *, E \in +} -(D, E) = \ell(: - \ell) \\
\forall D \in *, E \in + \text{ s.t. } \{D, E\} \notin & -(D, E) = 0 \\
& - \succeq 0
\end{cases}
\tag{7}
$$

Commentary on the SDP relaxation. We think of the SDP solution - as a matrix indexed by all the (left and right) vertices of the bipartite graph . We associate the first : = |*| rows and columns of - with the left vertices and the last = = |+| rows and columns of - with the right vertices of . If G ∈ $\{0,1\}^{|*|}$ and H ∈ $\{0,1\}^{|+|}$ indicate left and right subsets of vertices in a purported biclique of total size :, then - should be "thought of" as a relaxation of the constraints satisfied by the rank 1 matrix $(G, H)(G, H)^{\top}$. In particular, the first two constraints posit that - is non-negative in all its entries and that its trace (that equals $\|G\|_2^2 + \|H\|_2^2$ and thus the total size of the biclique) is :. The next three constraints posit that the left hand side vertices contribute ℓ to the trace (corresponding to the left hand side contributing ℓ vertices to the biclique), that the right hand side vertices contribute : − ℓ to the trace, and that $(\sum_8 G_8)(\sum_8 H_8) = \ell(: - \ell)$. The penultimate constraint posits that if D ∈ * and E ∈ + do not have an edge between them, then we cannot simultanesouly pick D, E to be in the biclique (capturing the "biclique" constraints).

For fixed :, =, the infeasibility of the SDP for some ℓ = ℓ(:, =) is equivalent to there being a degree 2 sum-of-squares certificate of the absence of ℓ × : − ℓ bicliques in . We will show that the above SDP is in fact *feasible* whp over the draw of so long as ℓ ≲ =/:. This corresponds to the basic SDP barrier at : = $=^{2/3}$ (a threshold obtained by balancing the above obtained trade-off – see Remark 2.5) encountered in prior works on the semi-random planted clique problem.

Lemma 6.1 (SDP lower bound for biclique certification). *Let = (*, +,) ∼ (:, =, 1/2) be a bipartite Erdős-Rényi random graph with edge probability* 1/2. *Then, with probability at least* 0.99 *over the draw of , for any* 100 = ≤ : ≤ $=/\sqrt{2}$ *and* ℓ ≤ 2=/: *for some constant* 2 > 0 *small enough, the SDP* (7) *is feasible.*

Proof. We will prove the lemma by exhibiting an explicit solution to the SDP (7). The unbalanced

setting requires a slightly more involved construction compared to the related SDP lower bounds for the clique number of (=, 1/2), where a natural shifted and scaled adjacency matrix yields a feasible solution.

The construction. In order to describe our construction, it is helpful to think of the solution - as being divided into $-_{C>?}$, the principal : × : block corresponding to first : rows and columns, $-_{1>C}$, the principal = × = block corresponding to the last = rows and columns, and $-_{2A>BB}$, the : × = off-diagonal block (and its transposed copy).

We will set every diagonal entry of $-_{C>?}$ to be ℓ/: and every off-diagonal entry of $-_{C>?}$ to be $(ℓ/:)^2$. Informally, $-_{C>?}$ is the 2nd moment matrix of the probability distribution that chooses every vertex on the left with probability ℓ/: independently.

We describe $-_{2A>BB}$ next. For every D ⍰ *, E ⍰ +, we set -(D, E) = -(E, D) = 0 if D is not connected to E in , and otherwise we set -(D, E) = -(E, D) = 2_1(ℓ/=) for some constant 2_1 > 0 to be chosen later. Notice that this is equivalent to setting $-_{2A>BB}$ = $2_1 \frac{ℓ}{=}$ where is the : by = bipartite adjacency matrix of .

Finally, we describe $-_{1>C}$. This is where we need to be a little more careful. Let $0_1, 0_2, \ldots, 0_:$ be =-dimensional vectors in $\{-1, 1\}^=$ such that $0_D(E)$ = 1 iff {D, E} is an edge in . That is, the 0_8s are the ±1 neighborhood indicators of the : left vertices in . Then, we set $-_{1>C}$ = $\frac{:+1}{=(\;)}$ ($\frac{:-ℓ}{8\,1}$ $0_8 0^⍰$ + **11**$^⍰$). Here **1** is the vector of all 1 coordinates. Note that the diagonal entries of $-_{1>C}$ exactly equal (: − ℓ)/= and thus tr($-_{1>C}$) = : − ℓ as required. Also note that $-_{1>C}$ is low rank, as it has rank at most : + 1.

We discuss now the choice of 2_1. We want to enforce $\sum_{D⍰*,E⍰+}$ -(D, E) = ℓ(: − ℓ), so we choose 2_1 = $\frac{ℓ}{(/=)}$ $\frac{ℓ(\;-ℓ)}{\sum_{D⍰*,E⍰+}(D,E)}$ = $\frac{-ℓ\;)=}{\sum_{D⍰*,E⍰+}(D,E)}$. Note that with probability at least 0.999 we have that Ω(:=) ≤ $\sum_{D⍰*,E⍰+}$ (D, E) ≤ :=, so with probability at least 0.999 we have that 2_1 is bounded below and above by absolute constants.

