Association schemes, non-commutative polynomial concentration, and sum-of-squares lower bounds for planted clique

Raghu Meka[*]

Institute for Advanced Study, Princeton & DIMACS, Rutgers

Avi Wigderson[*]

Institute for Advanced Study, Princeton

Abstract

Finding cliques in random graphs and the closely related "planted" clique variant, where a clique of size t is planted in a random $G(n, 1/2)$ graph, have been the focus of substantial study in algorithm design. Despite much effort, the best known polynomial-time algorithms only solve the problem for $t = \Theta(\sqrt{n})$. Here we show that beating $\sqrt{n}$ would require substantially new algorithmic ideas, by proving a lower bound for the problem in the sum-of-squares (or Lasserre) hierarchy, the most powerful class of semidefinite programming algorithms we know of: r rounds of the sum-of-squares hierarchy can only solve the planted clique for $t \geq \sqrt{n}/(C \log n)^{r^2}$. Previously, no nontrivial lower bounds were known. Our proof is formulated as a degree lower bound in the Positivstellensatz algebraic proof system, which is equivalent to the sum-of-squares hierarchy.

The heart of our (average-case) lower bound is a proof that a certain random matrix derived from the input graph is (with high probability) positive semidefinite. Two ingredients play an important role in this proof. The first is the classical theory of association schemes, applied to the average and variance of that random matrix. The second is a new large deviation inequality for matrix-valued polynomials. Our new tail estimate seems to be of independent interest and may find other applications, as it generalizes both the estimates on real-valued polynomials and on sums of independent random matrices.

1 Introduction

Finding cliques in random graphs has been the focus of substantial study in algorithm design. Let $G(n, p)$ denote Erdös-Renyi random graphs on n vertices where each edge is kept in the

---

graph with probability p. It is easy to check that in a random graph $G \leftarrow G(n, 1/2)$, the largest clique has size $(2 + o(1)) \log_2 n$ with high probability. On the other hand, the best known polynomial-time algorithms can only find cliques of size $(1+o(1)) \log_2 n$ and obtaining better algorithms remains a longstanding open problem: Karp [Kar76] even suggested that finding cliques of size $(1 + \varepsilon) \log_2 n$ could be a computationally hard problem.

Motivated by this, much attention has been given to the related *planted clique problem* or *hidden clique problem* introduced by Jerrum [Jer92] and Kucera [Kuc95]. Here, we are given a graph $G \leftarrow G(n, 1/2, t)$ generated by first choosing a $G(n, 1/2)$ random graph and placing a clique of size t in the random graph for $t \gg \log_2 n$. The goal is to recover the hidden clique for as small a t as possible given G. The best known polynomial-time algorithms can solve the problem for $t = \Theta(\sqrt{n})$ [AKS98] and improving on this bound has received significant attention.

In this work we show that indeed obtaining polynomial-time algorithms recognizing cliques of size $n^{1/2-\varepsilon}$ requires significantly new algorithmic ideas: we exhibit a lower bound for the problem in the powerful Lasserre [Las01] and "sum-of-squares" (SOS) [Par00] semi-definite programming hierarchies[1]. As it happens, showing such lower bounds for the planted clique problem reduces easily to showing integrality gap results for the natural formulation of the maximum clique problem in these hierarchies on $G(n, 1/2)$ graphs. Our main result then is the following average-case lower bound for maximum clique (we defer the formal definition of the semi-definite relaxation and hierarchies for now)[2]:

Theorem 1.1. *With high probability, for $G \leftarrow G(n, 1/2)$ the natural r-round SOS relaxation of the maximum clique problem has an integrality gap of at least $\sqrt{n}/(C \log n)^{r^2}$.*

As a corollary we obtain the following lower bound for the planted clique problem.

Corollary 1.2. *With high probability, for $G \leftarrow G(n, 1/2, t)$ the natural r-round SOS relaxation of the maximum clique problem has an integrality gap of at least $\sqrt{n}/t(C \log n)^{r^2}$.*

Linear and semi-definite hierarchies are one of the most powerful and well-studied techniques in algorithm design. The most prominent of these are the Sherali-Adams hierarchy (SA) [SA90], Lovasz-Schrijver hierarchy (LS) [LS91], their semi-definite versions SA_+, LS_+ and Lasserre and SOS hierarchies. The hierarchies present progressively stronger convex relaxations for combinatorial optimization problems parametrized by the number of *rounds r* and the r round relaxation can be solved in $n^{O(r)}$ time on instances of size n. In terms of relative power (barring some minor technicalities about how the numbering starts), it is known that $LS_+(r) < SA_+(r) < SOS(r)$. On the flip side, because they capture most powerful techniques for combinatorial optimization, lower bounds for hierarchies serve as strong unconditional evidence for computational hardness. Such lower bounds are even more relevant and compelling in situations where we do not have NP-hardness results.

---

[1]For brevity, in the following, we will use SOS hierarchy as a common term for the formulations of Lasserre [Las01] and Parrilo [Par00] which are essentially the same in our context.

[2]Throughout, c, C denote constants.

Broadly speaking, our understanding of the SOS hierarchy is more limited than those of LS$_+$ and SA$_+$ hierarchies and in fact the SOS hierarchy appears to be much more powerful. A particularly striking example of this phenomenon was provided by a recent work of Barak et al. [BBH$^+$12]. They showed that a constant number of rounds of the SOS hierarchy can solve the much studied *unique games problem* on instances which need super constant number of LS$_+$, SA$_+$ rounds. It was also shown by the works of [BRS11, GS11] that the SOS hierarchy captures the sub-exponential algorithm for unique games of [ABS10]. These results emphasize the need for a better understanding of the power and limitations of the SOS hierarchy.

From the perspective of proving limitations, all known lower bounds for the SOS hierarchy essentially have their origins in the works of Grigoriev [Gri01b, Gri01a] some of which were also independently rediscovered by Schoenebeck [Sch08]. These works show that even $\Omega(n)$ rounds of SOS hierarchy cannot solve random $3XOR$ or $3SAT$ instances leading to a strong unconditional average-case lower bound for a natural distribution.

Most subsequent lower bounds for SOS hierarchy such as those of [Tul09], [BCV$^+$12] rely on [Gri01b] and [Sch08] and gadget reductions. For example, Tulsiani [Tul09] shows that $2^{O(\sqrt{\log n})}$ rounds of SOS has an integrality gap of $n/2^{O(\sqrt{\log n})}$ for maximum clique in worst-case. This is in stark contrast to the average-case setting: even a single round of SOS gets an integrality gap of at most $O(\sqrt{n})$ for maximum clique on $G(n, 1/2)$ [FK00]. Thus, the worst-case and average-case problems have very different complexities. Finally, using reductions is problematic for us as they almost always induce distributions that are far from uniform and definitely not as natural as $G(n, 1/2)$.

For max-clique on random $G(n, 1/2)$ graphs, Feige and Krauthgamer [FK00] showed that LS$_+(r)$, and hence SOS(r), has an integrality gap of at most $\sqrt{n}/2^{O(r)}$ with high probability. Complementing this, they also showed [FK03] that the gap remains $\sqrt{n}/2^r$ for LS$_+(r)$ with high probability. However, there were no non-trivial lower bounds known for the stronger SOS hierarchy.

For the planted clique problem, Jerrum [Jer92] showed that a broad class of Markov chain Monte-Carlo (MCMC) based methods cannot solve the problem when the planted clique has size $O(n^{1/2-\delta})$ for any constant $\delta > 0$. In a recent work, Feldman et al. [FGR$^+$13] introduced the framework of *statistical algorithms* which generalizes many algorithmic approaches like MCMC methods and showed that such algorithms cannot find large cliques when the planted clique has size $O(n^{1/2-\delta})$ in less than $n^{\Omega(\log n)}$ time[3]. However, their framework seems quite different from hierarchy based algorithms. In particular, the statistical algorithms framework is not applicable to algorithms which first pick a sample, fix it, and then perform various operations (such as convex relaxations) on it, as is the case for hierarchies.

---

[3]The results of [FGR$^+$13] actually apply to the harder *bipartite* planted clique problem, but this assumption is not too critical.

1.1 Proof systems and SDP hierarchies

We approach the problem of SOS lower bounds from the proof-complexity and *positivstellensatz* proof system perspective of Grigoriev and Volobjov [GV01]. We explain this proof system next.

Suppose we are given a system of polynomial equations or "axioms"

$$f_1(x) = 0, \ f_2(x) = 0, \ \ldots, \ f_m(x) = 0,$$

where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is a n-variate polynomial. A positivstellensatz *refutation* of the system $\mathcal{F} = ((f_i))$ is an identity of the form

$$\sum_{i=1}^{m} f_i g_i \equiv 1 + \sum_{i=1}^{N} h_i^2,$$

where $\{g_1, \ldots, g_m\}$ and $\{h_1, \ldots, h_N\}$ are arbitrary n-variate polynomials. Clearly, the system $\mathcal{F}$ is infeasible over reals if there exists an identity as above. Starting with the seminal work of Artin on Hilbert's seventeenth problem [Art27], a long line of important results in real algebraic geometry – [Kri64, Ste73, Put93, Sch91]; cf. [BCR98] and references therein – showed that, under some (important) technical conditions[4], such certifying identities always exist for an infeasible system. This motivates the following notion of complexity for refuting systems of polynomial equations.

Definition 1.3 (Positivstellensatz Refutation, [GV01]). *Let $\mathcal{F} \equiv \{f_1, \ldots, f_n : \mathbb{R}^n \to \mathbb{R}\}$, be a system of axioms, where each f_i is a real n-variate polynomial. A positivstellensatz refutation of degree r (PS(r) refutation, henceforth) for $\mathcal{F}$ is an identity of the form*

$$\sum_{i=1}^{m} f_i g_i \equiv 1 + \sum_{i=1}^{N} h_i^2, \tag{1.1}$$

where $g_1, \ldots, g_m, h_1, \ldots, h_N$ are n-variate polynomials such that $deg(f_i g_i) \leq 2r$ for all $i \in [m]$ and $deg(h_j) \leq r$ for all $j \in [N]$.

Our interest in positivstellensatz refutations as above comes from the known relations between such identities and SOS hierarchy. Informally (under appropriate technical conditions), identities as above of degree r show that SOS hierarchy can *certify* infeasibility of the axioms in $2r + \Theta(1)$ rounds and vice versa. We will focus on showing degree lower bounds for identities as above and use them to get integrality gaps for the hierarchies. We formalize this in Section A. For a brief history of the different formulations from [GV01], [Las01], [Par00] and the relations between them and results in real algebraic geometry we refer the reader to [OZ13].

Given the above setup, we shall consider the following set of natural axioms to test if a graph G has a clique of size k.

---

[4]We avoid going into the details here as the conditions are easily met in the presence of Boolean axioms.

Definition 1.4. *Given a graph G, let* $\text{Clique}(G, k)$ *denote the following set of polynomial axioms:*

$$
\begin{aligned}
\textit{(Max-Clique):} \quad & x_i^2 - x_i, \quad \forall i \in [n] \\
& x_i \cdot x_j, \quad \forall \text{ pairs } \{i, j\} \notin G \\
& \sum_i x_i - k.
\end{aligned}
\tag{1.2}
$$

Here, the equations on the first line are *Boolean axioms* restricting feasible solutions to be in $\{0, 1\}^n$. The equations on the second line constrain the support of any feasible x to define a clique in G. Finally, the equation on the third line specifies the size of support of x. Thus, for any graph G, $\text{Clique}(G, k)$ is feasible if and only if G has a clique of size k. Our core result is to show lower bounds on positivstellensatz refutations for $\text{Clique}(G, k)$.

Theorem 1.5 (Main). *With high probability over $G \leftarrow G(n, 1/2)$, the system $\text{Clique}(G, k)$ defined by Equation 1.2 has no $\mathsf{PS}(r)$ refutation for $k \leq \sqrt{n}/(C \log n)^{r^2}$.*

Given the above theorem it is easy to deduce the integrality gap for the SOS hierarchy, Theorem 1.1: see Section A. We next highlight some of our techniques which may be of broader interest.

