

A Stress-Free Sum-Of-Squares Lower Bound for Coloring

Pravesh K. Kothari ✉

Carnegie Mellon University, Pittsburgh, PA, USA

Peter Manohar ✉

Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

We prove that with high probability over the choice of a random graph G from the Erdős-Rényi distribution $G(n, 1/2)$, a natural $n^{O(\varepsilon^2 \log n)}$ -time, degree $O(\varepsilon^2 \log n)$ sum-of-squares semidefinite program cannot refute the existence of a valid k -coloring of G for $k = n^{1/2+\varepsilon}$. Our result implies that the refutation guarantee of the basic semidefinite program (a close variant of the Lovász theta function) cannot be appreciably improved by a natural $o(\log n)$ -degree sum-of-squares strengthening, and this is tight up to a $n^{o(1)}$ slack in k . To the best of our knowledge, this is the first lower bound for coloring $G(n, 1/2)$ for even a single round strengthening of the basic SDP in any SDP hierarchy.

Our proof relies on a new variant of instance-preserving *non-pointwise complete reduction* within SoS from coloring a graph to finding large independent sets in it. Our proof is (perhaps surprisingly) short, simple and *does not* require complicated spectral norm bounds on random matrices with dependent entries that have been otherwise necessary in the proofs of many similar results [12, 33, 45, 28, 51].

Our result formally holds for a constraint system where vertices are allowed to belong to multiple color classes; we leave the extension to the formally stronger formulation of coloring, where vertices must belong to unique colors classes, as an outstanding open problem.

2012 ACM Subject Classification Theory of computation → Computational complexity and cryptography

Keywords and phrases Sum-of-Squares, Graph Coloring, Independent Set, Lower Bounds

Digital Object Identifier 10.4230/LIPIcs.CCC.2021.23

Related Version *Full Version*: <https://arxiv.org/abs/2105.07517>

Funding *Pravesh K. Kothari*: Supported by NSF CAREER Award #2047933.

Peter Manohar: NSF Graduate Research Fellowship Program and the ARCS Foundation. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1745016. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Acknowledgements We thank Xinyu Wu for taking part in early stages of this research, and the anonymous reviewers for providing valuable feedback.

1 Introduction

Starting with the seminal work of Arora, Bollobás, Lovász and Tzourakis [2], understanding the power of systematic hierarchies of linear and semidefinite programs for solving combinatorial optimization problems has been a foundational goal in complexity theory. This project has achieved many successes including sharp lower bounds for basic problems [58, 20, 54, 17, 15, 16, 27] in various hierarchies of linear and semidefinite programs [47, 52, 59, 49] (see [22, 26] for expositions).



© Pravesh K. Kothari and Peter Manohar;
licensed under Creative Commons License CC-BY 4.0
36th Computational Complexity Conference (CCC 2021).

Editor: Valentine Kabanets; Article No. 23; pp. 23:1–23:21



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



However, proving lower bounds for the sum-of-squares (SoS) semidefinite programming hierarchy – the strongest known hierarchy of efficiently solvable convex programs – has achieved only a limited amount of success. This is partially explained by the remarkable success of the SoS hierarchy in designing state-of-the-art algorithms for worst-case optimization problems such as max-cut [29], sparsest cut [4], unique games on general [1] and algebraic graphs [5, 32], quantum separability [13] and more recently, a string of successes in high-dimensional algorithmic statistics including robust estimation of moments [44], clustering spherical [34, 43] and non-spherical mixture models [8, 24], robust learning of all Gaussian mixtures [6, 48], list-decodable learning [39, 55, 7, 56], tensor decomposition [50], and sparse [25] and tensor principal component analysis [36], among others. Indeed, given the remarkable power of the SoS method in designing algorithms for such *average-case* settings, SoS lower bounds (and related restricted algorithmic techniques such as the low-degree polynomial method [33, 37, 46]) are increasingly used to ascertain average-case hardness and *algorithmic thresholds*.

In the last few years, there has been some progress in proving sum-of-squares lower bounds for average-case problems [30, 31, 57, 61, 11, 12, 42, 28, 51]. However, such progress has come about via fairly technical¹, problem-specific arguments and a host of natural questions, e.g. combinatorial optimization on sparse random graphs, remain out of reach of current techniques. In particular, a central challenge in this line of work has been to analyze the sum-of-squares semidefinite programs for refuting the existence of a k -coloring in Erdős-Rényi random graphs. Classical works [19] in probability showed that the chromatic number of $G \sim G(n, 1/2)^2$ is tightly concentrated around $n/2 \log_2 n$. However, the best known polynomial time algorithm (corresponding to the degree 2 SoS relaxation, a close variant of the famous Lovász theta function) can only *refute* the existence of a \sqrt{n} -coloring in such random graphs³. While it is natural to guess that higher-degree relaxations yield no significant improvement, establishing this has proved to be an elusive goal. Indeed, even the easier goal of establishing sharp SoS lower bounds for the clique number of $G \sim G(n, 1/2)$ required [12] the introduction of *pseudo-calibration* – a technique that has found several further uses in establishing SoS lower bounds for average-case problems. However, analyzing lower bound constructions based on pseudo-calibration requires understanding the spectra of complicated random matrices with dependent random entries. While this has been accomplished for a few select examples [33, 28], the case of graph coloring seems to be particularly unwieldy and has thus resisted progress so far.

In this paper, we establish a tight SoS lower bound for a natural higher-degree SoS relaxation of the graph coloring problem in $G(n, 1/2)$. Our proof circumvents pseudo-calibration entirely. Instead, we exhibit a *non-pointwise complete reduction* – a notion of reductions that departs from the standard framework introduced by Tulsiani [61] (and used in [18]) – that obtains a lower bound for the coloring problem from a lower bound for the independent set problem (see Section Section 1.3 for a detailed discussion). Somewhat surprisingly, our analysis does *not* require spectral analysis of complicated random matrices and instead succeeds whenever the lower bound construction for the independent set problem satisfies some natural covering properties. Our main result then follows by verifying these properties for the construction of [12].

¹ Almost all recent analyses run into ~ 50 pages!

² Recall $G \sim G(n, 1/2)$ is a graph on n vertices where each edge $\{i, j\}$ is independently included with probability $1/2$.

³ We note that a close variant of Lovász-theta function is also a crucial component in the current state-of-the-art algorithms for worst-case coloring of k -colorable graphs with a small polynomial number of colors [38, 3, 21].

1.1 Results

Our results apply to the following polynomial constraint system in the real-valued variables $\{x_{i,c}\}_{i \in [n], c \in [k]}$ that is satisfiable if and only if the graph G is k -colorable.

Color Constraints	
$x_{i,c}^2 = x_{i,c}$ for all $i \in [n], c \in [k]$	(Booleanity Constraints)
$x_{i,c}x_{j,c} = 0$ for all $c \in [k]$ and $\{i, j\} \in E(G)$	(Edge Constraints)
$\sum_c x_{i,c} \geq 1$ for all $i \in [n]$	(Sum Constraints)

In Color Constraints, the variable $x_{i,c}$ represents the 0-1 indicator of whether the i vertex is in the c -th color class. The booleanity constraints enforce that $x_{i,c} \in \{0, 1\}$, the edge constraints enforce that if $\{i, j\} \in E(G)$, then the subset of colors assigned to i is disjoint from the subset of colors assigned to j , and the sum constraints enforce that each vertex is in at least one color class.

Color Constraints allow for a vertex to be in *more than one* color class. Our lower bound technique does not currently succeed for the related set of constraints where each vertex must belong to exactly one color class. See Section 1.5 for a discussion on the difference between the formulations.

Our main result shows that with high probability over the draw of $G \sim G(n, 1/2)$, the degree $O(\varepsilon^2 \log n)$ SoS proof system cannot refute Color Constraints for G when $k = n^{\frac{1}{2} + \varepsilon}$.

► **Theorem 1.** *Let n be sufficiently large positive integer and $\varepsilon \in (\Omega(\sqrt{\frac{1}{\log n}}), \frac{1}{2})$. Then, for $k = n^{\frac{1}{2} + \varepsilon}$ and $d = O(\varepsilon^2 \log n)$, with probability $1 - 1/\text{poly}(n)$ over the draw of $G \sim G(n, 1/2)$, the $n^{O(d)}$ -time, degree d sum-of-squares relaxation of Color Constraints cannot refute the existence of a k -coloring of G .*

Equivalently, Theorem 1 says that with high probability over $G \sim G(n, 1/2)$, Color Constraints do not admit an $O(\varepsilon^2 \log n)$ -degree positivstellensatz refutation when $k = n^{\frac{1}{2} + \varepsilon}$. As was formally verified in [10]⁴, a degree 2 coloring pseudo-expectation is equivalent to a vector solution with value at least k to the semidefinite program that computes the Lovász theta function. To the best of our knowledge, this result gives the first lower bound for $\omega(1)$ rounds (or even a single round of strengthening of the basic SDP) in a natural SDP hierarchy.