Analysis. With probability at least 0.999 over the draw of , - immediately satisfies all the constraints except for positive semidefiniteness. We focus next on verifying the PSD-ness of -. Consider any "test" vector I ⍰ ⍰$^{:+=}$, which we will think of as ($I_!$, $I_'$) where $I_!$ is the projection of I to the first : coordinates (i.e., the left vertices) and $I_'$ the projection to the last = coordinates (i.e., the right vertices).

Now,

$$I^⍰ - I = I_!^⍰ -_{C>?} I_! + I_'^⍰ -_{1>C} I_' + 2 I_!^⍰ -_{2A>BB} I_' . \tag{8}$$

Let be the subspace of at most : + 1 dimensions spanned by the : rows of and the all 1s vector **1**. Now, notice that $-_{2A>BB} I_'$ = $-_{2A>BB} I_,$ where $I_,$ is the projection of $I_'$ to . Similarly, by design, $-_{1>C}$ has range space equal to , so $-_{1>C} I_'$ = $-_{1>C} I_,$. Thus, WLOG, we can assume that $I_'$ = $I_,$ in the following.

Let's write $I_!$ = $I_!^k$ + $I_!'$ and $I_'$ = $I_'^k$ + $I_'^{\prime}$ where $I_!^k$= h$I_!$, v$\frac{1}{=}$iv$\frac{1}{=}$is the component of $I_!$ along the all 1s direction (and similarly for $I_'^k$). Let $-'_{2A>BB}$ = $2_1\frac{ℓ}{=}$(− $\frac{ℓ}{2}$**11**$^⍰$) be the "centered" version of $-_{2A>BB}$. Also define the centered versions $-'_{C>?}$ = $-_{C>?}$ − $\frac{ℓ^2}{?}$**11**$^⍰$and $-_{1>}^{C'}$= $\frac{1-ℓ}{=(:+\;)}$ $\sum_{8=1}^{:}$ $0_8 0_8^⍰$.

Our argument is to simply "charge" the third term (which can be potentially negative) to the first two terms (that are always non-negative). We will use the following two standard random matrix facts (see Fact 3.15) in our analysis: for ′ = − $_2$ **11**⊺ we have k′k$_2$ ≤ $ (√=), and the :-th smallest singular value of both and ′ is at least Ω(√= − √:) = Ω(√=) as : ≤ =/2.

The potentially negative terms. Let's work with the potentially negative terms coming from the parallel components of I$_!$ and I$_'$. Observe that (I$^k_!$)⊺ -2A>BB I$^k_'$ = I$^k_!$ I$^k_2$ ^{2}1 ℓ + (I$^k_!$)⊺ -2A>BB I$^{k'}$ ≥ I$^k_!$ I$^k_2$ ' $_2$ ($\frac{2\ell}{2}$ − $ ($\frac{\ell}{=}$ √=)) > 0. Since (I$^k_!$)⊺ -C>? I$^k_!$ + (I$^k_'$)⊺ -1>C I^k > 0 we can conclude that (I^k)⊺ - I^k ≥ 0.

Let's analyze the potentially negative terms coming from the perpendicular components of I$_!$ and I$_'$. We have |(I$^!$)⊺ -2A>BB I$_'$′| = |(I$^!$)⊺ -2A>BB I$_'$′| ≤ 2$_2$ $\frac{\ell}{=}$ √= kI′$_!$k$_2$ kI′$_'$k$_2$ = 2$_2$ $\frac{\ell}{\sqrt{=}}$ kI′$_!$k$_2$ kI′$_'$ k2.

Finally, let's analyze the potentially negative terms coming from crossing the parallel and the perpendicular components of I$_!$ and I$_'$. We have: |(I$^k_!$)⊺ -2A>BB I′$_'$| = |(I$^k_!$)⊺ -2A>BB I$_'$′| ≤ kI$^k_!$k$_2$ kI′$_'$ k$_2$ $ (ℓ/ √=). Similarly, |(I$^!$)⊺ -2A>BB I$^k_'$| ≤ kI′$_!$k$_2$ kI$^k_'$ k$_2$ $ (ℓ/ √=).

The square terms. We now compute a lower bound on the non-negative terms in (8).

We have I⊺ -C>? I$_!$ = I⊺ -!'>? I$_!$ + $^{\ell}_2$ (I$^!$)⊺ **11**⊺ I^k = kI$_!$k^2 ($^{\ell}$ $\frac{}{2}$ − $\frac{\ell}{: 2}$) + $\frac{2}{:}$ kI$_!$k^2 k^2 $\frac{}{=2}$ ≥ $\frac{}{: 2}$ 2$_3$ kI$_!$k^2 $\frac{\ell}{2:}$.

Next, we lower bound I⊺$_'$ -1>C I$_'$. Now, I$_'$ ⊺. Recall that the :-th smallest singular value of and ′ is at least Ω(√=) if : ≤ =/2 with probability at least 0.999 over the draw of the graph , and 2) the matrix $\int_\varepsilon$ $_8$ 0$_8$0⊺$_8$ has the same eigenvalues as 4′′⊺ where, recall that ′ = − $\frac{1}{2}$ **11**⊺. Together this yields that all eigenvalues of $\int$ $_8$ 0$_8$0⊺$_8$ + **11**⊺ are at least 2$_4$= for some constant 2$_4$ > 0 when restricted to the subspace . Thus, I$_'$⊺ -1>C I$_'$ ≥ 2$_4$= kI$_'$k$^2_2$ $\frac{:-\ell 1}{=(:+)}$ = 2$_4$ kI$_'$k$^2_2$ $\frac{:-\ell}{+1}$ ≥ 2$_5$ kI$_'$k$^2_2$ recalling that : − ℓ > :/2.