1.2   Techniques: Association schemes

As we will soon see, the essence of proving Theorem 1.5 involves showing that a certain random matrix is positive semi-definite (PSD) with high probability. In our case, this calls for showing a relation of the form $A \prec B$[5] for two matrices A, B whose rows and columns are indexed by subsets of $[n]$ of size r. Often, in such situations and especially those involving random matrices, it suffices to show that the smallest eigenvalue of B is bigger than the norm of A, but not for us. Luckily, the matrices we study, though complicated to describe, will be *set-symmetric* - namely matrices whose entries depend only on the size of the intersection of the corresponding row and column indexing sets. The set of all such matrices, called the *Johnson scheme*, is quite well studied in combinatorics as a special case of *association schemes*. In particular, all such matrices commute with one another and their common eigenspaces are completely understood. This theory allows us to show that the eigenvalue of $B - A$ is non-negative on each of the specific eigenspaces of the Johnson scheme.

1.3   Techniques: Large deviation bounds for non-commutative polynomials

As part of showing the requisite PSD'ness relation, our lower bound argument naturally calls for understanding the spectrum of a random matrix. There is a lot of recent interest in such

---

[5]Here and henceforth $\prec$ denotes PSD ordering: $A \prec B$ if and only if $B - A$ is positive definite.

bounds, especially in the context of large deviation inequalities for sums of independent matrices. The specific random matrix we study is more complicated and to bound its spectrum we prove a large-deviation bound for *non-commutative* or *matrix-valued* polynomials .

Define $P : \mathbb{R}^n \to \mathbb{R}^{D \times D}$ to be a degree r *matrix polynomial* if it is a degree r multilinear polynomial in n variables where the coefficients are D-dimensional matrices: $P(x) = \sum_{I:I \subseteq [n], 0 < |I| \leq r} \prod_{i \in I} x_i A_I$, where $A_I \in \mathbb{R}^{D \times D}$. We say P is Hermitian if the individual matrices A_I are Hermitian. Let $\mathsf{Var}(P) = \sum_I A_I A_I^\dagger = \mathbb{E}_{x \in_u \{1,-1\}^n} P(x) P(x)^\dagger$.

Our main claim is the following large deviation inequality for matrix polynomials:

Theorem 1.6. *There exist constants $c, C > 0$ such that the following holds. Let $P : \mathbb{R}^n \to \mathbb{R}^{D \times D}$ be a degree r Hermitian polynomial. Then, for all $\Delta > 0$, with probability at least $1 - Cn^{r-1} \cdot D \cdot \exp(-c\,\Delta^{2/r})$ over $\varepsilon \in_u \{1, -1\}^n$,*

$$-\Delta\,\mathsf{Var}(P)^{1/2} \preceq P(\varepsilon) \preceq \Delta\,\mathsf{Var}(P)^{1/2}.$$

The above theorem combines two important classes of large-deviation inequalities. The first are matrix-valued Chernoff bounds [AW02, Oli10], and non-commutative Khinchine inequalities [LP86, LPP91]. We refer to the excellent survey of [Tro12] for more details and some of their applications. The second are the now well-known large deviation inequalities for polynomials which follow from *hypercontractivity* [Jan97]. Such inequalities and hypercontractivity of polynomials are of considerable importance in analysis of Boolean functions; see [O'D08] for instance. Our result gives a common generalization of these inequalities. We give a more detailed comparison of the bounds in Section 5.

2 Outline

We now give an outline of our arguments. As in most previous works (cf. [Gri01a], [Gri01b], [Sch08]) on showing lower bounds for $\mathsf{PS}(r)$ refutations, our main tool will be a *dual certificate*. Let $\mathcal{P}(n, 2r) : \mathbb{R}^n \to \mathbb{R}$ be the set of n-variate real polynomials of total degree at most $2r$.

Definition 2.1 (PSD Mappings). *A linear mapping $\mathcal{M} : \mathcal{P}(n, 2r) \to \mathbb{R}$ is said to be positive semi-definite (PSD) if $\mathcal{M}(P^2) \geq 0$ for all n-variate polynomials P of degree at most r.*

Definition 2.2 (Dual Certificates). *Given a set of axioms $f_1, \ldots, f_m$, a dual certificate for the axioms is a PSD mapping $\mathcal{M} : \mathcal{P}(n, 2r) \to \mathbb{R}$ such that $\mathcal{M}(f_i g) = 0$ for all $i \in [m]$ and all polynomials g such that $deg(f_i g) \leq 2r$.*

For readers more familiar with the optimization framework of SOS hierarchy, dual certificates correspond to feasible vector solutions for the r-round SOS-relaxation. Under reasonable technical conditions which ensure *strong duality*, the converse also holds. For the clique axioms from Equation 1.2, a dual certificate would correspond to a feasible vector solution for the r-round SOS relaxation for maximum clique (see Figure A for the exact formulation) with *value* k.

The following elementary lemma will be crucial.

Lemma 2.3 (Dual Certificate). *Given a system of axioms $((f_i))$, there does not exist a $\mathsf{PS}(r)$ refutation of the system if there exists a dual certificate $\mathcal{M} : \mathcal{P}(n, 2r) \to \mathbb{R}$ for the axioms.*

The existence of such a mapping trivially implies a lower bound for $\mathsf{PS}(r)$ refutations: apply $\mathcal{M}$ to both sides of a purported $\mathsf{PS}(r)$ identity as in Equation 1.1 to arrive at a contradiction.

The lemma suggests a general recipe for proving $\mathsf{PS}(r)$ refutation lower bounds:

- Design a dual certificate $\mathcal{M}$: For the clique axioms we care about, it is easy to figure out what the *right* dual certificate $\mathcal{M}$ "should be" by working backwards from the axioms. The same happens also for the $\mathsf{PS}(r)$ refutation lower bounds of [Gri01a, Gri01b]. The main hurdle then is to show that the obtained mapping $\mathcal{M}$ is indeed PSD. At a high level, this reduces to proving a certain random matrix $M \in \mathbb{R}^{N \times N}$ is PSD, where $N = \sum_{\ell=0}^{r} \binom{n}{\ell}$. We consider $E = \mathbb{E}[M]$ and (*variance*) $\Sigma^2 = \mathsf{Var}(M) = \mathbb{E}[(M - E)(M - E)^{\dagger}]$ and prove $M \succeq 0$ via $E - M \preceq E$ in two steps.

- Large deviation: with high probability $E - M \preceq \Delta\Sigma$, for some small factor Δ. In our case, M can be written as a low-degree matrix polynomial, and we use Theorem 1.6 to show the claim.

- (Deterministic) Matrix analysis: $\Delta\Sigma \preceq E$. In our case, both E, Σ have common eigenspaces, and we compare their eigenvalues on each eigenspace via the theory of association schemes.

The above high-level plan has various complications. For one, the matrices E, M will be singular, so we need to handle the kernel differently. Moreover, the actual certificate M has too high a variance. Instead, we work with a closely related matrix M' which suffices for us. We gloss over these (important) issues in the high-level discussion of this section.

We next demonstrate the above three steps on the toy example $r = 1$, in a way that will hint at generalizations.

Step 1: For $r = 1$, the matrix $M \in \mathbb{R}^{n \times n}$ we study will be $M = (k/2) \cdot D + \binom{k}{2}A$, where D denotes the diagonal matrix of degrees of the vertices and A is the adjacency matrix of G.

For higher r, we build a similar looking matrix where instead of degrees of vertices we look at a natural, generalized notion of *degree* for subsets of vertices.

Step 2: For $r = 1$, $E = \mathbb{E}[M] = (kn/2)I_n + k(k-1)/4 J_n$, where I_n is the $n \times n$ identity matrix and J_n is the $n \times n$ all the one's matrix. Ignoring the diagonal entries for now (which are not too important), we can write $E - M$ as a random matrix where each entry is uniformly random in $\{-k(k-1)/4, k(k-1)/4\}$. Thus,

$$E - M \sim (k^2/4) \sum_{i \neq j} \varepsilon_{ij} E_{ij},$$

where $\varepsilon_{ij} \in_u \{1, -1\}$ and E_{ij} is the symmetric matrix with $E_{ij}(p, q) = 1$ if $\{p, q\} = \{i, j\}$ and 0 otherwise. Thus, we can write $E - M$ as a sum of independent random matrices. Applying

7

the matrix Chernoff bounds of Oliveira [Oli10], it follows that $E - M \prec O(\sqrt{\log n}) \, \Sigma$ with high probability[6].

For higher r, we do not have a sum of independent matrices as above, but instead get a degree $O(r^2)$ matrix polynomial and we apply Theorem 1.6 to this polynomial to get a similar claim.

Step 3: For $r = 1$, an easy calculation shows that $\Sigma = \mathsf{Var}(M)^{1/2} \sim (k^2 \sqrt{n})/4 \cdot I_n$ and for $k < n/2$, $E \succ (kn/4) I_n$. Thus, by the above arguments,

$$E - M \prec O(\sqrt{\log n}) \Sigma \prec \Delta E,$$

where $\Delta = O(\sqrt{\log n}) k / \sqrt{n}$. It follows that $M \succ 0$, as long as $k \ll \sqrt{n/\log n}$, so that $\Delta < 1$. The above argument also suggests how $\sqrt{n}$ arises as a natural limit for the lower bound we seek.

For higher r, we show similar inequalities but with $\Delta \sim (\log n)^{r^2/2} \cdot k/\sqrt{n}$, so that we get $M \succ 0$ as long as $k \ll \sqrt{n}/(\log n)^{r^2/2}$. However, as indicated above, this step is not as simple and we in fact have to study *all* the eigenvalues of E, Σ; we do so, by appealing to theory of association schemes. Another important point is that, while the expectation matrix is easy to write explicitly for all r, the variance is quite complicated for higher r and we only give sufficiently good estimates for its entries.

We now start with some preliminaries.

3   Preliminaries

We shall use the following notations[7]:

1. $\mathcal{P}(n, 2r)$ denotes the set of n-variate polynomials of degree at most $2r$.

2. $\mathsf{PS}(r)$ denotes positivstellensatz refutations of degree at most r as defined in Definition 1.3.

3. A linear mapping $\mathcal{M} : \mathcal{P}(n, 2r) \to \mathbb{R}$ is said to be positive semi-definite (PSD) if $\mathcal{M}(P^2) \geq 0$ for all $P \in \mathcal{P}(n, r)$.

4. For $0 \leq r \leq n$, let $\binom{[n]}{r}$, $\binom{[n]}{\leq r}$ denote all subsets of size exactly and at most r, respectively.

5. For $0 \leq r \leq n$, $\mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ denotes matrices with rows and columns indexed by subsets of $[n]$ of size exactly r. Similarly, $\mathbb{R}^{\binom{[n]}{\leq r} \times \binom{[n]}{\leq r}}$ denotes matrices with rows and columns indexed by subsets of $[n]$ of size at most r.

---

[6]For this special case, we can remove the $\sqrt{\log n}$ factor–see [Ver] for instance, but this is not important for us.

[7]Some are repeated from the introduction so as to have them at one place.

6. We will view linear functionals $\mathcal{M} : \mathcal{P}(n, 2r) \to \mathbb{R}$ as matrices $M \in \mathbb{R}^{\binom{[n]}{\leq r} \times \binom{[n]}{\leq r}}$, where for $I, J \in \binom{[n]}{\leq r}$, $M_{IJ} = \mathcal{M}\left(\prod_{s \in I \cup J} x_s\right)$. In general, this correspondence is not bijective. However, as we only deal with mappings which are constant under multilinear extensions throughout, the correspondence is one-to-one. It is a standard fact that a mapping $\mathcal{M}$ is PSD if and only if the matrix M is PSD.