► **Remark 1 (Tightness of Theorem 1).** It is well-known [23, 9] that the degree 2 sum-of-squares relaxation of Color Constraints can refute the existence of k -coloring in $G \sim G(n, 1/2)$ for $k = O(\sqrt{n})$. Thus, our lower bound in Theorem 1 is tight up to a n^ε factor in k . On the other hand, we give a simple proof in Appendix B that shows that the degree $8(1 + o(1)) \log_2 n$ SoS relaxation of Color Constraints succeeds in refuting the existence of a k -coloring in $G(n, 1/2)$ (w.h.p.) for the nearly optimal [19] bound of $k \leq \frac{n}{e \cdot 2^{(1+o(1)) \log_2 n}}$. Hence, the upper bound on d in Theorem 1 is tight up to constants.

⁴ [10] proved this equivalence for a slightly different formulation of Color Constraints, which we will discuss in Section 1.5. However, the same proof works even for Color Constraints.

1.2 A non-pointwise complete SoS reduction from coloring to independent set

Using standard SDP duality, proving Theorem 1 is equivalent to proving the existence of a dual witness called a pseudo-expectation defined below (see lecture notes [14] and the monograph [26] for background).

► **Definition 2** (Pseudo-expectation for Coloring). *A degree d coloring pseudo-expectation $\tilde{\mathbb{E}}$ for G using k colors is a linear operator that maps polynomials of degree $\leq d$ in variables $\{x_{i,c}\}_{i \in [n], c \in [k]}$ to \mathbb{R} , satisfying the following three properties:*

1. **Normalization:** $\tilde{\mathbb{E}}[1] = 1$,
2. **Positivity:** $\tilde{\mathbb{E}}[f^2] \geq 0$ for every polynomial f of degree at most $d/2$,
3. **Coloring Constraints:** $\tilde{\mathbb{E}}$ satisfies Color Constraints.
 - (a) for every polynomial f of degree at most $d - 2$, $\tilde{\mathbb{E}}[f \cdot (x_{i,c}^2 - x_{i,c})] = 0$,
 - (b) for every polynomial f of degree at most $d - 2$ and any edge $\{i, j\} \in E(G)$, $\tilde{\mathbb{E}}[f \cdot x_{i,c} x_{j,c}] = 0$,
 - (c) for every polynomial f of degree at most $\frac{d-1}{2}$, $\tilde{\mathbb{E}}[f^2 \cdot (\sum_{c \leq k} x_{i,c} - 1)] \geq 0$.

In order to prove Theorem 1, it suffices to show that with high probability over the draw of $G \sim G(n, 1/2)$, there is a degree $O(\varepsilon^2 \log n)$ coloring pseudo-expectation for the graph G that uses $k = n^{\frac{1}{2} + \varepsilon}$ colors. Somewhat surprisingly, we prove the existence of such a pseudo-expectation essentially without any random matrix analysis. Instead, we construct a coloring pseudo-expectation $\tilde{\mathbb{E}}'$ for G from a pseudo-expectation $\tilde{\mathbb{E}}$ satisfying the related independent set constraints for *the same graph* G whenever $\tilde{\mathbb{E}}$ satisfies two additional natural “covering” properties. We recall the definition of an independent set pseudo-expectation below.

► **Definition 3** (Pseudo-expectation for Independent Set). *A degree d independent set pseudo-expectation $\tilde{\mathbb{E}}$ is a linear operator that maps polynomials of degree $\leq d$ in variables $\{x_i\}_{i \in [n]}$ to \mathbb{R} , satisfying the following three properties:*

1. **Normalization:** $\tilde{\mathbb{E}}[1] = 1$,
2. **Positivity:** $\tilde{\mathbb{E}}[f^2] \geq 0$ for every polynomial f of degree at most $d/2$,
3. **Independent Set Constraints:** For every polynomial f of degree at most $d - 2$, $\tilde{\mathbb{E}}[f \cdot (x_i^2 - x_i)] = 0$ and $\tilde{\mathbb{E}}[f \cdot x_i x_j] = 0$ for any edge $\{i, j\} \in E(G)$.

Our main result that constructs a reduction from coloring to independent set is described below.

► **Theorem 2.** *Let G be a graph on n vertices, and let $\tilde{\mathbb{E}}$ be a degree d independent set pseudo-expectation. Suppose further that $\tilde{\mathbb{E}}$ satisfies the two “covering” properties: (1) $\tilde{\mathbb{E}}[x_i] \geq \frac{1}{k_0}$ for some integer k_0 , and (2) there exists $\lambda \in \mathbb{R}_{>0}$ such that for all multilinear f with $\deg(f) \leq d/2$, $\tilde{\mathbb{E}}[f^2] \geq \lambda \|\Pi_G f\|_2^2$, where Π_G is the projection of f onto the linear subspace orthogonal to $\{g_{x_i x_j} : \{i, j\} \in E(G), \deg(g) \leq d - 2\}$ (viewed as a subset of coefficient vectors of polynomials with the Euclidean inner product), and $\|f\|_2$ denotes the ℓ_2 norm of the polynomial of f , viewed as a coefficient vector. Then, there is a degree $d' := 1 + d/2$ coloring pseudo-expectation $\tilde{\mathbb{E}}'$ using $k = O(k_0 d \log(n^d/\lambda))$ colors.*

Theorem 1 follows by verifying (see Section 3) that the independent set pseudo-expectation constructed in [12] satisfies the hypotheses of Theorem 2 with $k_0 = n^{\frac{1}{2} + \varepsilon}$ and $\lambda = n^{-O(d)}$.

Theorem 2 holds for *every* graph G that admits an independent set pseudo-expectation satisfying the two additional covering properties. Hence, Theorem 2 gives a reduction “within SoS” from the problem of coloring G to the problem of finding a large independent set in G . As a consequence of the modularity of Theorem 2, we have also reduced the task of proving

SoS lower bounds for coloring for $G(n, p)$ with $p \ll 1/2$ to the task of “merely” proving a similar lower bound for independent set for $G(n, p)$. The latter task, though challenging, appears significantly less daunting than attacking coloring directly.

To understand the two covering properties intuitively, note that even in “real-life” the existence of a single large independent set (say of size $\sim n/k$) does not imply the existence of a k -coloring of G . However, the existence of a k -coloring follows if we can prove that there is a collection of k independent sets that *cover* all vertices of G . The conditions appearing in Theorem 2 can be thought of as forcing two “low-degree” consequences of such a uniform covering property on the pseudo-expectation for independent sets. Informally, the first constraint says that each vertex i appears in the independent set with reasonable probability, and the second constraint says that the minimum eigenvalue of \mathbb{E} is not too small, once we ignore polynomials that are required to have pseudo-expectation 0 due to the independent set constraints.

1.3 Comparison with Tulsiani’s framework

Our proof of Theorem 2 requires a notion of reduction that departs from the standard framework introduced in [60]. Tulsiani’s method⁵ uses a *pointwise complete* reduction from problem B to problem A to construct a pseudo-expectation consistent with a polynomial formulation for B from a pseudo-expectation consistent with a polynomial formulation for A . Specifically, a *pointwise complete* SoS reduction from problem B to problem A is a map from instances I_A of problem A to instances I_B of problem B , along with a “solution map” $x \mapsto y$ that takes any solution x of instance I_A into a solution y of instance I_B that, in addition, satisfies: (1) each entry of the solution map $x \rightarrow y$ is computable by low-degree polynomials, and (2) there is a “low-degree sum-of-squares proof” that if x satisfies the constraint system A for instance I_A then y satisfies the constraint system B for instance I_B . In particular, if $y_i = p_i(x)$ for each i for polynomials p_1, p_2, \dots of degree most d_1 , then the framework allows us to transform a degree d pseudo-expectation consistent with A into a degree $\approx d/d_1$ pseudo-expectation consistent with B . Tulsiani used this machinery to prove several SoS lower bounds for *worst-case* combinatorial optimization problems such as constraint satisfaction, vertex cover, independent set and coloring.