Let's now complete the charging argument. Let's first observe that, by the AM-GM inequality, the square terms contribute at least 2$_6$ $\sqrt{\ell/:}$ kI$_!$k$_2$ kI$_'$k2. The potentially negative term from the perpendicular components is at most 2^{2}ℓ/ √= kI′$_!$k$_2$ kI′$_'$k$_2$ in magnitude, and the potentially negative term from crossing the components is at most kI$^k_!$k$_2$ kI$_'$′ k$_2$ $ (ℓ/ √=) + kI′$_!$k$_2$ kI$^k_'$ k$_2$ $ (ℓ/ √=).

Thus, the square terms dominate as long as ℓ ≤ $ (=/:).

This completes the proof.

6.2 Low-degree lower bound for ? = 1/2

Formally, we will prove that there are distributions over bipartite graphs that admit ℓ by : − ℓ cliques for appropriate parameters ℓ that are indistinguishable from (:, =, ?) — the distribution on random bipartite graphs with left vertex set of size :, right vertex set of size = and each bipartite edge included to be in the graph with probability ? independently. The choice of the planted model requires a bit of care, as we soon discuss. We will deal with the case of ? = 1/2 and general ? separately for clarity of exposition.

- $\mathcal{D}_{\text{null}} = (n, m, 1/2)$: the distribution on bipartite graphs $G = (L, R, E)$ where $|L| = n$, $|R| = m$ and each bipartite edge $\{u, v\}$ with $u \in L$ and $v \in R$ is included in G with probability 1/2.

- $\mathcal{D}_{\text{planted}} = (n, m, k, 1/2)$: the distribution on bipartite graphs $G = (L, R, E)$ where $|L| = n$, $|R| = m$, sampled as follows. Choose S by including each vertex from L in S with probability k/n. Choose T by including every vertex from R in T with probability $(n - k)/m$. Finally, include each edge $\{u, v\}$ with $u \in L$ and $v \in R$ with probability

$$\Pr_{\mathcal{D}_{\text{planted}}}[\{u, v\} \text{ is included}] = \begin{cases} 1 & \text{if } u \in S, v \in T, \\ \frac{m/2 - (m-k)}{m-(n-k)} & \text{if } u \in S, v \notin T, \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

Remark 6.2. $\mathcal{D}_{\text{planted}}$ is chosen so as to have a k by $n - k$ biclique in it while having the same distribution of degrees of *left* vertices as in $\mathcal{D}_{\text{null}}$. This is necessary since otherwise the average degree of the left vertices gives a distinguisher between the models.

Theorem 6.3. *Fix $\varepsilon > 0$ independent of m with $\varepsilon \le 0.001$. For $n = m^{1/2+\varepsilon}$ and $k \le m^{1/4-0.001}$, the norm of the degree-$\lfloor 0.001/\varepsilon \rfloor$ likelihood ratio between $(n, m, k, 1/2)$ and $(n, m, 1/2)$ is $1 + \Omega(1)$. On the other hand, for $n = m^{1/2+\varepsilon}$ and all k, the norm of the degree-$\Omega(1/\varepsilon)$ likelihood ratio between $(n, m, k, 1/2)$ and $(n, m, 1/2)$ is unbounded as $m \to \infty$.*

Remark 6.4. Information-theoretically, to identify a small list in the semi-random planted clique model, we need $n = \tilde{\Theta}(\sqrt{m})$ [Ste17]. If we set n to be this value, then the corresponding bipartite random graph has no k by $n - k$ clique for $k = \Omega(\log m)$. The above theorem shows that in the low-degree polynomial model, distinguishing between the case when $k = m$ vs $k = \Omega(\log m)$ requires polynomials of degree $\Omega(1/\varepsilon)$.

For a bipartite graph $G = (L, R, E)$ recall that $x_{u,v}$ is 1 if the edge $\{u, v\}$ is included and -1 otherwise. We also define $x^S = \prod_{\{u,v\} \in S} x_{u,v}$.