7. For $I \subseteq [n]$, let $X_I = \prod_{i \in I} x_i$.

8. By default all vectors are column vectors. For a set I, $\mathbb{1}(I)$ denotes the indicator vector of the set I.

9. For a matrix $A \in \mathbb{R}^{m \times n}$, $A^\dagger \in \mathbb{R}^{n \times m}$ denotes its conjugate matrix.

We shall use the following fact about matrices which follows easily from the variational characterization of eigenvalues.

Fact 3.1. *For $M \in \mathbb{R}^{N \times N}$, let $M' \in \mathbb{R}^{N' \times N'}$ be a principal submatrix of M. Then, M has at least N' eigenvalues that are at least $\lambda_{\min}(M')$.*

4   Johnson scheme

Association schemes is a classical area in combinatorics and coding theory (cf. for instance [vLW01]). We shall use a few classical results (lemmas 4.6, 4.7 below), about the eigenspaces and eigenvalues of association schemes and the *Johnson scheme* in particular. We also introduce two bases for the Johnson scheme, which will play a key role in bounding the eigenvalues of various matrices later.

We start with some basics about the Johnson scheme - some of our notations are non-standard but they fit better with the rest of the manuscript.

Definition 4.1 (Set-Symmetry). *A matrix $M \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ is set-symmetric if for every $I, J \in \binom{[n]}{r}$, $M(I, J)$ depends only on the size of $|I \cap J|$.*

Definition 4.2 (Johnson Scheme). *For $n, r \leq n/2$, let $\mathcal{J} \equiv \mathcal{J}_{n,r} \subseteq \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ be the subspace of all set-symmetric matrices. $\mathcal{J}$ is called the Johnson scheme.*

As we will soon see, $\mathcal{J}$ is also a commutative algebra. There is a natural basis for the subspace $\mathcal{J}$:

Definition 4.3 (D-Basis). *For $0 \leq \ell \leq r \leq n$, let $D_\ell \equiv D_{n,r,\ell} \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ be defined by[8]*

$$D_\ell(I, J) = \begin{cases} 1 & |I \cap J| = \ell \\ 0 & otherwise. \end{cases} \tag{4.1}$$

---

[8] We will often omit the subscripts n, r.

For example, D_0 is the well-studied disjointness matrix. Clearly, $\{D_\ell : 0 \le \ell \le r\}$ span the subspace $\mathcal{J}$. Also, it is easy to check that the D_ℓ's and hence all the matrices in $\mathcal{J}$, commute with one another.

Another important collection of matrices that come up naturally while studying PSD'ness of set-symmetric matrices is the following which gives a basis of PSD matrices for the Johnson scheme.

Definition 4.4 (P-Basis). *For $0 \le t \le r$, let $P_t \equiv P_{n,r,t} \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ be defined by*[9]

$$P_t(I, J) = \binom{|I \cap J|}{t}.$$

Equivalently, for $T \subseteq [n]$, if we let P_T be the PSD rank one matrix

$$P_T = \mathbb{1}\left(\{I : I \subseteq [n], I \supseteq T\}\right) \cdot \mathbb{1}\left(\{I : I \subseteq [n], I \supseteq T\}\right)^\dagger,$$

then

$$P_t = \sum_{T : T \subseteq [n], |T| = t} P_T. \tag{4.2}$$

The equivalence of the above two definitions follows from a simple calculation: there is a non-zero contribution to (I, J)'th entry from the T'th summand from Equation 4.2 if and only if $T \subseteq I \cap J$. Clearly, $P_t \succeq 0$ for $0 \le t \le r$. We will exploit this relation repeatedly by expressing matrices in $\mathcal{J}$ as linear combinations of P_t's. The following elementary claim relates the two bases $((D_\ell))$ and $((P_t))$ for fixed n, r.

Claim 4.5. *For fixed n, r, the following relations hold:*

1. *For $0 \le t \le r$, $P_t = \sum_{\ell=t}^r \binom{\ell}{t} D_\ell$.*

2. *For $0 \le \ell \le r$, $D_\ell = \sum_{t \ge \ell} (-1)^{t-\ell} \binom{t}{\ell} P_t$.*

Proof. The first relation follows immediately from the definition of P_t. The second relation follows from inverting the set of equations given in (1). $\square$

The main nontrivial result from the theory of association schemes we use is the following characterization of the eigenspaces of matrices in $\mathcal{J}$. The starting point for these characterizations is the fact that matrices in $\mathcal{J}$ commute with one another and hence are simultaneously diagonalizable. We refer the reader to Section 7.4 in [God] (the matrices P_t in our notation correspond to matrices C_t in [God]) for the proofs of these results.

Lemma 4.6. *Fix $n, r \le n/2$ and let $\mathcal{J} \equiv \mathcal{J}(n, r)$ be the Johnson scheme. Then, for P_t as defined by Equation 4.2, there exist subspaces $V_0, V_1, \ldots, V_r \in \mathbb{R}^{\binom{[n]}{r}}$ that are orthogonal to one another such that:*

---

[9]We will often omit the subscripts n, r.

1. $V_0, \ldots, V_r$ are eigenspaces for $\{P_t : 0 \le t \le r\}$ and consequently for all matrices in $\mathcal{J}$.

2. For $0 \le j \le r$, $dim(V_j) = \binom{n}{j} - \binom{n}{j-1}$.

3. For any matrix $Q \in \mathcal{J}$, let $\lambda_j(Q)$ denote the eigenvalue of Q within the eigenspace V_j. Then,

$$\lambda_j(P_t) = \begin{cases} \binom{n-t-j}{r-t} \cdot \binom{r-j}{t-j} & j \le t \\ 0 & j > t \end{cases}. \tag{4.3}$$

The above lemma helps us estimate the eigenvalues of any matrix in $Q \in \mathcal{J}$ if we can write Q as a linear combination of the P_t's or D_ℓ's. To this end, we shall also use the following estimate on the eigenvalues of such linear combinations.

Lemma 4.7. Let $Q = \sum_\ell \alpha_\ell D_\ell \in \mathcal{J}(n, r)$, and $\beta_t = \sum_{\ell \le t} \binom{t}{\ell} \alpha_\ell$, where $\alpha_\ell \ge 0$. Then, for $0 \le j \le r$,

$$\lambda_j(Q) \le \sum_{t \ge j} \beta_t \cdot \binom{n-t-j}{r-t} \cdot \binom{r-j}{t-j}.$$

Proof. By Claim 4.5,

$$\sum_\ell \alpha_\ell D_\ell = \sum_\ell \alpha_\ell \left(\sum_{t \ge \ell} (-1)^{t-\ell} \binom{t}{\ell} P_t \right) = \sum_t P_t \left(\sum_{\ell \le t} (-1)^{t-\ell} \binom{t}{\ell} \alpha_\ell \right)$$

$$\preceq \sum_t P_t \left(\sum_{\ell \le t} \binom{t}{\ell} \alpha_\ell \right) = \sum_t \beta_t P_t.$$

Therefore, as Q and P_t's have common eigenspaces, by Lemma 4.6,

$$\lambda_j(Q) \le \lambda_j \left(\sum_t \beta_t P_t \right) \le \sum_t \beta_t \lambda_j(P_t) = \sum_{t \ge j} \beta_t \cdot \binom{n-t-j}{r-t} \cdot \binom{r-j}{t-j}.$$

$\square$

5 Large deviation bounds for non-commutative polynomials

In this section we prove Theorem 1.6. For comparison, we first state the matrix deviation bound for the linear $r = 1$ case and the scalar large-deviation inequality for polynomials. The specific version follows from non-commutative Khinchine inequalities [LP86, LPP91] or from [Oli10][10] [11].

---

[10] The version here follows easily from the main statement of [Oli10] by applying Fact 5.5.

[11] An essential advantage of [Oli10] over [AW02] is that the former bounds the deviation in terms of the "norm-of-sum" $\| \sum_i A_i A_i^\dagger \|^{1/2}$, whereas [AW02] obtains a bound in terms of "sum-of-norm" $\left(\sum_i \|A_i A_i^\dagger\| \right)^{1/2}$. Our main theorem also gets the stronger norm-of-sum bound and having such a bound is critically important for us (even for the degree $r = 2$, $\mathsf{PS}(2)$ refutations, case). See Section 3.1 of [Oli10] for further discussion.

Theorem 5.1 ([Oli10]). *There exist constants $c, C > 0$ such that the following holds. For all Hermitian matrices $A_1, \ldots, A_n \in \mathbb{R}^{D \times D}$, and $\Delta > 0$, with probability at least $1 - CD \cdot \exp(-c\Delta^2)$ over $\varepsilon \in_u \{1, -1\}^n$,*

$$-\Delta \left(\sum_{i=1}^{n} A_i A_i^\dagger \right)^{1/2} \preceq \sum_i \varepsilon_i A_i \preceq \Delta \left(\sum_{i=1}^{n} A_i A_i^\dagger \right)^{1/2}.$$

The following classical large-deviation inequality for polynomials follows from *hypercontractivity*: see [Jan97] for instance.

Theorem 5.2. *There exist constants c, C such that the following holds. Let $P : \mathbb{R}^n \to \mathbb{R}$ be a polynomial of degree at most d. Then, for all $\Delta > 0$, with probability at least $1 - C \exp(-c\Delta^{2/d})$ over $\varepsilon \in_u \{1, -1\}^n$,*

$$-\Delta \, \mathsf{Var}(P)^{1/2} \leq P(\varepsilon) \leq \Delta \, \mathsf{Var}(P)^{1/2}.$$

Let us now compare Theorem 1.6 with the above results. To this end, let $1 - err(n, d, D, \Delta)$ be the minimum success probability for the conclusion in Theorem 1.6 to hold. Then, our bound is

$$err(n, d, D, \Delta) = n^{d-1} \cdot D \cdot \exp\left(-c\Delta^{2/d}\right).$$

The second factor is necessary because of the $d = 1$ case. The third factor is necessary from the $D = 1$ case. The first factor should be at least C^d for a constant C from the $D = 1$ case. However, for us the dimension of the matrices $D \approx n^{d-1}$ so it does not matter much. In addition, if one desires the error probability to be polynomially small (in the number of variables), which is often the case, the additional factor only costs us a constant factor in the deviation Δ. Nevertheless, we believe that the n^{d-1} factor is not needed and can be replaced by C^d.

To prove Theorem 1.6, we first show a similar claim for *decoupled* multi-linear polynomials. For any degree d multi-linear polynomial $P : \mathbb{R}^n \to \mathbb{R}^{D \times D}$, let $\tilde{P}(x^1, \ldots, x^d) : (\mathbb{R}^n)^d \to \mathbb{R}^{D \times D}$ denote the multi-linear polynomial, where every monomial $x_{i_1} \cdots x_{i_j}$, $i_1 < i_2 < \cdots < i_j$, $j \leq d$ in P is replaced by $x_{i_1}^1 x_{i_2}^2 \cdots x_{i_j}^j$. Further, we call a degree d polynomial *decoupled* if it has the above structure.

Lemma 5.3. *For a constant $C_1 > 0$ the following holds. Let $P : (\mathbb{R}^n)^d \to \mathbb{R}^{D \times D}$ be a decoupled degree d Hermitian polynomial. Then, for all $\delta \in (0, 1)$, $\Delta = (C_1 \log(n^{d-1} D/\delta))^{d/2}$, with probability at least $1 - \delta$, over $\varepsilon^1, \ldots, \varepsilon^d \in_u \{1, -1\}^n$,*

$$-\Delta \, \mathsf{Var}(P)^{1/2} \preceq P(\varepsilon^1, \ldots, \varepsilon^d) \preceq \Delta \, \mathsf{Var}(P)^{1/2}.$$

Proof. The proof will be by induction on the degree d with Theorem 5.1 serving as a base case. Suppose the statement is true for polynomials of degree d with $\Delta = (C_1 \log((n+1)^{d-1} D/\delta))^{d/2}$. Let P be a decoupled degree $d + 1$ polynomial, where $\varepsilon^1 = (\varepsilon_1^1, \varepsilon_2^1, \ldots, \varepsilon_n^1)$. Write

$$P(\varepsilon^1, \ldots, \varepsilon^d, \varepsilon^{d+1}) = \varepsilon_1^1 P_1(\varepsilon') + \varepsilon_2^1 P_2(\varepsilon') + \cdots + \varepsilon_n^1 P_n(\varepsilon'),$$

12

where $\varepsilon' \equiv (\varepsilon^2, \ldots, \varepsilon^{d+1})$ and each P_i is a decoupled polynomial of degree at most d. Let $\delta' = \delta/(n+1)$.