In average-case settings, however, we need tight control over the map between instances I_A and I_B in order to obtain a lower bound that applies to the target distribution over the instances of problem B . This makes Tulsiani’s method not directly applicable to our setting since (if we insist on instance-preserving reductions) there is provably no pointwise complete, instance-preserving reduction from k -coloring to independent set. This is because the existence of a large independent set in G does not, in general, imply the existence of a valid coloring of G with a small number of colors. Instead, as we discuss next, our reduction directly maps a pseudo-expectation for independent set into a pseudo-expectation for coloring as long as the pseudo-expectation for independent set satisfies the additional uniform covering conditions.

1.4 Proof overview: coloring by repeated sampling

We describe our construction and a couple of main insights that go into the proof of Theorem 2 here. These ideas make the proof of Theorem 1 “stress-free”: they allow us to completely sidestep the technical complexity of analyzing constructions based on pseudo-calibration that involve computing the spectra of certain random matrices (called graphical matrices) for proving SoS lower bounds.

⁵ What follows is an equivalent description of Tulsiani’s work in the language of pseudo-expectations.

We begin by describing the conceptual heart of the idea. In order to do this, it is helpful to consider the thought experiment (and very special case!) where the independent set pseudo-expectation $\tilde{\mathbb{E}}$ is in fact the expectation operator \mathbb{E}_μ associated with some *distribution* μ on independent sets of G . Further, suppose that $\mathbb{E}_\mu[x_i] = \Pr[i \in S] \geq \frac{1}{k_0}$. Then, observe that we can immediately derive that G must be k -colorable with $O(k_0 \log n)$ colors. In fact, we can produce a simple, explicit probability distribution on k -colorings of G : independently sample k independent sets S_1, \dots, S_k from μ and set each of them to be a new color class. Observe that the chance that a certain vertex is not included in any of the S_i 's is $\Pr[i \notin \cup_{j=1}^k S_j] = (1 - \frac{1}{k_0})^k \leq e^{-k/k_0} \ll \frac{1}{n}$ for $k = O(k_0 \log n)$, and hence by a union bound, using the S_i 's as color classes gives a valid k -coloring of G with high probability.⁶ To get a distribution μ' entirely supported over k -colorings of G , one can simply sample S_1, \dots, S_k from μ *conditioned* on the high probability event that $\cup_{j=1}^k S_j = V(G)$.

Our key idea is to replicate this “independent sampling” step within the sum-of-squares framework. For pseudo-expectations, independent sampling produces a pseudo-expectation on a tuple of k independent sets given by the k -th tensor $\tilde{\mathbb{E}}^{\otimes k}$. However, there is no natural way to perform the final “conditioning” step for low-degree pseudo-expectations⁷, which for distributions is the simplest way to ensure the “covering property”, that is, $i \in \cup_{j=1}^k S_j$ for every i , or equivalently to make $\mathbb{E}_{\mu'}$ satisfy the sum constraints $\sum_c x_{i,c} \geq 1$ for every i .

The sampling analogy suggests a way out, however: observe that when one draws S_1, \dots, S_k from an actual probability distribution μ on independent sets that satisfies $\mathbb{E}_\mu x_i \geq 1/k_0$, we expect each i to be in not just one but in fact in $\frac{k}{k_0} = O(\log n)$ of the subsets. Equivalently, we expect that $\mathbb{E}_\mu \sum_c x_{i,c} = \Omega(\log n)$. Because this expectation is large, if low-degree polynomials of μ are sufficiently well-concentrated around their expectations, we may expect that the influence of the points x in the support of μ where $\sum_c x_{i,c} \leq 1 \ll \mathbb{E}_\mu \sum_c x_{i,c}$ to be small. Thus, one may hope that expectations of low-degree ($\deg \leq d$) polynomials cannot “distinguish” between distributions μ where every point in the support of μ satisfies $\sum_c x_{i,c} > 1$ versus those where the probability of $\sum_c x_{i,c} = 0$ is non-zero for some i . In that case, one might expect $\tilde{\mathbb{E}}^{\otimes k}$ to satisfy the sum constraints.

Our actual proof establishes precisely such a statement even for pseudo-distributions whenever the smallest nontrivial eigenvalue of the pseudo-moment matrix of the independent set pseudo-expectation $\tilde{\mathbb{E}}$ is not too small. We show that this condition implies a non-trivial eigenvalue lower bound for the k -fold tensor power of $\tilde{\mathbb{E}}$ on polynomials of total degree⁸ d . A direct argument relying on spectra of the tensor product of matrices yields an estimate that decays exponentially in k , which is too weak for us. Instead, we show that the smallest eigenvalue of $\tilde{\mathbb{E}}^{\otimes k}$ *when restricted to the subspace of polynomials of total degree $\leq d$* decays only as an exponential in $d \log n$. While eventually elementary, this argument is both crucial and somewhat technical and is presented in full in Section 2.4.1.

Intuitively, a good enough lower bound on the smallest non-zero eigenvalue of $\tilde{\mathbb{E}}^{\otimes k}$ on the relevant subspace of polynomials is our “surrogate” for the concentration of low-degree polynomials that we needed in the case of actual probability distributions above. Concretely,

⁶ Note that a vertex i will, with high probability, belong to multiple color classes. In order to obtain a valid k -coloring, we simply remove each vertex from all but one of its assigned color classes.

⁷ There is a natural and standard way to import “conditioning” of probability distributions into the SoS framework via “polynomial reweightings” (see [13] for a formal treatment of such reweightings). However, the relevant polynomial $\prod_i (1 - \prod_c (1 - x_{i,c}))$ in our case has degree nk , and so we would need the independent set pseudo-expectation to have degree $\gg n$ in order for the reweighting to be well-defined!

⁸ Notice that $\tilde{\mathbb{E}}^{\otimes k}$ is defined and even positive semidefinite on a *larger* subspace of polynomials that includes some of total degree $\sim kd$!

we use this non-trivial eigenvalue lower bound on $\tilde{\mathbb{E}}^{\otimes k}$ as follows: let h_i be the indicator polynomial of the “bad event” $\sum_c x_{i,c} \leq 1$. Then, we prove that for a polynomial f to be able to “detect” this event, we must have $\tilde{\mathbb{E}}^{\otimes k}[f^2 h_i] = \Omega(\tilde{\mathbb{E}}^{\otimes k}[f^2])$. However, applying Cauchy-Schwarz, we have that $\tilde{\mathbb{E}}^{\otimes k}[f^2 h_i] \leq \sqrt{\tilde{\mathbb{E}}^{\otimes k}[f^4] \tilde{\mathbb{E}}^{\otimes k}[h_i^2]}$. We show that the smallest eigenvalue condition implies a $2 \rightarrow 4$ hypercontractive inequality on the pseudo-expectation operator on polynomials of total degree $\leq d$, i.e., $\tilde{\mathbb{E}}^{\otimes k}[f^4] \leq (\frac{n^{O(d)}}{\lambda})^d \tilde{\mathbb{E}}^{\otimes k}[f^2]^2$. Combined with the estimate (that one roughly expects to hold from the independent sampling based argument) $\tilde{\mathbb{E}}^{\otimes k}[h_i] \approx e^{-k/k_0}$, this yields that $\tilde{\mathbb{E}}^{\otimes k}$ indeed satisfies the constraints $\sum_c x_{i,c} \geq 1$ for every i , when $k = O(k_0 d \log(n^d/\lambda))$.

1.5 Weak vs. strong formulation for coloring

The coloring axioms are often stated with an equality (we call this the *strong form*) in the sum constraints along with the additional constraints $\{x_{i,c} x_{i,c'} = 0 : c \neq c'\}$, instead of an inequality (the *weak form*) as done in Color Constraints. Namely, the strong coloring constraints are the following.

(Strong) Color Constraints	
$x_{i,c}^2 = x_{i,c}$ for all $i \in [n], c \in [k]$	(Booleanity Constraints)
$x_{i,c} x_{j,c} = 0$ for all $c \in [k]$ and $\{i, j\} \in E(G)$	(Edge Constraints)
$\sum_c x_{i,c} = 1$ for all $i \in [n]$	(Sum Equality Constraints)
$x_{i,c} x_{i,c'} = 0$ for all $c \neq c' \in [k]$	(Same Color Constraints)

When viewed as a polynomial optimization problem, there is no difference between the weak and strong formulations: one is satisfiable if and only if the other is. Further, SoS relaxations of both formulations “converge” (i.e., refute k -coloring in $G(n, 1/2)$ for the right value of k) at $O(\log n)$ degree, and both imply corresponding lower bounds for independent set: a degree d coloring (weak or strong) pseudo-expectation with k colors can easily be transformed into a degree d independent set pseudo-expectation with independent set size $\geq \frac{n}{k}$. Thus, while the SoS relaxation of the strong form is *formally* stronger (for degrees > 2), Color Constraints do not appear to meaningfully weaken the strong formulation.