Lemma 6.5. *For G sampled from $\mathcal{D}_{\text{planted}}$, let n be the number of left vertices in G, m the number of right vertices in G, and $d_1, \ldots, d_m$ the number of edges in G incident to each of the right vertices. Then*

$$\mathbb{E}_{\mathcal{D}_{\text{planted}}}[x^S] = \begin{cases} \frac{k}{n} \cdot \frac{k-1}{n-1} \cdots \frac{k-|S|+1}{n-|S|+1} \cdot \prod_{j=1}^{m} \left(1 + \varepsilon \frac{k-|S|}{n-|S|}\right)^{d_j} & \text{if } d_1, \ldots, d_m > |S|, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Conditioned on the planted biclique (S, T), the edges are independent. For an edge $\{u, v\}$, we calculate

$$\mathbb{E}[x_{u,v} \mid \text{planted biclique is } (S, T)] = \begin{cases} 1 & \text{if } u \in S, v \in T, \\ \frac{\mathbb{E}_{\mathcal{D}_{\text{planted}}}}{m-(n-k)} & \text{if } u \in S, v \notin T, \\ 0 & \text{otherwise,} \end{cases}$$

where for the case D (, E %, we calculated the expectation as

$$\frac{=/2-(:-\ell)}{=-(:-\ell)}\cdot 1+\left(1-\frac{=/2-(:-\ell)}{=-(:-\ell)}\right)\cdot(-1)=\frac{-(:-\ell)}{=-(:-\ell)}.$$

We observe that if any of the left vertices in is not in the planted biclique, the conditional expectation of " is zero. Therefore, we condition on the event that all the left vertices in are in the planted biclique, which happens with probability $\frac{!}{:}\cdot\frac{\ell}{}$.

Conditioned on this event, for any particular right vertex, all the edges in incident to it are independent from the other edges in . Let $_8$ be the subset of edges in that are incident to the 8-th right vertex. Then

$$\begin{aligned}
&\cdots_{\text{planted}}[\text{"}_8 \mid \text{planted biclique contains all left vertices in }]:-\ell\\
&=\frac{}{=}\cdot 1+\left(1-\frac{:-\ell}{=}\right)\cdot\frac{-(:-\ell)}{=-(:-\ell)}^{3_8}\\
&=\begin{cases}\left(\frac{:-\ell}{:}\right)^{\ell}\left(1+\$\frac{:=\ell}{}\right) & \text{if } 3_8>1,\\ 0 & \text{if } 3_8=1,\end{cases}
\end{aligned}$$

so

$$\begin{aligned}
&\cdots_{\text{planted}}[\text{" }\mid \text{planted biclique contains all left vertices in }]\\
&=\begin{cases}\left(\frac{:-\ell}{=}\right)^{'}\left(1+\$\frac{:-\ell}{=}^{\ell}\right)^{'} & \text{if } 3_1,\ldots,3_{'}>1,\\ 0 & \text{otherwise.}\end{cases}
\end{aligned}$$

Therefore, overall,

$$\cdots_{\text{planted}}[\text{"}]=\begin{cases}\left(\frac{\ell}{:}\right)^{!}\left(\frac{:-\ell}{=}\right)^{'}\left(1+\$\frac{:-\ell}{=}\right)^{'} & \text{if } 3_1,\ldots,3_{'}>1,\\ 0 & \text{otherwise.}\end{cases}$$

Proof of Theorem 6.3. Let $!^{'\le}$ be the degree- likelihood ratio between $_{\text{planted}}$ and $_{\text{null}}$. Then, by standard results, $!^{'\le}-1^2=\sum_{0<|\|}\cdots_{\text{planted}}[\text{"}]^2$, where the norm is the one induced by $_{\text{null}}$. Therefore, if the right-hand side is >(1), then $!^{'\le}$ is 1 + >(1), and if the right-hand side is unbounded, then $!^{'\le}$ is also unbounded.

Consider all with ! left vertices and ' right vertices. The contribution from these is, by Lemma 6.5,

$$\cdots_{\text{planted}}[\text{"}]=2\left(\frac{\ell}{:}\right)^{2!}\left(\frac{:-\ell}{=}\right)^{2'}\left(1+\$\frac{:-\ell}{=}\right)^{2'}\cdot\binom{:}{!}\binom{=}{'}=\text{Bip}(!,')\,,$$

' ! left vertices right vertices

where Bip(!, ') is the number of bipartite graphs with ! left vertices and ' right vertices such that all left degrees are at least 1 and all right degrees are greater than 1.

Consider a choice of ! and ' such that Bip(!, ') ≠ 0. Because we are interested in the behavior of the sum as = goes to infinity, we ignore as negligible all factors that depend only on ! and ' . We also approximate : − ℓ ≈ : and 1 + $ $\frac{:-ℓ}{=}$ $^{2'}$ ≈ 1. Then we have

$$Õ \;\; ··_{\text{planted}}["] \;\; ⍰ \;\; : \;\; 2 \;\; \frac{ℓ}{=}^{2!} \;\; \frac{:}{-}^{2'} ! \;\; ' \;\; : \;\; =$$

' ! right vertices left vertices

$$⍰ \;\; \frac{ℓ}{:}^{2!} \;\; \frac{:}{=}^{2'} \;\; :^{!} =^{'}$$

$$= =_{\_}^{'} :^{2'}{}_{\_}^{!} ℓ^{2!} .$$

For : = $=^{1/2+}$, the above is equal to $=^{(2'\_!)\_!/2} ℓ_2{}^{!}$.

For ℓ = $=^{1/4-0.001}$, this is equal to $=^{(2'-!)-0.00}{}_2{}^{!}$. For || ≤ 0.001/, we have 1 ≤ !, ' ≤ 0.001/ and hence 2 ' − ! ≤ 0.002/ − 1. Then $=^{(}{}_2{}^{'-!)-0.00}{}_2{}^{!}$ ≤ $=^{-}$, which goes to zero as = goes to infinity. Therefore, the sum of all the terms with || ≤ 0.001/ is >(1).