Note that for any fixed ε', the resulting function P is linear in ε^1 and hence, we can apply the degree 1 case Theorem 5.1 to the polynomial $P_{\varepsilon'}(\varepsilon^1) = P(\varepsilon^1, \varepsilon')$. Doing so with $\Delta_1 = (C_1 \log(D/\delta'))^{1/2}$, we get that for any fixed $\varepsilon' \in (\{1, -1\}^n)^d$, with probability at least $1 - \delta'$ over ε^1,

$$-\Delta_1 \left(\sum_i P_i(\varepsilon') P_i(\varepsilon')^\dagger \right)^{1/2} \preceq P_{\varepsilon'}(\varepsilon^1) = P(\varepsilon^1, \varepsilon') \preceq \Delta_1 \left(\sum_i P_i(\varepsilon') P_i(\varepsilon')^\dagger \right)^{1/2}.$$

Further, as each $P_i(\varepsilon')$ is in turn a degree d decoupled polynomial, by the induction hypothesis, for $\Delta_d = (C_1 \log((n+1)^{d-1} D/\delta'))^{d/2}$, with probability at least $1 - \delta'$,

$$-\Delta_d \, \mathsf{Var}(P_i)^{1/2} \preceq P_i(\varepsilon') \preceq \Delta_d \, \mathsf{Var}(P_i)^{1/2}.$$

By combining the above inequalities and using a union bound, we get that with probability at least $1 - (n+1)\delta' = 1 - \delta$,

$$
\begin{aligned}
P(\varepsilon^1, \varepsilon^2, \ldots, \varepsilon^{d+1}) &\preceq \Delta_1 \left(\sum_i P_i(\varepsilon') P_i(\varepsilon')^\dagger \right)^{1/2} \\
&\preceq \Delta_1 \left(\sum_i \left(\Delta_d \, \mathsf{Var}(P_i)^{1/2} \right) \cdot \left(\Delta_d \, \mathsf{Var}(P_i)^{1/2} \right)^\dagger \right)^{1/2} \\
&= \Delta_1 \Delta_d \left(\sum_i \mathsf{Var}(P_i) \right)^{1/2} = \Delta_1 \Delta_d \, \mathsf{Var}(P)^{1/2}.
\end{aligned}
$$

By a similar argument for the lower bounds, we get that with probability at least $1 - \delta$,

$$-\Delta \, \mathsf{Var}(P)^{1/2} \preceq P(\varepsilon^1, \ldots, \varepsilon^{d+1}) \preceq \Delta \, \mathsf{Var}(P)^{1/2},$$

where

$$\Delta = \Delta_1 \Delta_d = (C_1 \log(D/\delta'))^{1/2} \cdot (C_1 \log((n+1)^{d-1} D/\delta'))^{d/2} \leq (C_1 \log((n+1)^d D/\delta))^{(d+1)/2}.$$

It follows by induction that the statement holds for $\Delta = (C_1 \log((n+1)^{d-1} D/\delta))^{d/2}$. We can get the bound stated in the theorem by choosing a slightly bigger constant C_1 so as to replace $(n+1)$ with n in the expression for Δ. $\qquad\square$

We now prove Theorem 1.6. To do so we shall use the following claim of de La Peña and Montgomery-Smith [dlPnMS95].

Lemma 5.4. *There exists a constant $C_0 > 0$ such that the following holds. Let $P : \mathbb{R}^n \to \mathbb{R}^{D \times D}$ be a degree d polynomial and let $\tilde{P}$ be the decoupled version of P. Then, for every $t > 0$, and $\varepsilon, \varepsilon^1, \ldots, \varepsilon^d \in_u \{1, -1\}^n$,*

$$\Pr\left[\|P(\varepsilon)\| > t \right] \leq C_0^d \Pr\left[\left\| \tilde{P}(\varepsilon^1, \ldots, \varepsilon^d) \right\| > t/C_0^d \right].$$

Finally, we shall also use the following elementary fact that helps us relate bounding spectral norms to obtaining PSD relations.

Fact 5.5. *For any $A, B \in \mathbb{R}^{D \times D}$, with B Hermitian and $B \succ 0$, $\|B^{-1}A\| \leq \Delta$ if and only if $-\Delta B \preceq A \preceq \Delta B$.*

Proof. Note that $\|B^{-1}A\| = \|B^{-1/2}AB^{-1/2}\| \leq \Delta$. Given any $x \in \mathbb{R}^D$, let $y = B^{1/2}x$. Then,

$$x^\dagger A x = y^\dagger (B^{-1/2}AB^{-1/2})y \leq \Delta \cdot y^\dagger y = \Delta \cdot x^\dagger B x.$$

This shows that $A \preceq \Delta B$. The inequality $-\Delta B \preceq A$ follows similarly. $\qquad\square$

Proof of Theorem 1.6. Let P be the given polynomial and let $\tilde{P}$ be its decoupling. Let $\delta' = \delta/C_0^d$ and $t = (C_1 \log(n^{d-1}D/\delta'))^{d/2} \cdot C_0^d \cdot \|\mathsf{Var}(P)^{1/2}\|$. Then, from Lemma 5.4 and the tail bound for decoupled polynomials, Lemma 5.3, for $\varepsilon, \varepsilon^1, \ldots, \varepsilon^d \in_u \{1, -1\}^n$,

$$\Pr\left[\|P(\varepsilon)\| > t\right] \leq C_0^d \Pr\left[\|\tilde{P}(\varepsilon^1, \ldots, \varepsilon^d)\| > t/C_0^d\right] \leq C_0^r \cdot \delta' = \delta.$$

Therefore, for C a sufficiently large constant, we get that

$$\Pr\left[\|P(\varepsilon)\| > (C \log(n^{d-1}D/\delta))^{d/2} \cdot \|\mathsf{Var}(P)^{1/2}\|\right] \leq \delta.$$

Now, let $Q = \mathsf{Var}(P)^{-1/2}P$. Then, $\mathsf{Var}(Q) = \mathbb{I}_D$ (the $D \times D$ identity matrix). Applying the above inequality for Q, we get that for $\Delta = (C \log(n^{d-1}D/\delta))^{d/2}$, with probability at least $1 - \delta$, $\|\mathsf{Var}(P)^{-1/2}P(\varepsilon)\| \leq \Delta$. The claim now follows by applying Fact 5.5 and writing the error probability δ in terms of the deviation Δ. $\qquad\square$

6 Dual certificate for $\mathsf{PS}(r)$ refutations of max-clique

We will specify the dual certificate $\mathcal{M}$ by defining it for polynomials where each individual variable has degree at most 1 and *extend* $\mathcal{M}$ multi-linearly to all polynomials: for any polynomial P, $\mathcal{M}(P) = \mathcal{M}(\tilde{P})$ where $\tilde{P}$ is obtained from P by reducing the individual degrees of all variables to 1. We can do this without loss of generality because of the Boolean axioms.

As mentioned in the introduction, we can often work out what the dual certificate should be from the axioms and basic linear algebra. As an example, we first work out the case where the graph G is the complete graph; this will also help us draw a concrete connection to the work of [Gri01a].

6.1 Complete graph and knapsack

For complete graph, the clique axioms simplify to

$$\text{(Max-Clique):} \quad x_i^2 - x_i, \quad \forall i \in [n]$$
$$\sum_i x_i - k.$$

14

These incidentally also correspond to proving lower bounds for knapsack as studied by Grigoriev [Gri01a] (and was what lead us to the specific dual certificate we study). However, in the context of lower bounds for knapsack, the axioms are mainly interesting for non-integer k and Grigoriev shows that for non-integer $k \leq n/2$, the above system has no $\mathsf{PS}(r)$ refutation for $r < k$.

The above axioms tell us that any candidate dual certificate $\mathcal{M}_{Gr} \equiv: \mathcal{P}(n, 2r) \to \mathbb{R}$ should satisfy:

$$\mathcal{M}_{Gr}\left(\left(\sum_{i=1}^n x_i - k\right)\left(\prod_{i \in I} x_i\right)\right) = 0, \ \forall I, \ |I| < 2r.$$

Now, as the above equation is symmetric, it is natural to assume that $\mathcal{M}_{Gr}$ is also symmetric in the sense that $\mathcal{M}_{Gr}(X_I) = f(|I|)$ for some function $f : \{0, \ldots, 2r\} \to \mathbb{R}_+$. Working from this assumption, Grigoriev derives the following recurrence relation for $f : \{0, \ldots, 2r\} \to \mathbb{R}_+$,

$$f(i+1) = \frac{k-i}{n-i} f(i).$$

From the above it follows that we can define f and hence $\mathcal{M}$ as follows:

$$\mathcal{M}_{Gr}(X_I) = f(|I|) = f(0) \cdot \frac{k(k-1) \cdots (k-|I|)}{n(n-1) \cdots (n-|I|)}$$

Grigoriev takes $f(0) = 1$. Here we set $f(0) = \binom{n}{2r}$ with a view towards what is to come. Thus, the final certificate is

$$\mathcal{M}_{Gr}(X_I) = \binom{n}{2r} \cdot \frac{k(k-1) \cdots (k-|I|)}{n(n-1) \cdots (n-|I|)} = \binom{n-|I|}{2r-|I|} \cdot \frac{\binom{k}{|I|}}{\binom{2r}{|I|}}. \tag{6.1}$$

Grigoriev shows the following:

Theorem 6.1 ([Gri01a]). *For $k < n/2$, the mapping $\mathcal{M}_{Gr}$ defined above is PSD for $r < k$.*

6.2 Certificate for clique axioms

Following a similar approach, we now derive the dual certificate for the clique axioms from Equations 1.2, which we restate below for convenience: given a graph G on n vertices, $k \leq n$, the axioms of Clique(G, k) are

$$\begin{aligned} \text{(Max-Clique):} \quad & x_i^2 - x_i, \ \ \forall i \in [n] \\ & x_i \cdot x_j, \ \ \forall \text{ pairs } \{i, j\} \notin G \\ & \sum_i x_i - k. \end{aligned} \tag{6.2}$$

The above axioms tell us that any candidate dual certificate $\mathcal{M} \equiv \mathcal{M}_G : \mathcal{P}(n, 2r) \to \mathbb{R}$ should satisfy:

$$\mathcal{M}(X_I) = 0, \ \forall I, |I| \leq 2r, \ I \text{ is not a clique in } G,$$

$$\mathcal{M}\left(\left(\sum_{i=1}^{n} x_i - k\right) X_I\right) = 0, \ \forall I, |I| < 2r. \tag{6.3}$$

The above equations give us a system of linear equations that $\mathcal{M}$ needs to satisfy. By working with the equations, it is easy to *guess* a natural solution for the system.

Given a graph G on $[n]$, and $I \subseteq [n], |I| \leq 2r$, let

$$deg_G(I) = |\{S \subseteq [n] : I \subseteq S, \ |S| = 2r, \ S \text{ is a clique in } G\}|.$$

For instance, if $r = 1$ and $v \in G$, then $deg_G(\{v\})$ is the degree of vertex v.