However, at the moment our technique does not succeed in constructing a pseudo-expectation that satisfies the constraints in the strong formulation. This is an important technical issue encountered in proving several prior SoS lower bounds where it turns out to be unwieldy to handle “hard” constraints such as those formulated by an exact polynomial equality. For example, in the planted clique problem, one may naturally wish for the pseudo-expectation to satisfy the clique-size constraint “ $\sum_i x_i = \omega$ ” *exactly*. While this is achieved for the degree 4 pseudo-expectation of [35, 53], the degree $\sim \log n$ pseudo-expectation constructed in [12] does *not* satisfy this as a constraint. This technical deficiency can sometimes even be crucial in downstream applications. For example, the construction of the hardness result for finding Nash equilibria in two player games in [41] (see also the discussion in [40]) needs elaborate work-arounds in order to work without satisfying such exact constraints.

Informally speaking, this is because the proofs of positivity of candidate pseudo-expectations rely on “collecting terms” together in the *graphical matrix* (a class of structured random matrices) decomposition in order to form PSD matrices. This aggregation step needs coefficients on various graphical matrices appearing in the decomposition to satisfy certain exact relationships. Modifying such coefficients to satisfy hard constraints while maintaining positivity appears challenging.

Such technical difficulty has been dealt with in some special cases (where the analyses did not need pseudo-calibration in the first place). For example, for the (much simpler) case of constraint satisfaction problems with a single global equality constraint, this problem was addressed via certain *ad hoc* methods in a recent work [40]. That work also includes a longer discussion on the issues arising in constructing pseudo-expectations satisfying hard constraints. Finding general techniques to design and analyze pseudo-expectations that exactly satisfy multiple hard constraints *simultaneously* – such as those arising in the strong formulation of graph coloring – is an important and challenging open problem.

2 Reduction to SoS Lower Bounds for Independent Set

In this section, we prove Theorem 2 (restated below).

► **Theorem** (Theorem 2, restated). *Let G be a graph on n vertices, and let $\tilde{\mathbb{E}}$ be a degree d independent set pseudo-expectation. Suppose further that $\tilde{\mathbb{E}}$ satisfies the two “covering” properties: (1) $\tilde{\mathbb{E}}[x_i] \geq \frac{1}{k_0}$ for some integer k_0 , and (2) there exists $\lambda \in \mathbb{R}_{>0}$ such that for all multilinear f with $\deg(f) \leq d/2$, $\tilde{\mathbb{E}}[f^2] \geq \lambda \|\Pi_G f\|_2^2$, where Π_G is the projection of f onto the linear subspace orthogonal to $\{gx_i x_j : \{i, j\} \in E(G), \deg(g) \leq d - 2\}$ (viewed as a subset of coefficient vectors of polynomials with the Euclidean inner product), and $\|f\|_2$ denotes the ℓ_2 norm of the polynomial of f , viewed as a coefficient vector. Then, there is a degree $d' := 1 + d/2$ coloring pseudo-expectation $\tilde{\mathbb{E}}'$ using $k = O(k_0 d \log(n^d/\lambda))$ colors.*

2.1 Coloring degree of polynomials

Before proceeding with the proof, we first introduce some notation. Let f be a polynomial in the variables $\{x_{i,c}\}_{i \in [n], c \in [k]}$. We define the *coloring degree* of f , denoted by $\text{cdeg}(f)$, to be the maximum, taken over all monomials $\prod_{c=1}^k \prod_{i=1}^n x_{i,c}^{\alpha_{i,c}}$ for which f has a nonzero coefficient, of $\max_{c \in [k]} \deg(\prod_{i=1}^n x_{i,c}^{\alpha_{i,c}})$. As an example, the polynomial $x_{1,1}x_{1,2}$ has degree 2 and coloring degree 1, while the polynomial $x_{1,1}x_{2,1}$ has degree 2 and coloring degree 2. Informally, the coloring degree only “charges” a polynomial for degrees in variables of a single color.

Let \mathcal{P}_d denote the set of polynomials in the variables $\{x_i\}_{i \in [n]}$ of degree at most d . Then, the set of coloring degree $\leq d$ polynomials is precisely $\mathcal{P}_d^{\otimes k}$. Recall that the operator Π_G is the projection of $f \in \mathcal{P}_d$ to the subspace orthogonal to $\{gx_i x_j : \{i, j\} \in E(G), \deg(g) \leq d - 2\}$. We let $\Pi_G^{\otimes k}$ denote the k -th tensor of Π_G . Namely, $\Pi_G^{\otimes k}$ is the projection of $f \in \mathcal{P}_d^{\otimes k}$ to the subspace orthogonal to $\{gx_{i,c} x_{j,c} : \{i, j\} \in E(G), c \in [k], \text{cdeg}(gx_{i,c} x_{j,c}) \leq d\}$.

Recall that a degree d pseudo-expectation is a linear operator $\tilde{\mathbb{E}}: \mathcal{P}_d \rightarrow \mathbb{R}$ such that $\tilde{\mathbb{E}}[1] = 1$, and $\tilde{\mathbb{E}}[f^2] \geq 0$ for all f with $\deg(f) \leq d/2$. For a pseudo-expectation $\tilde{\mathbb{E}}: \mathcal{P}_d \rightarrow \mathbb{R}$, we define $\tilde{\mathbb{E}}^{\otimes k}: \mathcal{P}_d^{\otimes k} \rightarrow \mathbb{R}$ to be the k -th tensor of $\tilde{\mathbb{E}}$. Concretely, $\tilde{\mathbb{E}}^{\otimes k}$ is a pseudo-expectation in the variables $\{x_{i,c}\}_{i \in [n], c \in [k]}$, defined as follows. For polynomials f_1, \dots, f_k where (1) f_c is a polynomial in the variables $\{x_{i,c}\}_{i \in [n]}$ for each c , and (2) $\deg(f_c) \leq d$ for all c , we first define $\tilde{\mathbb{E}}^{\otimes k}[\prod_{c=1}^k f_c] \stackrel{\text{def}}{=} \prod_{c=1}^k \tilde{\mathbb{E}}[f_c]$, and then extend $\tilde{\mathbb{E}}^{\otimes k}$ to be defined on all $f \in \mathcal{P}_d^{\otimes k}$ via linearity. We define a *coloring degree d pseudo-expectation* to be a linear operator $\tilde{\mathbb{E}}: \mathcal{P}_d^{\otimes k} \rightarrow \mathbb{R}$ such that $\tilde{\mathbb{E}}[1] = 1$ and $\tilde{\mathbb{E}}[f^2] \geq 0$ for all f with $\text{cdeg}(f) \leq d/2$. If $\tilde{\mathbb{E}}$ is a degree d pseudo-expectation, then $\tilde{\mathbb{E}}^{\otimes k}$ is a coloring degree d pseudo-expectation.

It is well-known that degree d pseudo-expectations satisfy the Cauchy-Schwarz inequality:

► **Fact 4** (See [14]). *Let f, g be polynomials with $\deg(f), \deg(g) \leq d/2$, and let $\tilde{\mathbb{E}}$ be a degree d pseudo-expectation. Then $\tilde{\mathbb{E}}[fg] \leq \sqrt{\tilde{\mathbb{E}}[f^2]\tilde{\mathbb{E}}[g^2]}$.*

We observe that a similar fact also holds for coloring degree d pseudo-expectations.

► **Fact 5.** *Let f, g be polynomials with $\text{cdeg}(f), \text{cdeg}(g) \leq d/2$, and let $\tilde{\mathbb{E}}$ be a coloring degree d pseudo-expectation. Then $\tilde{\mathbb{E}}[fg] \leq \sqrt{\tilde{\mathbb{E}}[f^2]\tilde{\mathbb{E}}[g^2]}$.*

The proof of Fact 5 is nearly identical to the proof of Fact 4, as the proof of Fact 4 merely requires that $\tilde{\mathbb{E}}$ is a pseudo-expectation where $\tilde{\mathbb{E}}[f^2], \tilde{\mathbb{E}}[g^2], \tilde{\mathbb{E}}[fg]$ and $\tilde{\mathbb{E}}[(f - g)^2]$ are all well-defined.

2.2 Proof of Theorem 2

Construction of the pseudo-expectation. Fix a graph G and degree bound d , and let $\tilde{\mathbb{E}}: \mathcal{P}_d \rightarrow \mathbb{R}$ be a degree d independent set pseudo-expectation for G such that (1) $\tilde{\mathbb{E}}[x_i] \geq \frac{1}{k_0}$, and (2) $\tilde{\mathbb{E}}[f^2] \geq \lambda \|\Pi_G f\|_2^2$. Let $k \in \mathbb{N}$ to be chosen later, and assume without loss of generality that $k < n$.