For || ≥ $(1/), consider the term corresponding to ! = 2 and some ' = $(1/). Note that the term satisfies Bip(!, ') ≠ 0 (e.g., the complete bipartite graph on 2 left vertices and $(1/) right vertices is a valid choice). For this term, $=^{(2'\_!)-!/2}$ = $=^{2'-2-1}$ ≥ = for ' = $(1/) large enough. Therefore, this term goes to infinity as = goes to infinity, and then the same is true for the sum of all the terms.

6.3 Low-degree lower bound for general densities

In this section, we will prove that the following two distributions on bipartite random graphs are indistinguishable by low-degree polynomials.

- $_{\text{null}}$ = (:, =, ?): the distribution on bipartite graphs = (* , + ,) where |*| = :, |+| = =, and each bipartite edge {D, E} with D ⍰ * and E ⍰ + is included in with probability ?.
- $_{\text{planted}}$ = (:, =, ℓ, ?): the distribution on bipartite graphs = (* , + ,) where |*| = :, |+| = =, sampled as follows. Choose (by including each vertex from * in (with probability ℓ/:. Choose % by including every vertex from + in ' with probability (: − ℓ)/=. Finally, include each edge {D, E} with D ⍰ * and E ⍰ + with probability

$$\Pr_{\text{planted}}[\{D, E\} \text{ is included}] = \begin{cases} 1 & \text{if } D ⍰ (, E ⍰ \%, \\ \frac{=?-(\;\;ℓ)}{=-(:-ℓ)} \;\; :- & \text{if } D ⍰ (, E ⍰ \%, \\ ? & \text{otherwise}. \end{cases}$$

Theorem 6.6. *Fix > 0 independent of =. Let ? ≥ 1/2 and @= 1 − ?, and define such that @= $=^{-}$. For ≤ /2 and : = $=_1{}^{/}{}_2{}^{+}/@_1{}^{/}{}_2$ and ℓ ≤ $=^{1/4-0.001}$, the norm of the degree-* 5 (1/) *likelihood ratio between*

$(:, =, \ell, ?)$ *and* $(:, =, ?)$ *is* $1 + >(1)$*, for any function* 5 *independent of* $=$*. On the other hand, for* $\geq$ *and* $: = =_1^{/2}{}_+/@_1^{/2}$ *and all* ℓ*, the norm of the degree-*$\$(1/)$ *likelihood ratio between* $(:, =, \ell, ?)$ *and* $(:, =, ?)$ *is unbounded as* $= \to \infty$*.*

Remark 6.7. Information-theoretically, to identify a small list in the semi-random planted clique model with a general $?$, we need $: \geq \tilde{\Theta}(\sqrt{=/@})$ [Ste17]. If we set $:$ to be this allegedly optimal value, then the corresponding bipartite random graph has no ℓ by $: - \ell$ clique for $\ell = \$(\log =)$. The above theorem shows that in the low-degree polynomial model, distinguishing between the case when $\ell = =$ vs $\ell = \$(\log =)$ requires polynomials of degree growing faster than any function (independent of $=$) of $1/.4$

In this section, for a bipartite graph $= (*, +,)$, we define $"_{D,E}$ to be $\sqrt{\frac{1-?}{?}}$ if the edge $\{D, E\}$ is included and $-\sqrt{\frac{?}{1-?}}$ otherwise. We also define $" = \prod_{\{D,E\} \subseteq } "_{D,E}$.

Lemma 6.8. *For* $$ *sampled from* $_{\text{planted}}$*, let* $!$ *be the number of left vertices in* $,$ $'$ *the number of right vertices in* $,$ *and* $3_1, \ldots, 3_{'}$ *the number of edges in* $$ *incident to each of the right vertices. Then*

$$\cdots_{\text{planted}}[" \;] = \begin{cases} \dfrac{\binom{'}{:}}{\binom{=}{:}} \prod_{8=1}^{'} \dfrac{\binom{!}{3_8}}{\binom{=}{3_8}} \left(1 + \$\left(\dfrac{:^2}{=^2}\right)\right)^{\frac{-}{2}} \left(1 - \dfrac{:^2}{=^2}\right)^{\frac{-}{2}} & \text{if } 3_1, \ldots, 3_{'} > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Conditioned on the planted biclique $((, \%)$, the edges are independent. For an edge $\{D, E\}$, we calculate

$$\cdots_{\text{planted}}["_{D,E} \mid \text{planted biclique is } ((, \%)] = \begin{cases} \sqrt{\frac{1-?}{?}} & \text{if } D \in (, E \in \%, \\ -\frac{:^2}{=^2}\sqrt{\frac{?}{1-?}} & \text{if } D \in (, E \notin \%, \\ 0 & \text{otherwise,} \end{cases}$$

where for the case $D \in (, E \notin \%$, we calculated the expectation as

$$\frac{=? - (: - \ell)}{= - (: - \ell)} \cdot \sqrt{\frac{1-?}{?}} + \left(1 - \frac{=? - (: - \ell)}{= - (: - \ell)}\right) \cdot \left(-\sqrt{\frac{?}{1-?}}\right) = \frac{-(: - \ell)}{= - (: - \ell)} \sqrt{\frac{1-?}{?}}.$$

We observe that if any of the left vertices in $$ is not in the planted biclique, the conditional expectation of $"$ is zero. Therefore, we condition on the event that all the left vertices in $$ are in the planted biclique, which happens with probability $\frac{\binom{=}{:}}{\binom{!}{:}}$.