We define $\mathcal{M} \equiv \mathcal{M}_G : \mathcal{P}(n, 2r) \to \mathbb{R}$ for monomials as follows: for $I \subseteq [n], |I| \leq 2r$, let

$$\mathcal{M}\left(\prod_{i \in I} x_i\right) = deg_G(I) \cdot \frac{k(k-1)\cdots(k-|I|+1)}{2r(2r-1)\cdots(2r-|I|+1)} = deg_G(I) \cdot \frac{\binom{k}{|I|}}{\binom{2r}{|I|}}. \tag{6.4}$$

It is easy to check the following claim:

Claim 6.2. *For any graph G, $\mathcal{M} \equiv \mathcal{M}_G$ defined by Equation 6.4 satisfies Equations 6.3.*

Proof. The first equation in Equation 6.3 follows immediately from the definition of $\mathcal{M}$. Now, for $I \subseteq [n], |I| < 2r$,

$$\mathcal{M}\left(\left(\sum_{i} x_i - k\right) X(I)\right) = (|I| - k)\mathcal{M}(X(I)) + \sum_{j \notin I} \mathcal{M}(X(I \cup \{j\}))$$

$$= (|I| - k) \cdot deg_G(I) \cdot \frac{\binom{k}{|I|}}{\binom{2r}{|I|}} + \sum_{j \notin I} deg_G(I \cup \{j\}) \cdot \frac{\binom{k}{|I|+1}}{\binom{2r}{|I|+1}}$$

$$= \frac{\binom{k}{|I|+1}}{\binom{2r}{|I|+1}} \cdot \left(-(2r - |I|) \cdot deg_G(I) + \sum_{j \notin I} deg_G(I \cup \{j\})\right).$$

Observe that our notion of *degree*, deg_G, satisfies the following recurrence: for $|I| < 2r$,

$$deg_G(I) = \frac{1}{2r - |I|} \cdot \sum_{j \notin I, \ j \text{ adjacent to all of } I} deg_G(I \cup \{j\}) = \frac{1}{2r - |I|} \sum_{j \notin I} deg_G(I \cup \{j\}).$$

The above two equations imply that $\mathcal{M}$ satisfies the second equation in 6.3. $\qquad\square$

Thus, to prove our main theorem [Theorem 1.5](), it suffices to show that $\mathcal{M}$ as defined above is PSD with high probability. This is equivalent to showing that the following matrix $M \equiv M_G \in \mathbb{R}^{\binom{[n]}{\leq r} \times \binom{[n]}{\leq r}}$ is PSD with high probability for $G \leftarrow G(n, 1/2)$: for $I, J \in \binom{[n]}{\leq r}$,

$$M(I, J) = deg_G(I \cup J) \cdot \frac{\binom{k}{|I \cup J|}}{\binom{2r}{|I \cup J|}}. \tag{6.5}$$

We will show that M is PSD for $k \leq \sqrt{n}/(C \log n)^{r^2}$ over the next two sections:

Lemma 6.3 (Main Technical Lemma). *With high probability over $G \leftarrow G(n, 1/2)$, the matrix M_G defined by [Equation 6.5]() is PSD for $k \leq \sqrt{n}/(C \log n)^{d^2}$.*

7 PSD'ness of the expectation matrix

As first steps and as a warmup for the final proof we first show that the expectation matrix $\mathbb{E}[M]$ is PSD. We start by writing down $E \equiv \mathbb{E}[M]$.

Claim 7.1. *For $I, J \in \binom{[n]}{\leq r}$, and $E \equiv \mathbb{E}[M]$,*

$$E(I, J) = \binom{n - |I \cup J|}{2r - |I \cup J|} \cdot \frac{\binom{k}{|I \cup J|}}{\binom{2r}{|I \cup J|}} \cdot 2^{-\binom{2r}{2}}. \tag{7.1}$$

Proof. The claim follows from observing that for any set K, $\mathbb{E}[deg_G(K)] = \binom{n-|I|}{2r-|I|} \cdot 2^{-\binom{2r}{2}}$. For, all K, there are $\binom{n-|K|}{2r-|K|}$ sets of size $2r$ containing K, and each is a clique with probability $2^{-\binom{2r}{2}}$. $\qquad\square$

Note that, the expectation matrix above is just a scalar multiple of $\mathcal{M}_{Gr}$ (viewed as a matrix) as defined in [Equation 6.1](). Therefore, by [Theorem 6.1](), E as defined above is PSD for $r < k$. We give a simpler proof of this claim here. We then build on these ideas to get lower bounds not only for the least eigenvalue of E, but for all of them, which will be critical for what follows.

The main claim of the section is the following.

Theorem 7.2. *The matrix E is PSD for $r < \min(k, n - k)$.*

As E is singular, one needs to exhibit extra care while trying to show E is PSD. Fortunately, the kernel of E has a simple structure and we exploit this. The next lemma is proved in [GHP02], and we repeat the simple proof here as we will also use a similar argument later.

Lemma 7.3. *The kernel of E has dimension at least $\sum_{i=0}^{r-1} \binom{n}{i}$.*

Proof. For any $J \subseteq [n]$, $|J| < r$, let $f_J \in \mathbb{R}^{\binom{[n]}{\leq r}}$ be the vector defined by

$$f_J(K) = \begin{cases} 1 & K = J \cup \{j\}, j \notin J \\ (|J| - k) & K = J \\ 0 & \text{otherwise} \end{cases}.$$

We claim that the vectors $\{f_J : J \subseteq [n], |J| < r\}$ are (1) in the kernel of E and (2) are linearly independent.

Proof of (1): For any f_J and K, $|K| \leq r$,

$$(E \cdot f_J)(I) = \sum_{K:|K| \leq r} E(I,K) f_J(K) = (|J| - k) E(I,J) + \sum_{j \notin J} E(I, J \cup \{j\})$$

$$= (|J| - k) E(I,J) + \sum_{j \in I \setminus J} E(I,J) + \sum_{j \notin (I \cup J)} E(I, J \cup \{j\})$$

$$\text{(E(I,J) only depends on } I \cup J.\text{)}$$

$$= (|I \cup J| - k) E(X_{I \cup J}) + \sum_{j \notin (I \cup J)} E(x_j \cdot X_{I \cup J}),$$

where in the last equation we used the interpretation of E as a linear mapping. It now follows that f_J is in the kernel as, E satisfies the relation

$$0 = E\left(\left(\sum_i x_i - k \right) X_{I \cup J} \right) = \sum_i E(x_i X_{I \cup J}) - kE(X_{I \cup J})$$

$$= (|I \cup J| - k) E(X_{I \cup J}) + \sum_{j \notin I} E(X_{I \cup J \cup j}).$$

Proof of (2): Look at the vectors $((f_I))$ in decreasing order of the sizes $|I|$. Then, each vector f_I introduces a new non-zero coordinate, $f_I(I)$, that was not covered before. Thus the collection of vectors is linearly independent. $\qquad\square$

Now that we have a handle on the kernel of E, we use a trick due to Laurent [Lau03] to show PSD'ness of E. Let E_r be the principal sub-matrix of E corresponding to the subsets of size exactly r. We will show that E_r is positive definite and conclude that E is PSD by appealing to the interlacing eigenvalues theorem, Lemma 3.1. Technically, the proof of the next lemma is where we differ in the proof Theorem 7.2 from similar claims in [Lau03], [GHP02] who use various results about hyper-geometric series.

Lemma 7.4. *For $r < \min(k, n - k)$, the matrix E_r is positive definite.*

Proof. We will show this by writing E_r as a suitable positive linear combination of the PSD matrices P_t's from Section 3. More concretely, for any $\alpha_0, \ldots, \alpha_t > 0$, we have

$$0 \prec \sum_t \alpha_t P_t = \sum_{\ell=0}^r \left(\sum_{t=0}^\ell \alpha_t \binom{\ell}{t} \right) D_\ell.$$

Now, let $e_\ell = E(X(I))$ for any I, $|I| = 2r - \ell$, i.e.,

$$e_\ell = 2^{-\binom{2r}{2}} \cdot \binom{n - 2r + \ell}{d - 2r + \ell} \cdot \frac{\binom{k}{2r-\ell}}{\binom{d}{2r-\ell}}.$$

Then, $E = \sum_{\ell=0}^{r} e_\ell D_\ell$. Therefore, we will be done if we can find α_t's such that for every $0 \le \ell \le r$, $e_\ell = \sum_{t=0}^{\ell} \alpha_t \binom{\ell}{t}$. By examining the first values of ℓ, it is easy to guess what the α_t should be. First observe that $e_\ell = e_0 \cdot \binom{n-2r+\ell}{\ell}/\binom{k-2r+\ell}{\ell}$ and let $\alpha_t = e_0 \binom{n-k}{t}/\binom{k-2r+t-1}{t}$. Then,

$$e_0 \binom{n - 2r + \ell}{\ell} = e_0 \sum_{t=0}^{\ell} \binom{n - k}{t} \cdot \binom{k - 2r + \ell - 1}{\ell - t}$$

$$= \sum_{t=0}^{\ell} \alpha_t \cdot \binom{k - 2r + t - 1}{t} \binom{k - 2r + \ell - 1}{\ell - t}$$

$$= \sum_{t=0}^{\ell} \alpha_t \cdot \binom{\ell}{t} \cdot \binom{k - 2r + \ell}{\ell}.$$

Therefore, $e_\ell = \sum_t \binom{\ell}{t} \alpha_t$ and the lemma now follows:

$$E_r = \sum_{\ell=0}^{r} e_\ell D_\ell = \sum_{\ell=0}^{r} \left(\sum_{t=0}^{\ell} \alpha_t \binom{\ell}{t} \right) D_\ell \succeq \alpha_r \mathbb{I}.$$

$\square$

Proof of Theorem 7.2. From the previous claim and Lemma 3.1, E has at least $\binom{n}{r}$ eigenvalues which are positive. The claim now follows by Lemma 7.3. $\square$

8 PSD'ness of dual certificate

We are now ready to prove our main result, Theorem 1.5, by showing that the mapping M will be PSD with high probability (Theorem 6.3). This will bring together the tools developed in the previous sections.

As it turns out, the matrix M is a little unwieldy because of high variance entries. We will instead work with a new matrix M'_r such that $M'_r \succeq 0$ implies M is PSD. We then show PSD'ness of M'_r by applying the high-level approach from the introduction to M'_r (instead of M as sketched originally).

In Section 8.1 we define our main object, the matrix M'_r, a variant of the principal minor of M corresponding to subsets of size exactly r. In Section 8.2.1 we show how to write M'_r as a matrix polynomial in independent, uniform $\{1, -1\}$-valued random variables. In Section 8.2.2 we compute the eigenvalues of the expectation of M'_r. In Section 8.3 we estimate the entries of the variance matrix of M'_r, and in Section 8.2.3 we use the estimates to bound

19

the eigenvalues of the variance. It will be clear from these bounds that the corresponding eigenvalues of the expectation and variance are roughly within a factor of $(k^2)/n$. We put all the pieces together to show the PSD'ness of dual certificate in Sections 8.2.4 and 8.4.

8.1   Reduction to positive definiteness of M_r'

As was done in Section 7, we will lower bound the dimension of kernel of M and then find a large principal submatrix which is positive definite. To this end, let M_r be the principal submatrix of M corresponding to the rows and columns of size exactly r.

We shall use the following notations for brevity: For any set $I \subseteq [n]$, let $\mathcal{E}(I) = \{\{i,j\} : i \neq j \in I\}$. For $0 \leq i \leq r$, let $k(i) = \binom{k}{2r-i}/\binom{2r}{2r-i}$.

A problem with the matrix M_r is that the diagonal entries have very large variance: The expected value of each diagonal entry is roughly $\Theta_r(n^r k^r)$. However, with probability $\Theta_r(1)$, the actual entry will be zero (if the corresponding set is not a clique) which in turn will also lead to the entire row and column being zero as well. This leads to strong technical difficulties. To overcome this, we study a closely related matrix which does not have the same variance issues.