Let $\tilde{\mathbb{E}}^{\otimes k}: \mathcal{P}_d^{\otimes k} \rightarrow \mathbb{R}$ be the k -fold tensor power of $\tilde{\mathbb{E}}$, and let $\tilde{\mathbb{E}}'$ be the pseudo-expectation defined over all polynomials f that have *degree* at most $1 + d/2$ obtained by restricting $\tilde{\mathbb{E}}^{\otimes k}$ to this subspace.

Analysis of the constraints. We first observe that both $\tilde{\mathbb{E}}'$ and $\tilde{\mathbb{E}}^{\otimes k}$ trivially satisfy the booleanity constraints, edge constraints, and positivity constraint (over their respective domains), since $\tilde{\mathbb{E}}$ satisfies these constraints. We verify these simple facts in Appendix A. As a consequence, if f has $\Pi_G^{\otimes k} f = 0$, then $\tilde{\mathbb{E}}^{\otimes k}$ satisfies $f = 0$ as a constraint; namely, for any g with $\text{cdeg}(fg) \leq d$, it holds that $\tilde{\mathbb{E}}^{\otimes k}[fg] = 0$.

It remains to show that $\tilde{\mathbb{E}}'$ satisfies the sum constraints, i.e., for all f with $\deg(f) \leq \frac{d}{4}$ and for every i , $\tilde{\mathbb{E}}'[f^2(\sum_c x_{i,c} - 1)] \geq 0$. Fix i , and let $h_i := \sum_c x_{i,c} \prod_{c' \neq c} (1 - x_{i,c'}) + \prod_c (1 - x_{i,c})$. Note that h_i is the indicator of the event that $\sum_{c=1}^k x_{i,c} \leq 1$, and when written as a polynomial, has *coloring* degree 1.

We will rely on the following two technical lemmas in our proof. The first informally shows that $\tilde{\mathbb{E}}^{\otimes k}$ “thinks” that $\sum_c x_{i,c} \geq 2$ when the event indicated by h_i , namely “ $\sum_c x_{i,c} \leq 1$ ”, does not occur. Intuitively, this should clearly hold.

► **Lemma 6.** *$\tilde{\mathbb{E}}^{\otimes k}$ satisfies the constraint $(1 - h_i)(\sum_c x_{i,c} - 2) \geq 0$. Namely, for every polynomial f with $\text{cdeg}(f) \leq \frac{d-2}{2}$, it holds that $\tilde{\mathbb{E}}^{\otimes k}[f^2(1 - h_i)(\sum_c x_{i,c} - 2)] \geq 0$.*

The second lemma shows that the linear operator $\tilde{\mathbb{E}}$ satisfies a hypercontractive inequality – that is, the expectations of 4th powers of low-degree polynomials can be upper-bounded in terms of the expectations of their 2nd powers. Readers familiar with Fourier analysis over the hypercube may observe that the “scaling” in our estimate grows as $\exp(O(d \log n))$ in contrast to the $\exp(O(d))$ scaling in the usual hypercontractive inequality over the uniform measure on the Boolean hypercube. However, this worse bound will be sufficient for our purposes.

► **Lemma 7 (Hypercontractivity).** *For any multilinear f with $\text{cdeg}(f) \leq d/4$ satisfying $f = \Pi_G^{\otimes k} f$, we have $\tilde{\mathbb{E}}^{\otimes k}[f^4] \leq n^{O(\deg(f))} \cdot \tilde{\mathbb{E}}^{\otimes k}[f^2]^2 / (\lambda n^{-O(d)})^{2 \deg(f)}$.*

23:10 A Stress-Free Sum-Of-Squares Lower Bound for Coloring

We postpone the proofs of Lemmas 6 and 7 to Sections 2.3 and 2.4, respectively, and finish the proof assuming these two claims. Let f be any polynomial with $\text{cdeg}(f) \leq \frac{d}{4}$. We lower bound $\tilde{\mathbb{E}}^{\otimes k}[f^2 \sum_c x_{i,c}]$. Without loss of generality, we can assume that f is multilinear, as if f is not multilinear, then we can reduce it modulo the booleanity constraints. We can also assume that $f = \Pi_G^{\otimes k} f$, as if this does not hold then we write $f = f_1 + f_2$ where $f_1 = \Pi_G^{\otimes k} f_1$ and $\Pi_G^{\otimes k} f_2 = 0$, and then we observe that $\tilde{\mathbb{E}}^{\otimes k}[(f_1 + f_2)^2 \sum_c x_{i,c}] = \tilde{\mathbb{E}}^{\otimes k}[f_1^2 \sum_c x_{i,c}]$ (because $f_2 = 0$ is satisfied by $\tilde{\mathbb{E}}^{\otimes k}$ as a constraint) and $\text{cdeg}(f_1) \leq \text{cdeg}(f) \leq \frac{d}{4}$, as the projection operation can only decrease coloring degree. We note that $\tilde{\mathbb{E}}^{\otimes k}[h_i^2] = \tilde{\mathbb{E}}^{\otimes k}[h_i]$, since $h_i^2 \equiv h_i$ modulo the booleanity constraints $\{x_{i,c}^2 = x_{i,c}\}$. We have that

$$\begin{aligned} \tilde{\mathbb{E}}^{\otimes k}[f^2 \sum_c x_{i,c}] &= \tilde{\mathbb{E}}^{\otimes k}[f^2 h_i \sum_c x_{i,c}] + \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - h_i) \sum_c x_{i,c}] \\ &= \tilde{\mathbb{E}}^{\otimes k}[f^2 h_i^2 \sum_c x_{i,c}^2] + \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - h_i) \sum_c x_{i,c}] \quad (\text{as } h_i^2 \equiv h_i \text{ and } x_{i,c}^2 \equiv x_{i,c}) \\ &\geq 0 + \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - h_i) \sum_c x_{i,c}] \quad (\text{by positivity of } \tilde{\mathbb{E}}^{\otimes k}) \\ &\geq \tilde{\mathbb{E}}^{\otimes k}[f^2 \cdot 2(1 - h_i)] \quad (\text{by Lemma 6}) \\ &= 2(\tilde{\mathbb{E}}^{\otimes k}[f^2] - \tilde{\mathbb{E}}^{\otimes k}[f^2 h_i]) . \end{aligned}$$

Note that this is well-defined because $\tilde{\mathbb{E}}^{\otimes k}$ is defined on each of the terms in the above inequalities since $\text{cdeg}(f) \leq d/4 \leq (d-4)/2$ and $\text{cdeg}(h_i) = \text{cdeg}(\sum_c x_{i,c}) = 1$.

Now, we observe that

$$\begin{aligned} \tilde{\mathbb{E}}^{\otimes k}[f^2 h_i] &\leq \sqrt{\tilde{\mathbb{E}}^{\otimes k}[f^4]} \sqrt{\tilde{\mathbb{E}}^{\otimes k}[h_i^2]} \quad (\text{by Fact 5}) \\ &\leq \left(\frac{n^{O(d)}}{\lambda}\right)^{2 \deg(f)} \tilde{\mathbb{E}}^{\otimes k}[f^2] \cdot \sqrt{\tilde{\mathbb{E}}^{\otimes k}[h_i^2]} \quad (\text{by Lemma 7}) \\ &\leq \left(\frac{n^{O(d)}}{\lambda}\right)^{2 \deg(f)} \tilde{\mathbb{E}}^{\otimes k}[f^2] \cdot \sqrt{\tilde{\mathbb{E}}^{\otimes k}[h_i]} \quad (\text{since } h_i^2 \equiv h_i) . \end{aligned}$$

Next, we observe that $\tilde{\mathbb{E}}^{\otimes k}[h_i] = \tilde{\mathbb{E}}^{\otimes k}[\prod_c (1 - x_{i,c})] + \sum_c \tilde{\mathbb{E}}^{\otimes k}[x_{i,c} \prod_{c' \neq c} (1 - x_{i,c'})] = (1 - \tilde{\mathbb{E}}[x_i])^k + (k-1)(\tilde{\mathbb{E}}[x_i](1 - \tilde{\mathbb{E}}[x_i])^{k-1}) \leq k \cdot e^{-k/k_0}$ using the tensor structure of $\tilde{\mathbb{E}}^{\otimes k}$ and that $\tilde{\mathbb{E}}[x_i] \geq \frac{1}{k_0}$. Hence,

$$\begin{aligned} \tilde{\mathbb{E}}^{\otimes k}[f^2 h_i] &\leq \left(\frac{n^{O(d)}}{\lambda}\right)^{2 \deg(f)} \tilde{\mathbb{E}}^{\otimes k}[f^2] \cdot k \cdot e^{-k/k_0} \\ \implies \tilde{\mathbb{E}}^{\otimes k}[f^2 \sum_c x_{i,c}] &\geq 2\tilde{\mathbb{E}}^{\otimes k}[f^2] \left(1 - k \cdot e^{-k/k_0} \left(\frac{n^{O(d)}}{\lambda}\right)^{2 \deg(f)}\right) . \end{aligned}$$