Conditioned on this event, for any particular right vertex, all the edges in $$ incident to it are independent from the other edges in $$. Let $_8$ be the subset of edges in $$ that are incident to the 8-th right vertex. Then

$$\cdots_{\text{planted}}["_8 \mid \text{planted biclique contains all left vertices in }]$$

[4] The theorem leaves open the possibility of distinguishing with low degree in the case $> /2$. However, if $> /2$, then $: \geq \sqrt{=/@}$, which is also suboptimal.

$$= \frac{:-\ell}{=}\cdot\sqrt{\frac{1-?}{?}}^{3_8} + 1-\frac{:-\ell}{=}\cdot\sqrt{\frac{1-?}{?}}^{3_8}$$

$$= \begin{cases} \frac{:-\ell}{=}\sqrt{\frac{1-?}{?}} + \$\sqrt{\frac{:-\ell}{=}\frac{1-?}{?}} & \text{if } 3_8 > 1, \\ 0 & \text{if } 3_8 = 1, \end{cases}$$

so

$$\cdots_{\text{planted}}[" \mid \text{planted biclique contains all left vertices in }]$$

$$= \begin{cases} \frac{:-\ell}{=}\prod_{8=1}^{1'}\sqrt{\frac{1-?}{?}}^{3_8} + \$\sqrt{\frac{:-\ell}{=}\frac{1-?}{?}} & \text{if } 3_1, \ldots, 3_{1'} > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, overall,

$$\cdots_{\text{planted}}[\text{"}] = \begin{cases} \frac{\ell}{:}\left(\frac{:-\ell}{=}\right)^{'}\prod_{8=1}^{'}\sqrt{\frac{1-?}{=}}^{3_8} + \$\sqrt{?\frac{:-\ell}{=}\frac{1-?}{=}}^{?} & \text{if } 3_1, \ldots, 3_{1'} > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Proof of Theorem 6.6. Let $!'^{\leq}$ be the degree- likelihood ratio between $_{\text{planted}}$ and $_{\text{null}}$. Then, by standard results, $!'^{\leq} - 1^2 = \sum_{0<|\text{|}\leq} \cdots_{\text{planted}}[\text{"}]^2$, where the norm is the one induced by $_{\text{null}}$. Therefore, if the right-hand side is >(1), then $!'^{\leq}$ is 1 + >(1), and if the right-hand side is unbounded, then $!'^{\leq}$ is also unbounded.

Consider all with ! left vertices and ' right vertices. The contribution from these is, by Lemma 6.8,

$$\tilde{O}\left(\sum_{\cdots_{\text{planted}}[\text{"}]^2}\right)$$

$$= \tilde{O}\left(\sum_{3_1,\ldots,3_{1'}>1}\left(\frac{\ell}{:-\ell}\right)^{2!}\left(\frac{:-\ell}{=}\right)^{2'}\ddot{O}^{'}\left(\sum_{8=1}^{\sqrt{\frac{1-?}{?}}^{3_8}} + \$\sqrt{\frac{:-\ell}{=}\frac{1-?}{?}}\right)^{2}\cdot\left(\frac{:}{=}\right)^{!}\right)\text{Bip}(!, ', 3_1, \ldots, 3_{1'}),$$

where $3_1, \ldots, 3_{1'}$ represent the number of edges in incident to each of the right vertices, and $\text{Bip}(!, ', 3_1, \ldots, 3_{1'})$ is the number of bipartite graphs with ! left vertices and ' right vertices such that all left degrees are at least 1 and the right degrees are $3_1, \ldots, 3_{1'}$.

Consider a choice of ! and ' such that $\sum_{3_1,\ldots,3_{1'}} \text{Bip}(!, ', 3_1, \ldots, 3_{1'}) \neq 0$. Because we are interested in the behavior of the sum as = goes to infinity, we ignore as negligible all factors that depend only on ! and '. In particular, because $\sum_{3_1,\ldots,3_{1'}} \text{Bip}(!, ', 3_1, \ldots, 3_{1'})$ can be bounded in terms of only ! and ', we can focus on the term corresponding to $3_1 = \ldots = 3_{1'} = 2$, which

maximizes the contribution of the terms $\frac{1}{?}$? $^{3}{}_{8}$ and is therefore proportional to the entire sum up to factors that depend only on ! and ' . We also approximate : − ℓ ≈ : and $\frac{1\ ?}{-?}$ + \$: $\frac{\ell\ 1\ ?}{=\ -\ ?}$ $^{2'}$ ≈ $\frac{1\ ?}{?}$ $^{2'}$. Then we have

Õ ‥planted ["] ²⍰ ℓ $^{2!}$: $^{2'}$ 1 − ? $^{2'}$ · ¡ =

' ! right vertices left vertices

⍰ ℓ $^{2!}$: $^{2'}$ (1 − ?)$^{2'}$: $_{!}$ = '

= = ' : $_{2-}$ ' ! ℓ $_{2}$! (1 − ?)$^{2'}$.