For every $T \subseteq [n]$, let $M_T \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$, with $M_T(I,J) = k(|I \cap J|)$ if $I \cup J \subseteq T$, and G contains every edge in $\mathcal{E}(T) \setminus \mathcal{E}(I) \cup \mathcal{E}(J)$ (i.e., the only edges in T missing in G are those with both end points in one of I or J). We will study the matrix

$$M_r' = \sum_{T:|T|=2r} M_T. \tag{8.1}$$

Intuitively, for every I, J, $M_r'(I,J)$ is what $M(I,J)$ would be had we added cliques on the subsets I, J to the graph. The above definition avoids the problem of the whole row and column corresponding to I or J becoming zero if either was not a clique, and controls the variance.

We now reduce PSD'ness of M to that of M_r'.

Lemma 8.1. *For any graph G, the kernel of M_G has dimension at least $\sum_{\ell=0}^{r-1} \binom{n}{\ell} + |\{I : |I| = r, I \text{ not a clique in } G\}|$.*

Compared to Lemma 7.3, the difference is the last additive term.

Proof. The same proof as in Lemma 7.3 shows that the vectors $\{f_J : |J| < r\}$ defined there also lie in the kernel of M_G. The reason is that the proof of Lemma 7.3 only relied on the equations defining the dual certificate - last equation in 6.3, which M_G also satisfies.

Let $N = \sum_{\ell}^{r} \binom{n}{\ell}$ and for a set I, let $e_I \in \mathbb{R}^N$ be the vector with 1 in the I'th coordinate and 0's elsewhere. Observe that if a set $I, |I| \leq r$, is not a clique then, the I'th row and I'th column of M_G are identically zero. Therefore, for every such vector I, e_I is in the kernel of M_G.

Let $V = \{f_J : |J| < r\} \cup \{e_J : |J| = r, J \text{ not a clique in } G\}$. We claim that the vectors in V are linearly independent. Look at the vectors in V in decreasing order of the sizes

20

$|I|$. Then, each vector $f \in V$ introduces a new non-zero coordinate (corresponding to the set describing f) that was not covered before. Thus the collection of vectors is linearly independent. $\square$

Lemma 8.2. *If M'_r is positive definite, then the mapping M is PSD.*

Proof. First note that, for any I, J, $M_r(I, J) = M'_r(I, J)$ if I and J are cliques in G and $M_r(I, J) = 0$ otherwise. For, suppose that I and J are cliques in G. Then, $M_T(I, J) = k(|I \cap J|)$ if $I \cup J \subseteq T$ and T is a clique and 0 otherwise. Therefore,

$$M'_r(I, J) = \sum_T M_T(I, J) = k(|I \cap J|) \cdot |\{T : I \cup J \subseteq T, T \text{ clique}\}| = M_r(I, J).$$

Therefore, the principal submatrix of M_r corresponding to cliques of size exactly r is the same as in M'_r and hence is positive definite by the assumption. Thus, if we let N_r be the number of cliques of size exactly r in G, then by Lemma 3.1, M_r and hence M, has at least N_r positive eigenvalues. The claim now follows from the fact that the kernel has dimension at least $\sum_{\ell=0}^{r} \binom{n}{\ell} - N_r$ by the previous lemma. $\square$

8.2 Positive definiteness of M'_r

We now show that the M'_r defined in the last section is positive definite with high probability.

Lemma 8.3 (Main Technical Lemma). *For c a sufficiently large constant the following holds. The matrix $M'_r \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ defined by Equation 8.1 is positive definite with probability at least $1 - \delta$, for $k < \sqrt{n}/(c \log(n^r/\delta))^{r^2}$.*

The proof follows the two-step strategy described in the introduction: show that $E[M'_r] - M'_r \prec E[M'_r]$ with high probability. To do so, we first write $M'_r - E[M'_r]$ as a matrix polynomial in independent random variables. We then apply our tail bounds for matrix polynomials, Theorem 1.6, to argue that $E[M'_r] - M'_r \prec \Delta \cdot \Sigma$, where Δ is not too large and Σ is the *variance* of $M'_r - E[M'_r]$. Finally, we exploit the set-symmetry of the *variance* and *expectation* matrices $\Sigma, E[M'_r]$ to characterize their eigenvalues using the theory of association schemes. Using these estimates we conclude that $\Delta\Sigma \prec E[M'_r]$, thus proving the theorem. While the entries of $E[M'_r]$ are easy to get a handle on, estimating the eigenvalues of Σ requires some careful calculations to estimate the entries of Σ.

For the remainder of this section, we shall use the following additional notations:

- For $0 \leq i \leq r$, let

$$k'(i) = k(i)/2^{r^2 + \binom{i}{2}} = \binom{k}{2r-i}/\binom{2r}{2r-i} \cdot 2^{r^2 + \binom{i}{2}}.$$

- In the following we will adopt the convention that I, J, K denote elements of $\binom{[n]}{r}$ and T, T' denote elements of $\binom{[n]}{2r}$.

- In the following H denotes a graph on at most $2r$ vertices from $[n]$.

- All matrices considered below will be over $\mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$.

- We write $A \approx_r B$ if there exist constants c, C such that $c^{r^2} B \leq A \leq C^{r^2} B$.

8.2.1 Writing M'_r as a matrix polynomial

We first write M'_r as a matrix polynomial in independent random variables. For every edge $e \in \binom{[n]}{2}$, let y_e denote the indicator random variable that is 1 if $e \in G$ and 0 otherwise. Let $\varepsilon_e = 2y_e - 1$. Then, by Equation 8.1,

$$M'_r(I, J) = k(|I \cap J|) \cdot \sum_T \left(\prod_{e \in \mathcal{E}(T) \setminus \mathcal{E}(I) \cup \mathcal{E}(J)} y_e \right)$$

$$= k(|I \cap J|) \cdot \sum_T \left(\prod_{e \in \mathcal{E}(T) \setminus \mathcal{E}(I) \cup \mathcal{E}(J)} \frac{1 + \varepsilon_e}{2} \right).$$

For any I, J, T with $I \cup J \subseteq T$, let $H(I, J, T)$ denote all graphs whose edges are contained in $\mathcal{E}(T) \setminus \mathcal{E}(I) \cup \mathcal{E}(J)$. Note that $|\mathcal{E}(T) \setminus \mathcal{E}(I) \cup \mathcal{E}(J)| = r^2 + \binom{|I \cap J|}{2}$. Then, we can expand the above equation further as

$$M'_r(I, J) = \frac{k(|I \cap J|)}{2^{r^2 + \binom{|I \cap J|}{2}}} \cdot \sum_T \sum_{H \in H(I,J,T)} \left(\prod_{e \in H} \varepsilon_e \right) = k'(|I \cap J|) \cdot \sum_T \sum_{H \in H(I,J,T)} \left(\prod_{e \in H} \varepsilon_e \right)$$

$$= k'(|I \cap J|) \sum_H \left(\prod_{e \in H} \varepsilon_e \right) \left(\sum_{T: H(I,J,T) \ni H} 1 \right).$$

Now, for any graph H on at most $2r$ vertices, define the matrix A_H where

$$A_H(I, J) = k'(|I \cap J|) \cdot |\{T : H \in H(I, J, T)\}|. \tag{8.2}$$

Then, from the above two equations

$$M'_r = \sum_H \left(\prod_{e \in H} \varepsilon_e \right) A_H. \tag{8.3}$$

Note that the right hand side is a matrix polynomial in independent, uniformly random $\{1, -1\}$ variables $\{\{\varepsilon_e\}\}$; a setting, where we can apply Theorem 1.6. But to do so, we need to first understand the expectation matrix $\mathbb{E}[M'_r]$ and the variance matrix

$$\Sigma^2 = \sum_{H: H \neq \emptyset} A_H^2. \tag{8.4}$$

Note that both $\mathbb{E}[M'_r]$ and Σ are clearly set-symmetric. Thus, we can apply the results of Section 4.

22

8.2.2 Estimating the eigenvalues of the expectation matrix

We now estimate the spectrum of $E[M'_r]$. In the following, we recall the notation from Lemma 4.6.

Lemma 8.4. *If $k < n/2^{cr^2}$ for a sufficiently large constant c, then, for $0 \leq j \leq r$, $\lambda_j(\mathbb{E}[M'_r]) \geq k^{2r-j}n^r/2^{O(r^2)}$.*

Proof. Let $Q = \mathbb{E}[M'_r]$. Then, by Equation 8.1, $Q(I, J) = k'(|I \cap J|) \cdot \binom{n-|I \cup J|}{2r-|I \cup J|}$ as there are $\binom{n-|I \cup J|}{2r-|I \cup J|}$ many T's that contain I, J and every such T contributes to $M_T(I, J)$ by an additive $k'(|I \cap J|)$ amount. Therefore, $Q = \sum_\ell \alpha_\ell D_\ell$, where $\alpha_\ell \approx_r k^{2r-\ell}n^\ell$. We next express Q as a linear combination of P_t's: $Q = \sum_t \beta_t P_t$. Then, by Claim 4.5,

$$\beta_t = \sum_{\ell \leq t}(-1)^{t-\ell}\binom{t}{\ell}\alpha_\ell.$$

Now, if $k < n/2^{O(r^2)}$ for a sufficiently big constant, then α_ℓ's increase geometrically and the above sum will be dominated by the last term so that $\beta_t \geq \alpha_t/2 \approx_r k^{2r-t}n^t$. Therefore, by Lemma 4.6 and the fact that P_t's are PSD,

$$\lambda_j(Q) = \sum_t \beta_t\lambda_j(P_t) = \sum_{t \geq j}\beta_t\binom{n-t-j}{r-t}\binom{r-t}{t-j}$$
$$\geq 2^{-O(r^2)}\sum_{t \geq j}k^{2r-\ell}n^t \cdot n^{r-t} \geq k^{2r-j}n^r/2^{O(r^2)}.$$

$\square$

8.2.3 Estimating the eigenvalues of the variance matrix

We now estimate the spectrum of Σ^2. Again, we use the notation from Lemma 4.6.

Lemma 8.5. *For Σ as defined in Lemma 8.6, $\lambda_j(\ )$ defined as in Lemma 4.6, if $k^2 < n$, then $\lambda_j(\Sigma) \leq 2^{O(r^2)}k^{2r-j}n^r(k/\sqrt{n})$.*

The proof of the above lemma involves first estimating the entries of Σ up to factors depending only on r.

Lemma 8.6. *Let Σ be defined by Equation 8.4. Then, Σ is symmetric and if $k < \sqrt{n}/2$, the following holds: For any I, J, $\Sigma^2(I, J) \approx_r \left(n^r k^{4r}\left(\frac{n}{k^2}\right)^{|I \cap J|-1}\right)$.*

The proof of the lemma involves a somewhat tedious calculation. We remark that the -1 appearing in the exponent of (n/k^2) above will be significant. We defer the proof of the lemma for now and proceed with the proof of Lemma 8.5.

Proof of Lemma 8.5. Note that Σ^2 is set-symmetric and hence has eigenspaces $V_0, \ldots, V_r$ as in Lemma 4.6. Further by Lemma 8.6, if we write $\Sigma^2 = \sum_\ell \alpha_\ell D_\ell$, then $\alpha_\ell \approx_r n^r k^{4r} (n/k^2)^{\ell-1}$. Further, if $\beta_t = \sum_{\ell \leq t} \binom{t}{\ell} \alpha_\ell$, then as $k^2 < n$,

$$\beta_t \approx_r n^r k^{4r} \sum_{\ell \leq t} (n/k^2)^{\ell-1} \approx_r n^r k^{4r} (n/k^2)^{t-1}.$$

Therefore, by Lemma 4.7, we get that for $0 \leq j \leq r$,

$$\lambda_j(\Sigma^2) \leq \sum_{t \geq j} \beta_t \cdot \binom{n-t-j}{r-t} \cdot \binom{r-j}{t-j}$$

$$\leq C_r \sum_{t \geq j} n^r k^{4r} \left(\frac{n}{k^2}\right)^{t-1} \cdot n^{r-t}$$

$$\leq C_r n^{2r-1} k^{4r} \sum_{t \geq j} \frac{1}{k^{2(t-1)}} \leq C_r n^{2r-1} k^{4r} \cdot \frac{O(1)}{k^{2(j-1)}}$$

$$\leq C_r k^{4r-2j} n^{2r} \left(\frac{k^2}{n}\right),$$

where $C_r = 2^{O(r^2)}$. The claim now follows by noting that $\lambda_j(\Sigma) = \sqrt{\lambda_j(\Sigma^2)}$. $\qquad \square$

We defer the proof of Lemma 8.6 till the end of this section.