Now, suppose that f is any polynomial with $\deg(f) \leq \frac{d}{4}$. This implies that (1) $\tilde{\mathbb{E}}'[f^2 \sum_c x_{i,c}]$ is defined, and (2) $\text{cdeg}(f) \leq \frac{d}{4}$, and so we have that (using $\lambda \leq 1$)

$$\begin{aligned} \tilde{\mathbb{E}}'[f^2 \sum_c x_{i,c}] &= \tilde{\mathbb{E}}^{\otimes k}[f^2 \sum_c x_{i,c}] \geq 2\tilde{\mathbb{E}}^{\otimes k}[f^2] \left(1 - k \cdot e^{-k/k_0} \left(\frac{n^{O(d)}}{\lambda}\right)^{d/2}\right) \\ &= \tilde{\mathbb{E}}'[f^2] \cdot 2 \left(1 - k \cdot e^{-k/k_0} \left(\frac{n^{O(d)}}{\lambda}\right)^{d/2}\right) . \end{aligned}$$

Choosing $k = O(k_0 d \log(n^d/\lambda))$, it follows that $1 - k \cdot e^{-k/k_0} \left(\frac{n^{O(d)}}{\lambda}\right)^{d/2} \geq \frac{1}{2}$ and so $\tilde{\mathbb{E}}'[f^2(\sum_c x_{i,c} - 1)] \geq 0$ for all f with $\deg(f) \leq \frac{d}{4}$. Since $\tilde{\mathbb{E}}'$ is a degree $1 + \frac{d}{2}$ pseudo-expectation, this means that $\tilde{\mathbb{E}}'$ satisfies the constraint $\sum_c x_{i,c} - 1 \geq 0$, which finishes the proof.

2.3 Proof of Lemma 6

Let f be any polynomial with $\text{cdeg}(f) \leq (d-2)/2$. It suffices to show that $\tilde{\mathbb{E}}^{\otimes k}[f^2(1 - h_i)(\sum_c x_{i,c} - 2)] \geq 0$. For $2 \leq t \leq k$, let $g_i^{(t)} = \prod_{c \leq t} (1 - x_{i,c})$, $g_{i,c}^{(t)} = \prod_{c' \neq c, c' \leq t} (1 - x_{i,c'})$, and $h_i^{(t)} := \sum_{c \leq t} x_{i,c} g_{i,c}^{(t)} + g_i^{(t)}$. We show by induction on t that for each $t \geq 0$, it holds that $\tilde{\mathbb{E}}^{\otimes k}[f^2(1 - h_i^{(t)})(\sum_{c \leq t} x_{i,c} - 2)] \geq 0$ for every f where the coloring degree of f on the first t colors is at most $(d-2)/2$, and $\text{cdeg}(f) \leq d/2$.

The base case is when $t = 2$. In this case, we have $1 - h_i^{(2)} = 1 - x_{i,1}(1 - x_{i,2}) - x_{i,2}(1 - x_{i,1}) - (1 - x_{i,1})(1 - x_{i,2}) = x_{i,1}x_{i,2}$, so $\tilde{\mathbb{E}}^{\otimes k}[f^2(1 - h_i^{(2)})(\sum_{c \leq 2} x_{i,c} - 2)] = \tilde{\mathbb{E}}^{\otimes k}[f^2(x_{i,1}x_{i,2})(x_{i,1} + x_{i,2} - 2)] = \tilde{\mathbb{E}}^{\otimes k}[f^2(2x_{i,1}x_{i,2} - 2x_{i,1}x_{i,2})] = 0$. Note that since f has coloring degree at most $(d-2)/2$ on the first colors, $\tilde{\mathbb{E}}^{\otimes k}$ is always defined on each of these polynomials.

We now show the induction step. We observe that $h_i^{(t+1)} = (1 - x_{i,t+1})h_i^{(t)} + x_{i,t+1}g_i^{(t)}$. Let f be a polynomial that has coloring degree $\leq (d-2)/2$ on the first $t+1$ colors, and $\text{cdeg}(f) \leq d/2$. We have

$$\begin{aligned} & \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - h_i^{(t+1)})(x_{i,t+1} + \sum_{c \leq t+1} x_{i,c} - 2)] \\ &= \tilde{\mathbb{E}}^{\otimes k}[f^2\left((1 - x_{i,t+1})(1 - h_i^{(t)}) + x_{i,t+1}(1 - g_i^{(t)})\right) \cdot (x_{i,t+1} + \sum_{c \leq t} x_{i,c} - 2)] \\ &= \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - x_{i,t+1})(1 - h_i^{(t)})(\sum_{c \leq t} x_{i,c} - 2)] + \tilde{\mathbb{E}}^{\otimes k}[f^2 x_{i,t+1}(1 - g_i^{(t)})(1 + \sum_{c \leq t} x_{i,c} - 2)] \\ &= \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - x_{i,t+1})^2(1 - h_i^{(t)})(\sum_{c \leq t} x_{i,c} - 2)] + \tilde{\mathbb{E}}^{\otimes k}[f^2 x_{i,t+1}^2(1 - g_i^{(t)})(\sum_{c \leq t} x_{i,c} - 1)] . \end{aligned}$$

Since f has coloring degree at most $(d-2)/2$ on the first $t+1$ colors, $f \cdot (1 - x_{i,t+1})$ has coloring degree at most $(d-2)/2$ on the first t colors, and also $\text{cdeg}(f \cdot (1 - x_{i,t+1}))$ is at most $d/2$, as on the $(t+1)$ -th color it has degree at most $(d-2)/2 + 1$, and on every other color it is either at most $(d-2)/2$ or $d/2$. So, $\tilde{\mathbb{E}}^{\otimes k}[f^2(1 - x_{i,t+1})^2(1 - h_i^{(t)})(\sum_{c \leq t} x_{i,c} - 2)] \geq 0$ by the induction hypothesis. We also observe that $f \cdot x_{i,t+1}$ has coloring degree at most $(d-2)/2$ on the first t colors, and has coloring degree at most $d/2$.

It remains to show that $\tilde{\mathbb{E}}^{\otimes k}[f^2(1 - g_i^{(t)})(\sum_{c \leq t} x_{i,c} - 1)] \geq 0$ for all t and for all f with $\text{cdeg}(f) \leq d/2$ and coloring degree at most $(d-2)/2$ in the first t colors. We observe that $g_i^{(t)} x_{i,c} \equiv 0$ for all $c \leq t$, and so it suffices to show that $\tilde{\mathbb{E}}^{\otimes k}[f^2 \sum_{c \leq t} x_{i,c}] \geq \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - g_i^{(t)})]$. We do this by induction on t . In the base case, we have $\tilde{\mathbb{E}}^{\otimes k}[f^2 x_{i,1}] = \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - (1 - x_{i,1}))]$. For the induction step, we have

$$\begin{aligned} & \tilde{\mathbb{E}}^{\otimes k}[f^2(x_{i,t+1} + \sum_{c \leq t} x_{i,c})] \geq \tilde{\mathbb{E}}^{\otimes k}[f^2 x_{i,t+1}] + \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - g_i^{(t)})] \\ & \geq \tilde{\mathbb{E}}^{\otimes k}[f^2 x_{i,t+1} g_i^{(t)}] + \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - g_i^{(t)})] \\ & = \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - (1 - x_{i,t+1})g_i^{(t)})] = \tilde{\mathbb{E}}^{\otimes k}[f^2(1 - g_i^{(t+1)})] , \end{aligned}$$

where we use the fact that $f x_{i,t+1} g_i^{(t)}$ has coloring degree $\leq d/2$ since f has coloring degree at most $(d-2)/2$ in the first $t+1$ colors, and that $(g_i^{(t)})^2 \equiv g_i^{(t)}$ modulo the hypercube constraints. This finishes the proof.