For : = =$^{1/2+}$/(1 − ?)$_{1}{}^{/2}$, the above is equal to =$^{(2'-!)-!/2}$ℓ$_{2}$!(1 − ?)'$_{+}$!/2. For 1 − ? = @= =⁻, this is equal to =$^{(2'-!)-!/2-('+!/2)}$ℓ$_{2}$!. Finally, for ℓ = =, this is equal to =$^{(2'-!)+2!-!/2-('+!/2)}$.

For ≤ /2, we have that the above is at most =$^{-!+2!-!/2-!/2}$. For the exponent to be non-negative, we need 2! ≥ !/2, so ≥ 1/4. In particular, for ≤ 1/4 − 0.001, the term goes to zero as = goes to infinity regardless of how large ! is. Therefore, the sum of all the terms with || ≤ 5(1/) is >(1), for any function 5 independent of =.

For ≥ , consider the term corresponding to ! = 2 and some ' = \$(1/). We have that the term is at least =$^{'-!+2!-!/2-!/2}$ = =$^{'-2+4-1-}$ ≥ = for ' = \$(1/) large enough. Therefore, this term goes to infinity as = goes to infinity, and then the same is true for the sum of all the terms.

References

[ABS15] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. *J. ACM*, 62(5):Art. 42, 25, 2015.

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. In *Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan, 1997)*, volume 13, pages 457–466, 1998.

[BDH+20] Ainesh Bakshi, Ilias Diakonikolas, Samuel B. Hopkins, Daniel Kane, Sushrut Karmalkar, and Pravesh K. Kothari. Outlier-robust clustering of gaussians and other non-spherical mixtures. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 149–159. IEEE, 2020.

[BDJ+20] Ainesh Bakshi, Ilias Diakonikolas, He Jia, Daniel M. Kane, Pravesh K. Kothari, and Santosh S. Vempala. Robustly learning mixtures of k arbitrary gaussians. *CoRR*, abs/2012.02119, 2020.

[BHK+16] Boaz Barak, Samuel B. Hopkins, Jonathan A. Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *FOCS*, pages 428–437. IEEE Computer Society, 2016.

[BHK+19] 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.

[BK20] Ainesh Bakshi and Pravesh Kothari. Outlier-robust clustering of non-spherical mixtures. *CoRR*, abs/2005.02970, 2020.

[BK21] Ainesh Bakshi and Pravesh K. Kothari. List-decodable subspace recovery: Dimension independent error in polynomial time. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 1279–1297. SIAM, 2021.

[BKar] Ainesh Bakshi and Pravesh Kothari. Outlier-robust clustering of non-spherical mixtures, 2020 (conference version merged with "Robustly Learning any Clusterable Mixture of Gaussians" by Diakonikolas, Hopkins, Kane, and Karmalkar).

[BKS15] Boaz Barak, Jonathan A. Kelner, and David Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method [extended abstract]. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 143–151. ACM, New York, 2015.

[BKS17] Boaz Barak, Pravesh K. Kothari, and David Steurer. Quantum entanglement, sum of squares, and the log rank conjecture. In *STOC*, pages 975–988. ACM, 2017.

[BP21] Ainesh Bakshi and Adarsh Prasad. Robust linear regression: optimal rates in polynomial time. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 102–115. ACM, 2021.

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science—FOCS 2011*, pages 472–481. IEEE Computer Soc., Los Alamitos, CA, 2011.

[BS95] A. Blum and J. Spencer. Coloring random and semi-random k-colorable graphs. *Journal of Algorithms*, 19(2):204 – 234, 1995.

[BS16] Boaz Barak and David Steurer. Proofs, beliefs, and algorithms through the lens of sum-of-squares, 2016. Lecture notes in preparation, available on http://sumofsquares.org.

[CSV17] Moses Charikar, Jacob Steinhardt, and Gregory Valiant. Learning from untrusted data. In *STOC*, pages 47–60. ACM, 2017.

[DHKK20] Ilias Diakonikolas, Samuel B. Hopkins, Daniel Kane, and Sushrut Karmalkar. Robustly learning any clusterable mixture of gaussians. *CoRR*, abs/2005.06417, 2020.

[DKWB19] Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Subexponential-time algorithms for sparse pca. *arXiv preprint arXiv:1907.11635*, 2019.

[DMM09] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[Fei19] Uriel Feige. Introduction to semirandom models. In Tim Roughgarden, editor, *Beyond Worst-case Analysis of Algorithms*, chapter 10, pages 266–290. Oxford, 2019.

[FGR+13] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. In *STOC*, pages 655–664. ACM, 2013.

[FGR+17] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S. Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM*, 64(2):Art. 8, 37, 2017.

[FK98] Uriel Feige and Joe Kilian. Heuristics for finding large independent sets, with applications to coloring semi-random graphs. In *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No. 98CB36280)*, pages 674–683. IEEE, 1998.

[FK00] Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures Algorithms*, 16(2):195–208, 2000.

[FK01] Uriel Feige and Joe Kilian. Heuristics for semirandom graph problems. *Journal of Computer and System Sciences*, 63(4):639 – 671, 2001.

[FK03] 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.