8.2.4 Positive definiteness of M'_r

We are now ready to show positive definiteness of M'_r and prove Lemma 8.3.

Proof of Lemma 8.3. By Equation 8.3,

$$M'_r = \mathbb{E}[M'_r] + \sum_{H:H \neq \emptyset} \left(\prod_{e \in H} \varepsilon_e\right) A_H.$$

We now apply Theorem 1.6 to the matrix polynomial from the second term. Note that the degree of the polynomial is at most $(2r)^2/2 = 2r^2$. Fix $\delta > 0$, and let $N = \binom{n}{r}$. Then, for C large enough, and $\Delta = (C \log(N/\delta))^{r^2}$, by Theorem 1.6, with probability at least $1 - \delta$,

$$M'_r \succeq \mathbb{E}[M'_r] - \Delta \left(\sum_H A_H^2\right)^{1/2} = \mathbb{E}[M'_r] - \Delta\Sigma.$$

Let $Q = \mathbb{E}[M'_r] - \Delta\Sigma$. Then, Q is set-symmetric and by Lemmas 8.4 and 8.5, for $0 \leq j \leq r$,

$$\lambda_j(Q) \geq \lambda_j(\mathbb{E}[M'_r]) - \Delta \cdot \lambda_j(\Sigma)$$

$$\geq k^{2r-j} n^r / 2^{O(r^2)} - (C \log(N/\delta))^{r^2} \cdot k^{2r-j} n^r (k/\sqrt{n}) \cdot 2^{O(r^2)}$$

$$= \frac{k^{2r-j} n^r}{2^{O(r^2)}} \cdot \left(1 - \left(\frac{k}{\sqrt{n}}\right) \cdot 2^{O(r^2)} \cdot (C \log(N/\delta))^{r^2}\right).$$

24

Therefore, if $k < \sqrt{n}/(c \log(N/\delta))^{r^2}$ for a sufficiently large constant c, the last expression above will be positive so that $\lambda_j(Q) > 0$. Thus, $M'_r \succeq Q \succ 0$, proving the lemma. $\qquad \square$

8.3 Estimating the entries of the variance matrix

We now prove Lemma 8.6.

Proof of Lemma 8.6. Σ is clearly set-symmetric as we sum over all possible graphs H. We have,

$$\Sigma^2(I, J) = \sum_{H:H\neq\emptyset} A_H^2(I, J)$$

$$= \sum_{H:H\neq\emptyset} \sum_K A_H(I, K) A_H(J, K)$$

$$= \sum_{H:H\neq\emptyset} \sum_K k'(|I \cap K|)\, k'(|J \cap K|) \cdot |\{T : H(I, K, T) \ni H\}| \cdot |\{T' :\in H(J, K, T') \ni H\}|,$$

where the last equality follows from the definition of A_H from Equation 8.2. Continuing with the above computation, we have

$$\Sigma^2(I, J) = \sum_{H:H\neq\emptyset} \sum_K k'(|I \cap K|) \cdot k'(|J \cap K|) \left(\sum_{T:H\in H(I,K,T)} 1 \right) \left(\sum_{T':H\in H(J,K,T)} 1 \right)$$

$$= \sum_{T,T'} \sum_K k'(|I \cap K|) \cdot k'(|J \cap K|) \left(\sum_{H:H\in H(I,K,T)\cap H(J,K,T')} 1 \right)$$

(Interchanging the order of summation)

$$= \sum_{T,T'} \sum_K k'(|I \cap K|) \cdot k'(|J \cap K|) \cdot |\{H \neq \emptyset : H \in H(I, K, T) \cap H(J, K, T')\}|$$

Now, the last term above counts the number of non-empty graphs in $H(I, K, T) \cap H(J, K, T')$. First observe that there exist such graphs only if $(I \cup K) \subseteq T, (J \cup K) \subseteq T'$. In particular, only if $I \subseteq T, J \subseteq T'$, and $K \subseteq T \cap T'$. Further, any graph in $H(I, K, T) \cap H(J, K, T')$ must have its vertex set contained in $T \cap T'$. Thus, for fixed T, T', the number of such graphs is at most $2^{O(r^2)}$. We will essentially just use this trivial bound but for one important exception when $|T \cap T'| = r$. We claim that when $|T \cap T'| = r$, there are no non-empty graphs in $H(I, K, T) \cap H(J, K, T')$ so that the inner summand is actually 0. For, if $H \in H(I, K, T) \cap H(J, K, T')$, then all edges of H must (1) be in $T \cap T'$, and (2) cannot be in $\mathcal{E}(K)$ (by definition). These two conditions are contradictory as if $|T \cap T'| = r$ and $K \subseteq T \cap T'$, then $K = T \cap T'$.

Therefore, combining the above arguments, we get

$$\Sigma^2(I, J) = 2^{O(r^2)} \sum_{\substack{T,T':|T\cap T'|>r \\ I\subseteq T, J\subseteq T'}} \sum_{K\subseteq T\cap T'} k'(|I \cap K|) \cdot k'(|J \cap K|). \tag{8.5}$$

25
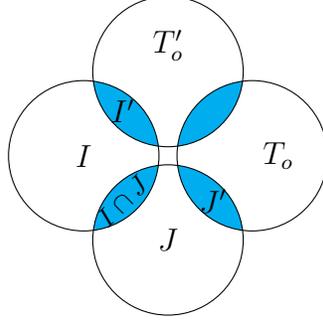
Figure 1: Enumerating over the sets $T = I \cup T_0$ and $T' = J \cup T'_o$. R is the blue shaded region and K should be in R. Notation: $|R| = r + e$, $|I'| = i$, $|J'| = j$.

We next estimate the last summation which is somewhat tricky. Our final goal will be to get an appropriate geometrically decreasing series parameterized by $|T \cap T'|$. We then exploit the fact that we only sum over T, T' with $|T \cap T'| > r$.

We will choose T, T' as follows. Let $R = T \cap T'$. Then, $I \cap J \subseteq R$. Now, pick T, T' by picking in order (see Figure 8.3): $I' = (R \cap I) \setminus (I \cap J)$, $J' = (R \cap J) \setminus (I \cap J)$, $R \setminus (I \cup J)$, $T \setminus (R \cup I)$ and $T' \setminus (R \cup J)$. See Figure 8.3 for a pictorial representation: R is the blue shaded region, T is the two circles on the horizontal axis and T' is the two circles on the vertical axis.

We will use the following notations for brevity:

- $R = T \cap T'$, $|R| = r + e$. Recall that $e \geq 1$.

- $|I \cap J| = m$, $|I'| = i$ and $|J'| = j$.

We will estimate the right hand side of Equation 8.5 by parametrizing over $1 \leq e \leq r$, and $0 \leq i, j \leq r$. To this end, first note that $K \subseteq R$ and $|K| = r$. Therefore,

$$|I \cap K| + |J \cap K| \geq 2|K \cap I \cap J| \geq 2\left(|I \cap J| - e\right) = 2(m - e).$$

Hence,

$$k'(|I \cap K|) \cdot k'(|J \cap K|) \leq k^{2r-|I \cap K|} k^{2r-|J \cap K|} \leq k^{4r-2m+2e}.$$

Thus, if we fix T, T' and sum over all possible $K \subseteq T \cap T'$, the inner summand from Equation 8.5 becomes

$$\sum_{K \subseteq T \cap T'} k'(|I \cap K|) \cdot k'(|J \cap K) \leq 2^{2r} k^{4r-2m+2e}. \tag{8.6}$$

We have to now enumerate the number of ways of choosing T, T'. If we fix e, i, j as above, then the number of sets T, T' achieving these sizes is exactly the number of ways to choose $I' \subseteq I \setminus J$, $J' \subseteq J \setminus I$, $R \setminus (I \cup J)$, $T \setminus (R \cup I)$, $T' \setminus (R \cup J)$. Further,

$$|R \setminus (I \cup J)| = r + e - |I \cap J| - |I'| - |J'| = r + e - m - i - j,$$
$$|T \setminus (R \cup I)| = 2r - |R \cup I| = 2r - (2r + e - m - i) = m + i - e,$$
$$|T' \setminus (R \cup J)| = 2r - |R \cup J| = 2r - (2r + e - m - j) = m + j - e.$$

26

Thus, the number of ways to pick T, T' is at most (again, this is easier to see pictorially from Figure 8.3)

$$\binom{I}{i} \cdot \binom{J}{j} \cdot \binom{n}{r + e - m - i - j} \cdot \binom{n}{m + i - e} \cdot \binom{n}{m + j - e} \leq$$
$$2^{2r} \cdot n^{r+e-m-i-j} \cdot n^{m+i-e} \cdot n^{m+j-e} = 2^{2r} n^{r+m-e}. \quad (8.7)$$

Combining Equations 8.5, 8.6, 8.7, we get

$$\Sigma^2(I, J) = 2^{O(r^2)} \sum_{\substack{1 \leq e \leq r, \\ 0 \leq i, j \leq r}} 2^{4r} k^{4r - 2m + 2e} \cdot n^{r + m - e}$$

$$= 2^{O(r^2)} \cdot k^{4r - 2m} \cdot n^{r + m} \sum_{e \geq 1} \left(\frac{k^2}{n} \right)^e$$

$$\leq 2^{O(r^2)} \cdot k^{4r - 2m} \cdot n^{r + m} \left(\frac{k^2}{n} \right)$$

$$= 2^{O(r^2)} k^{2r} n^{2r} \left(\frac{k^2}{n} \right)^{r + 1 - m},$$

where we used the fact that if $k < \sqrt{n}/2$ we have a geometrically decreasing series. The claim now follows. $\qquad \square$

8.4 Putting things together

We bring the arguments from previous sections together to prove our main results Theorem 6.3 and Theorem 1.5.

Proof of Theorem 6.3. Follows immediately from Lemma 8.2 and Lemma 8.3. $\qquad \square$

Proof of Theorem 1.5. Follows immediately from Lemma 2.3, Claim 6.2 and Theorem 6.3. $\qquad \square$

Theorems 1.1 and 1.2 follow immediately from our $\mathsf{PS}(r)$-refutation lower bound using standard arguments. We defer these to the appendix.

9 Conclusion and future work

In this work we showed a lower bound for the maximum clique problem on random $G(n, 1/2)$ graphs in the SOS hierarchy and positivstellensatz proof system. Besides the specific application to clique lower bounds, the PSD'ness of the matrix M from Equation 6.5 seems to carry further information that could be potentially useful elsewhere, perhaps for studying various sub-graph statistics. Further, the arguments related to association schemes and non-commutative polynomial tail inequalities could also be useful elsewhere, especially for other

SOS hierarchy lower bounds. One natural and interesting candidate is the densest subgraph problem.

The most obvious open problem is to solve the planted clique problem for $k = n^{1/2-\delta}$ for any constant $\delta > 0$ in polynomial time. As our results show, this may require some very new techniques.