2.4 Proof of Lemma 7: hypercontractivity

Let f be a multilinear polynomial with $\text{cdeg}(f) \leq d/4$ with $\Pi_G^{\otimes k} f = f$. Suppose that:

$$\tilde{\mathbb{E}}^{\otimes k}[f^2] \geq \lambda_1 \|f\|_2^2, \quad (2.1)$$

$$\tilde{\mathbb{E}}^{\otimes k}[f^4] \leq \lambda_2 \|f^2\|_2^2, \quad (2.2)$$

$$\|f^2\|_2^2 \leq C \|f\|_2^4. \quad (2.3)$$

Then it follows that $\tilde{\mathbb{E}}^{\otimes k}[f^4] \leq \lambda_2 \|f^2\|_2^2 \leq \lambda_2 C \|f\|_2^4 \leq \frac{\lambda_2 C}{\lambda_1^2} \tilde{\mathbb{E}}^{\otimes k}[f^2]^2$.

Equation (2.1) follows from the following lemma with $\lambda_1 := (\lambda n^{-O(d)})^{\text{deg}(f)}$.

► **Lemma 8** (Eigenvalue lower bound for $\tilde{\mathbb{E}}^{\otimes k}$). *Suppose that for any multilinear g with $\text{deg}(g) \leq d/2$, $\tilde{\mathbb{E}}[g^2] \geq \lambda \|\Pi_G g\|_2^2$. Then for any multilinear f of coloring degree $\leq d/2$, it holds that $\tilde{\mathbb{E}}^{\otimes k}[f^2] \geq (\lambda n^{-O(d)})^{\text{deg}(f)} \cdot \|\Pi_G^{\otimes k} f\|_2^2$.*

We postpone the proof of Lemma 8 for now, and finish the proof of Lemma 7.

Let $g = \sum_m g_m \cdot m$ be a polynomial of degree $\text{deg}(g)$ with $\text{cdeg}(g) \leq d/2$, where m is a monomial and g_m is the coefficient of g for the monomial m . Since $\tilde{\mathbb{E}}^{\otimes k}$ satisfies the booleanity constraints, it follows that $\tilde{\mathbb{E}}[m] \leq 1$ for all monomials m . Hence, $\tilde{\mathbb{E}}[g^2] \leq \sum_{m_1, m_2} |g_{m_1} g_{m_2}| = (\sum_m |g_m|)^2 \leq (nk)^{O(\text{deg}(g))} \|g\|_2^2$, by Cauchy-Schwarz, as g is supported on at most $(nk)^{O(\text{deg}(g))}$ distinct monomials. Since $(nk)^{O(\text{deg}(g))} \leq n^{O(\text{deg}(g))}$ as $k < n$, it follows that $\tilde{\mathbb{E}}^{\otimes k}[g^2] \leq n^{O(\text{deg}(g))} \|g\|_2^2$, and so (setting $g = f^2$) Equation (2.2) holds with $\lambda_2 := n^{O(\text{deg}(f))}$.

Finally, for a polynomial g and monomial m let g_m be the coefficient of g on m . For any m of degree $\leq 2 \text{deg}(f)$, we have that $f_m^2 = \sum_{m_1, m_2: m_1 \cdot m_2 = m} f_{m_1} f_{m_2}$. We observe that this is equal to $\langle v^{(m)}, f \rangle$, where $v^{(m)}$ is the vector defined as $v_{m_2}^{(m)} \stackrel{\text{def}}{=} f_{m_1}$ where $m_1 \cdot m_2 = m$ (and is 0 if no such m_1 exists). It follows that $\|v^{(m)}\|_2 \leq \|f\|_2$, and hence that $|\langle v^{(m)}, f \rangle| \leq \max_{v: \|v\|_2 \leq \|f\|_2} |\langle v, f \rangle| = \|f\|_2^2$. Hence, $|f_m^2| \leq \|f\|_2^4$, and so $\|f^2\|_2^2 = \sum_{m: \text{deg}(m) \leq 2 \text{deg}(f)} |f_m^2| \leq (nk)^{O(\text{deg}(f))} \|f\|_2^4 = n^{O(\text{deg}(f))} \|f\|_2^4$, and so Equation (2.3) holds with $C := n^{O(\text{deg}(f))}$.

Combining, we conclude that $\tilde{\mathbb{E}}^{\otimes k}[f^4] \leq \tilde{\mathbb{E}}^{\otimes k}[f^2]^2 / (\lambda n^{-O(d)})^{2 \text{deg}(f)}$, which finishes the proof.

2.4.1 Proof of Lemma 8: eigenvalue lower bound for $\tilde{\mathbb{E}}^{\otimes k}$

Proof outline. The proof proceeds in three steps. First, we show that the moment matrix of the independent set pseudo-expectation $\tilde{\mathbb{E}}$, when written in a basis so that the constant polynomial 1 is an eigenvector, has an eigenvalue lower bound of $\lambda n^{-O(d)}$. To show that this property implies the desired eigenvalue lower bound, we observe that any f of total degree $\leq d$ is a linear combination of monomials that use at most $\text{deg}(f)$ colors. Further, (the coefficient vector of) each such monomial is a linear combination of tensor products of eigenvectors of the $\tilde{\mathbb{E}}$ that use a “non-1” eigenvector in at most $\text{deg}(f)$ modes of the tensor and thus is in the span of eigenvectors of $\tilde{\mathbb{E}}^{\otimes k}$ (in the new basis) with eigenvalue at least $(\lambda n^{-O(d)})^{\text{deg}(f)}$. This reasoning immediately implies that f , when written in the chosen basis, has the desired eigenvalue lower bound. To finish the proof, we argue that the change of basis does not modify $\|f\|_2$ by too much.

We now proceed with implementing the above proof plan. For every $S \subseteq [n]$ with $|S| \leq d$, recall that we can express any multilinear polynomial g with degree $\leq d$ as a linear combination of the monomials $x_S \stackrel{\text{def}}{=} \prod_{i \in S} x_i$. Let g_S be the coefficient of g on the

monomial S , so that $g = \sum_{|S| \leq d} g_S x_S$. Let e_S be the S -th standard basis vector in $\mathbb{R}^{\binom{n}{\leq d}}$. Then g (as a vector of coefficients) is $\sum_S g_S e_S$. For $S \neq \emptyset$, define e'_S as $e_S - \tilde{\mathbb{E}}[x_S] \cdot e_\emptyset$. We can write g uniquely in the e'_S basis as $g = \sum_S g'_S e'_S$, where $g'_S = g_S$ for $S \neq \emptyset$, and $g'_\emptyset = g_\emptyset + \sum_{S \neq \emptyset} g_S \tilde{\mathbb{E}}[x_S] = \tilde{\mathbb{E}}[g]$. Note that, if we let $x'_S := x_S - \tilde{\mathbb{E}}[x_S]$ for $S \neq \emptyset$ and $x'_\emptyset := x_\emptyset = 1$, then $g = \sum_S g'_S x'_S$ as a polynomial.

Let \mathcal{M} be the moment matrix of $\tilde{\mathbb{E}}$ in the x' basis. This matrix is indexed by sets $S, S' \subseteq [n]$ with $|S|, |S'| \leq d/2$, and $\mathcal{M}(S, S') = \tilde{\mathbb{E}}[x'_S x'_{S'}]$, which is equal to $\tilde{\mathbb{E}}[(x_S - \tilde{\mathbb{E}}[x_S])(x_{S'} - \tilde{\mathbb{E}}[x_{S'}])] = \tilde{\mathbb{E}}[x_S x_{S'}] - \tilde{\mathbb{E}}[x_S] \tilde{\mathbb{E}}[x_{S'}]$ if $S, S' \neq \emptyset$, equal to 0 if exactly one of S, S' is \emptyset , and equal to 1 if $S = S' = \emptyset$. This implies that e'_\emptyset is an eigenvector of \mathcal{M} with eigenvalue 1. We also observe that if g has degree $\leq d/2$ and g' is the coefficient vector of g in the e' basis, then $\tilde{\mathbb{E}}[g^2] = g'^\top \mathcal{M} g'$.

We now prove the following eigenvalue lower bound on \mathcal{M} .

▷ **Claim 9.** $\mathcal{M} \succeq \lambda n^{-O(d)} \Pi_G$.

Proof. Let S with $|S| \leq d/2$ be a set that is not an independent set in G , i.e. that $\Pi_G e_S = 0$. We observe that $\mathcal{M} e'_S = 0$. Indeed, the T -th entry of $\mathcal{M} e'_S$ is $\mathcal{M}(T, S) = \tilde{\mathbb{E}}[x'_T x'_S] = \tilde{\mathbb{E}}[x_T x_S] - \tilde{\mathbb{E}}[x_T] \tilde{\mathbb{E}}[x_S] = 0 - 0 = 0$ for $T \neq \emptyset$, and is 0 if $T = \emptyset$ because $\mathcal{M}(\emptyset, S) = 0$ for $S \neq \emptyset$.