[FKP19] Noah Fleming, Pravesh Kothari, and Toniann Pitassi. Semialgebraic proofs and efficient algorithm design. *Foundations and Trends® in Theoretical Computer Science*, 14(1-2):1–221, 2019.

[GJW20] David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Low-degree hardness of random optimization problems. *arXiv preprint arXiv:2004.12063*, 2020.

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42(6):1115–1145, 1995.

[Hås99] Johan Håstad. Clique is hard to approximate within $=^{1-}$. *Acta Math.*, 182(1):105–142, 1999.

[HKP+17] 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 *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE, 2017.

[HL18] Samuel B. Hopkins and Jerry Li. Mixture models, robustness, and sum of squares proofs. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1021–1034. ACM, 2018.

[Hop18] Samuel Hopkins. Statistical inference and the sum of squares method. *PhD thesis, Cornell University*, 2018.

[HS17] Samuel B Hopkins and David Steurer. Efficient bayesian estimation from few samples: community detection and related problems. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 379–390. IEEE, 2017.

[IK22] Misha Ivkov and Pravesh K. Kothari. List-decodable covariance estimation. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1276–1283. ACM, 2022.

[Jer92] Mark Jerrum. Large cliques elude the metropolis process. *Random Struct. Algorithms*, 3(4):347–360, 1992.

[JPR+22] Chris Jones, Aaron Potechin, Goutham Rajendran, Madhur Tulsiani, and Jeff Xu. Sum-of-squares lower bounds for sparse independent set. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science—FOCS 2021*, pages 406–416. IEEE Computer Soc., Los Alamitos, CA, [2022] ©2022.

[Kar72] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.

[Kho14] Subhash Khot. Hardness of approximation. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. 1*, pages 711–728. Kyung Moon Sa, Seoul, 2014.

[KKK19] Sushrut Karmalkar, Adam R. Klivans, and Pravesh Kothari. List-decodable linear regression. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 7423–7432, 2019.

[KKM18] Adam R. Klivans, Pravesh K. Kothari, and Raghu Meka. Efficient algorithms for outlier-robust regression. In *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*, pages 1420–1430, 2018.

[KMOW17] Pravesh K. Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *STOC*, pages 132–145. ACM, 2017.

[KS17] Pravesh K. Kothari and Jacob Steinhardt. Better agnostic clustering via relaxed tensor norms. 2017.

[KSS18] Pravesh K. Kothari, Jacob Steinhardt, and David Steurer. Robust moment estimation and improved clustering via sum of squares. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1035–1046. ACM, 2018.

[Kuc95] Ludek Kucera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57(2-3):193–212, 1995.

[KWB19] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. *arXiv preprint arXiv:1907.11636*, 2019.

[Las01] Jean B. Lasserre. New positive semidefinite relaxations for nonconvex quadratic programs. In *Advances in convex analysis and global optimization (Pythagorion, 2000)*, volume 54 of *Nonconvex Optim. Appl.*, pages 319–331. Kluwer Acad. Publ., Dordrecht, 2001.

[LM21] Allen Liu and Ankur Moitra. Settling the robust learnability of mixtures of gaussians. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 518–531. ACM, 2021.

[LM22] Allen Liu and Ankur Moitra. Minimax rates for robust community detection. *arXiv preprint arXiv:2207.11903*, 2022.

[MMT20] Theo McKenzie, Hermish Mehta, and Luca Trevisan. A new algorithm for the robust semi-random independent set problem. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms*, pages 738–746, 2020.

[MPW16] Ankur Moitra, William Perry, and Alexander S. Wein. How robust are reconstruction thresholds for community detection? In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 828–841. ACM, New York, 2016.

[MS16] Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 814–827. ACM, New York, 2016.

[Nes00] Yurii Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, volume 33 of *Appl. Optim.*, pages 405–440. Kluwer Acad. Publ., Dordrecht, 2000.

[Par00] Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

[RT12] Prasad Raghavendra and Ning Tan. Approximating CSPs with global cardinality constraints using SDP hierarchies. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 373–384. ACM, New York, 2012.

[RY20a] Prasad Raghavendra and Morris Yau. List decodable learning via sum of squares. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 161–180. SIAM, 2020.

[RY20b] Prasad Raghavendra and Morris Yau. List decodable subspace recovery. In Jacob D. Abernethy and Shivani Agarwal, editors, *Conference on Learning Theory, COLT 2020, 9-12 July 2020, Virtual Event [Graz, Austria]*, volume 125 of *Proceedings of Machine Learning Research*, pages 3206–3226. PMLR, 2020.

[Sho87] N. Z. Shor. Quadratic optimization problems. *Izv. Akad. Nauk SSSR Tekhn. Kibernet.*, (1):128–139, 222, 1987.

[Ste17] Jacob Steinhardt. Does robustness imply tractability? a lower bound for planted clique in the semi-random model. *arXiv preprint arXiv:1704.05120*, 2017.

[SW20] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *arXiv preprint arXiv:2008.02269*, 2020.

[Tao12] Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[Wei20] Alexander S Wein. Optimal low-degree hardness of maximum independent set. *arXiv preprint arXiv:2010.06563*, 2020.

[Zuc07] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory Comput.*, 3:103–128, 2007.