Another obvious open problem arising from this work is to close the gap between the upper and lower bounds for the SOS hierarchy in terms of number of rounds, r: $1/2^{O(r^2)}$ vs $1/2^r$. Concretely, we could ask if $r = O(\sqrt{\log n})$ rounds of SOS hierarchy can beat the $\sqrt{n}$ barrier for planted clique. In our current arguments, we lose factors of $2^{O(r^2)}$ mainly in two places:

1. In Sections 8.2.2 and 8.2.3, the latter of which in turn comes from the estimates in Section 8.3.

2. In Section 8.2.1, when we apply the tail inequality for matrix polynomials to a polynomial of degree r^2.

We believe that the losses in (1) can be improved to losses of $2^{O(r)}$ with more tenuous calculations. In fact, the improvements are not too hard for the bounds in Section 8.2.2.

On the other hand, the losses coming from (2) seem more critical and intrinsic to the current analysis. Large deviation inequalities of the form we use in Section 8.2.1 need to lose an exponential factor in the degree even for the scalar case. However, if the polynomial has a special form, then better large deviation bounds may be possible. For example consider the polynomial $P(x_1, \ldots, x_n) = x_1 x_2 \cdots x_d + x_{d+1} x_{d+2} \cdots x_{2d} + \cdots + x_{n-d+1} x_{n-d+2} \cdots x_n$. Then, using tail inequalities for degree d polynomials gives you losses of $exp(d)$, however P is effectively a linear polynomial and hence much stronger bounds independent of the degree d hold. Intuitively, the worst degree d polynomials in terms of large deviation bounds are polynomials which can be expressed as powers (or other functions) of few linear polynomials and hence are quite "structured". The work of Latala [Lat06] (also see [Kan11] for $\{1, -1\}$-valued random variables) in fact gives tight large deviation bounds for (scalar) polynomials taking into account the *structure* of the specific polynomial. One plausible avenue for improving the bounds in Section 8.2.1 is to see if the polynomial we study has any special structure (or lack thereof).

Another question is to remove the dependence on the number of variables n from our non-commutative polynomial tail inequality, Theorem 1.6, which seems unnecessary.

References

[ABS10]   Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *FOCS*, pages 563–572, 2010.

[AKS98]   Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Struct. Algorithms*, 13(3-4):457–466, 1998.

[Art27]   Emil Artin. Uber die zerlegung deniter funktionen in quadrate. *Abhandlungen aus dem Mathematischen Seminar der Universitat Hamburg*, 5(1):100–115, 1927.

[AW02]   Rudolf Ahlswede and Andreas Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

[BBH+12]   Boaz Barak, Fernando G. S. L. Brandão, Aram Wettroth Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *STOC*, pages 307–326, 2012.

[BCR98]   Jacek Bochnak, Michel Coste, and Marie-Francoise Roy. *Real algebraic geometry*. Springer, 1998.

[BCV+12]   Aditya Bhaskara, Moses Charikar, Aravindan Vijayaraghavan, Venkatesan Guruswami, and Yuan Zhou. Polynomial integrality gaps for strong SDP relaxations of densest k-subgraph. In *SODA*, pages 388–405, 2012.

[BRS11]   Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *FOCS*, pages 472–481, 2011.

[dlPnMS95]   Victor H. de la Peña and S.J. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate U- statistics. *Ann. Probab.*, 23(2):806–816, 1995.

[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, 2013.

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

[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.

[GHP02]   Dima Grigoriev, Edward Hirsch, and Dmitrii. Pasechnik. Complexity of semialgebraic proofs. In *STACS 2002*, volume 2285, pages 419–430. Springer Berlin Heidelberg, 2002.

[God]   Chris Godsil. Association schemes. Lecture Notes available at http://quoll.uwaterloo.ca/mine/Notes/assoc1.pdf.

[Gri01a]   Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. *Computational Complexity*, 10(2):139–154, 2001.

[Gri01b]   Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001.

[GS11]   Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with PSD objectives. In *FOCS*, pages 482–491, 2011.

[GV01]   Dima Grigoriev and Nicolai Vorobjov. Complexity of null-and positivstellensatz proofs. *Ann. Pure Appl. Logic*, 113(1-3):153–160, 2001.

[Jan97]   S. Janson. *Gaussian Hilbert Spaces*. Cambridge Tracts in Mathematics. Cambridge University Press, 1997.

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

[Kan11]   Daniel M. Kane. k-independent gaussians fool polynomial threshold functions. In *IEEE Conference on Computational Complexity*, pages 252–261, 2011.

[Kar76]   R. M. Karp. Probabilistic analysis of some combinatorial search problems. *In: Algorithms and Complexity: New Directions and Recent Results*, pages 1–19, 1976.

[Kri64]   Jean-Louis Krivine. Anneaux preordonnes. *Journal d'Analyse Mathematique*, 12(1):307–326, 1964.

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

[Las01]   Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[Lat06]   Rafal Latala. Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.*, 34(6):2315–2331, 2006.

[Lau03]   Monique Laurent. Lower bound for the number of iterations in semidefinite hierarchies for the cut polytope. *Mathematics of Operations Research*, 28(4):871–883, 2003.

[LP86]   F. Lust-Piquard. Inegalites de Khintchine dans C_p $(1 < p < \infty)$. *C. R. Acad. Sci. Paris*, 303:298–292, 1986.

[LPP91]   F. Lust-Piquard and G. Pisier. Non commutative Khintchine and Paley inequalities. *Ark. Mat*, 29(2):241–260, 1991.

[LS91] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

[O'D08] Ryan O'Donnell. Some topics in analysis of boolean functions. In *STOC*, pages 569–578, 2008.

[Oli10] Roberto Oliveira. Sums of random Hermitian matrices and an inequality by Rudelson. *Electron. Commun. Probab.*, 15(19):203–212, 2010.

[OZ13] Ryan O'Donnell and Yuan Zhou. Approximability and proof complexity. In *SODA*, pages 1537–1556, 2013.

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

[Put93] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

[SA90] H. Sherali and W. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990.

[Sch91] Konrad Schmudgen. The k-moment problem for compact semi-algebraic sets. *Mathematische Annalen*, 289(1):203–206, 1991.

[Sch08] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *FOCS*, pages 593–602, 2008.

[Ste73] Gilbert Stengle. A nullstellensatz and a positivstellensatz in semialgebraic geometry. *Mathematische Annalen*, 207(2):87–97, 1973.

[Tro12] JoelA. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

[Tul09] Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *STOC*, pages 303–312, 2009.

[Ver] Roman Vershynin. Lecture 6: Norm of a random matrix. Lecture Notes on Non-Asymptotic Random Matrix Theory - Available online.

[vLW01] J.H. van Lint and R.M. Wilson. *A Course in Combinatorics*. Cambridge University Press, 2001.

SOS-relaxation for Max-Clique. Input: Graph $G = (V, E)$, r - number of rounds. Variables of the SDP are vectors $\boldsymbol{U}_S$, where $S \subseteq [n], |S| \leq r$.

$$\text{maximize } \sum_{i \in V} \|\boldsymbol{U}_{\{i\}}\|_2^2.$$

such that $\quad \langle \boldsymbol{U}_{\{i\}}, \boldsymbol{U}_{\{j\}} \rangle = 0, \qquad\qquad \forall i, j, \ \{i, j\} \notin E$

$\qquad\qquad\quad \langle \boldsymbol{U}_{S_1}, \boldsymbol{U}_{S_2} \rangle = \langle \boldsymbol{U}_{S_3}, \boldsymbol{U}_{S_4} \rangle, \qquad S_1 \cup S_2 = S_3 \cup S_4, |S_1 \cup S_2| \leq r$

$\qquad\qquad\quad \langle \boldsymbol{U}_{S_1}, \boldsymbol{U}_{S_2} \rangle \in [0, 1], \qquad\qquad |S_1|, |S_2| \leq r$

$\qquad\qquad\quad \|\boldsymbol{U}_\emptyset\|_2^2 = 1$

Figure 2: r-round SOS-relaxation for Maximum Clique

A Hierarchy Gaps and Positivstellensatz Refutations

For a detailed discussion of the hierarchies and $\mathsf{PS}(r)$-refutations we refer the reader to the discussions in [OZ13]. The basic principle is that, typically, $\mathsf{PS}(r)$-refutations are more robust and stronger than the hierarchy formulations.

The SOS (or Lasserre) relaxation for maximum clique is stated in Figure A (cf. [Tul09]). Although, the formulation itself is not in terms of an SDP, it is a standard fact that as the program only involves inner products of vectors, the optimization can be done by semidefinite programming. The connection between Figure A and $\mathsf{PS}(r)$-refutations comes from the following straightforward lemma stating that a certificate for $\mathsf{PS}(r)$-refutations is simply a primal solution to the standard r-round SOS-relaxation of the problem.

Lemma A.1. *Let* $G = (V, E)$ *be a graph and let* $\text{Clique}(G, k)$ *denote the clique axioms as defined by Equations 1.2. Suppose that there exists a dual certificate* $\mathcal{M} : \mathcal{P}(n, 2r) \to \mathbb{R}$ *for* $\text{Clique}(G, k)$ *as defined in Definition 2.2. Then, the value of the r-round SOS-relaxation for maximum clique given by Figure A is at least k.*

Proof. Let $\mathcal{M} : \mathcal{P}(n, 2r) \to \mathbb{R}$ be the dual certificate and $M \in \mathbb{R}^{\binom{[n]}{\leq r} \times \binom{[n]}{\leq r}}$ be the corresponding PSD matrix. Without loss of generality suppose that $M(\emptyset, \emptyset) = 1$. Let $M = UU^\dagger$, where $U = \mathbb{R}^{\binom{[n]}{\leq r} \times N}$ for some N. Finally, for $S \in \binom{[n]}{\leq r}$, let $\boldsymbol{U}_S$ be the S'th row of U. We claim that the collection $(\boldsymbol{U}_S, |S| \leq r)$ gives a feasible solution for the SDP in Figure A.

Observe that for any two subsets $S_1, S_2 \in \binom{[n]}{\leq r}$,

$$\langle \boldsymbol{U}_{S_1}, \boldsymbol{U}_{S_2} \rangle = M(S_1, S_2) = \mathcal{M}(X_{S_1 \cup S_2}).$$

Therefore, the vectors $(\boldsymbol{U}_S : |S| \leq r)$ satisfy the first two constraints of Figure A as $\mathcal{M}$ is a dual certificate. Further, $\|\boldsymbol{U}_\emptyset\|^2 = M(\emptyset, \emptyset) = 1$ and for any set S,

$$\|\boldsymbol{U}_S\|_2^2 = \langle \boldsymbol{U}_S, \boldsymbol{U}_S \rangle = \langle \boldsymbol{U}_S, \boldsymbol{U}_\emptyset \rangle \leq \|\boldsymbol{U}_S\|_2,$$

so that $\|\boldsymbol{U}_S\| \leq 1$. Thus, $(\boldsymbol{U}_S: |S| \leq r)$ give a feasible solution for the program in Figure A. Finally, the value of the solution is

$$\sum_{i \in V} \|\boldsymbol{U}_{\{i\}}\|_2^2 = \sum_{i \in V} \mathcal{M}(X_{\{i\}}) = k.$$

This proves the lemma. $\qquad \square$

Our main theorems now follows.

Proof of Theorem 1.1. Let $G \leftarrow G(n, 1/2)$. Then, from the above lemma and the proof of Theorem 1.5 (where we showed the existence of a dual certificate for the clique axioms), the value of the r-round SOS-relaxation for max-clique on G is at least $\sqrt{n}/(C \log n)^{r^2}$ with high probability. The claim follows as the integral value is $(2 + o(1)) \log_2 n$ with high probability. $\qquad \square$

Proof of Corollary 1.2. The value of the relaxation in Figure A is clearly monotone with respect to adding edges. Therefore, from the above argument, for $G \leftarrow G(n, 1/2, t)$ the value of the r-round SOS-relaxation for max-clique on G is at least $\sqrt{n}/(C \log n)^{r^2}$ with high probability. The claim follows as the integral value is t with high probability. $\qquad \square$