Now, let $g' = \sum_{S: |S| \leq d} g'_S e'_S$ be arbitrary. By the above, without loss of generality we may assume that $g'_S = 0$ for all S that is not an independent set in G . Let g be the corresponding polynomial in the x basis, so that $g = \sum_S g_S x_S$, where $g_\emptyset = g'_\emptyset - \sum_{S \neq \emptyset} g'_S \tilde{\mathbb{E}}[x_S]$ and $g_S = g'_S$ for all $S \neq \emptyset$. Notice that $\tilde{\mathbb{E}}[g] = g'_\emptyset$. We observe that $\Pi_G g = g$, as $g_S = g'_S = 0$ for all S that is not an independent set in G . Now, we have that $g'^\top \mathcal{M} g' = \tilde{\mathbb{E}}[g^2] \geq \lambda \|\Pi_G g\|_2^2 = \lambda \|g\|_2^2$, by our eigenvalue lower bound assumption on $\tilde{\mathbb{E}}$.

It remains to relate $\|g\|_2^2$ and $\|g'\|_2^2$. We have that $\|g'\|_2^2 = \sum_{|S| \leq d/2} g'^2_S = \tilde{\mathbb{E}}[g]^2 + \sum_{S \neq \emptyset} g'^2_S \leq \tilde{\mathbb{E}}[g]^2 + \|g\|_2^2 \leq \tilde{\mathbb{E}}[g^2] + \|g\|_2^2 \leq (n^{O(d)} + 1) \|g\|_2^2$, as $\tilde{\mathbb{E}}[g^2] \leq n^{O(d)} \|g\|_2^2$ since $0 \leq \tilde{\mathbb{E}}[x_S x_T] \leq 1$ for all S, T , and there are at most $n^{O(d)}$ such pairs. Hence, $g'^\top \mathcal{M} g' \geq \lambda n^{-O(d)} \|g'\|_2^2$ when $g' = \Pi_G g'$, and so $\mathcal{M} \succeq \lambda n^{-O(d)} \Pi_G$. ◁

We have already shown that e'_\emptyset is an eigenvector of \mathcal{M} with eigenvalue 1, and that the zero eigenvectors of \mathcal{M} are the vectors e'_S where S is not an independent set in G . Let $f_0 = e'_\emptyset, f_1, \dots, f_t$ be the eigenvectors of \mathcal{M} with nonzero eigenvalues $\lambda_0 = 1, \lambda_1, \dots, \lambda_t$, where $\lambda_i \geq \lambda n^{-O(d)}$ for $1 \leq i \leq t$. Let $\mathcal{M}^{\otimes k}$ be the k -th tensor of \mathcal{M} . Let $f_i^{(c)}$ denote the i -th eigenvector in the c -th component of the tensor. The eigenvectors of $\mathcal{M}^{\otimes k}$ are the vectors $\bigotimes_{c=1}^k f_i^{(c)}$. We additionally observe that $\mathcal{V}^{(c)} \stackrel{\text{def}}{=} \mathbf{Span}(f_i^{(c)} : i > 0) = \mathbf{Span}(e'_S{}^{(c)} : |S| > 0, \Pi_G e_S = e_S)$, as $f_0^{(c)} = e'_\emptyset{}^{(c)}$.

Let f be a multilinear polynomial with $\text{cdeg}(f) \leq d/2$ in the variables $\{x_{i,c}\}_{i \in [n], c \in [k]}$. That is, f is a vector in $\mathbf{Span}\left(\bigotimes_{c=1}^k e_{S_c}^{(c)} : |S_c| \leq d/2 \forall c \in [k]\right)$, where $e_S^{(c)}$ denotes the S -th standard basis vector in the c -th component of the tensor. As before, we can write f as a vector f' in the e' basis, so $f' = \sum_{(S_1, \dots, S_k) : |S_c| \leq d/2 \forall c \in [k]} f'_{(S_1, \dots, S_k)} \bigotimes_{c=1}^k e_{S_c}^{(c)}$. We again observe that $\tilde{\mathbb{E}}^{\otimes k}[f^2] = f'^\top \mathcal{M}^{\otimes k} f'$, because the $((S_1, \dots, S_k), (T_1, \dots, T_k))$ -th entry of $\mathcal{M}^{\otimes k}$ is exactly $\prod_{c=1}^k \tilde{\mathbb{E}}[x'_{S_c} x'_{T_c}] = \tilde{\mathbb{E}}^{\otimes k}[\prod_{c=1}^k x'_{S_c} x'_{T_c}]$. Note that by the structure of the zero eigenvectors of \mathcal{M} , if f satisfies $\Pi_G^{\otimes k} f = 0$, then f is an eigenvector of $\mathcal{M}^{\otimes k}$ with eigenvalue 0. In particular, without loss of generality we can assume that $f = \Pi_G^{\otimes k} f$, as by the above we can discard the component of f in the kernel of $\Pi_G^{\otimes k}$.

Let (S_1, \dots, S_k) be such that $f'_{(S_1, \dots, S_k)} \neq 0$. We must have $S_c = \emptyset$ for all but at most $\deg(f)$ of the c 's. This is because f has degree $\deg(f)$, and so in particular every monomial in f can only use at most $\deg(f)$ distinct colors. This shows that $f' \in \mathcal{V} := \text{Span} \left(\bigotimes_{c \in C} \mathcal{V}^{(c)} \otimes_{c \notin C} e'_\emptyset^{(c)} : C \subseteq [k], |C| \leq \deg(f) \right)$. We observe that \mathcal{V} is the span of eigenvectors of \mathcal{M} of the form $\bigotimes_{c \in C} f'_{i_c}^{(c)} \otimes_{c \notin C} e'_\emptyset^{(c)}$ for $|C| \leq \deg(f)$. Since each of these vectors is an eigenvector with eigenvalue at least $(\lambda n^{-O(d)})^{|C|} \cdot 1^{k-|C|} \geq (\lambda n^{-O(d)})^{\deg(f)}$, it follows that $f'^T \mathcal{M}^{\otimes k} f' \geq (\lambda n^{-O(d)})^{\deg(f)} \|f'\|_2^2$. Thus, $\tilde{\mathbb{E}}^{\otimes k}[f^2] \geq (\lambda n^{-O(d)})^{\deg(f)} \|f'\|_2^2$.

It remains to relate $\|f'\|_2^2$ and $\|f\|_2^2$. Fix (S_1, \dots, S_k) with $|S_c| \leq d/2$. Let (T_1, \dots, T_k) with $|T_c| \leq d/2$. We say that (T_1, \dots, T_k) *extends* (S_1, \dots, S_k) if for every c , either $T_c = S_c$ or $T_c \neq \emptyset$ and $S_c = \emptyset$. The *parity* of the extension is the parity of the number of c where $T_c \neq \emptyset$ and $S_c = \emptyset$. We observe that $f_{(S_1, \dots, S_k)} = \sum_{(T_1, \dots, T_k) \text{ extending } (S_1, \dots, S_k)} (\text{parity of extension}) \cdot f'_{(T_1, \dots, T_k)}$. This is because $e'_{(T_1, \dots, T_k)} = \bigotimes_{c=1}^k e'_{T_c} = \bigotimes_{c=1}^k (e_{T_c} - e_\emptyset)$. We thus see that $\|f\|_1 \leq \sum_{(T_1, \dots, T_k)} |f'_{(T_1, \dots, T_k)}| \cdot n_{(T_1, \dots, T_k)}$, where $n_{(T_1, \dots, T_k)}$ is the number of (S_1, \dots, S_k) that (T_1, \dots, T_k) extends. We have shown that if $f'_{(T_1, \dots, T_k)} \neq 0$ then it must be the case that $T_c \neq \emptyset$ for at most $\deg(f)$ of the c 's. Hence, such (T_1, \dots, T_k) can only extend at most $2^{\deg(f)}$ of the (S_1, \dots, S_k) 's, as each of the (S_1, \dots, S_k) 's is obtained by changing a subset of the T_c 's to be empty. Hence, $\|f\|_1 \leq 2^{\deg(f)} \|f'\|_1$. Since f' has at most $(nk)^{\deg(f)} \leq n^{2\deg(f)}$ nonzero coefficients, we get that $\|f\|_2 \leq n^{\deg(f)} \cdot 2^{\deg(f)} \|f'\|_2$, and so we conclude that $\|f\|_2^2 \leq n^{O(\deg(f))} \|f'\|_2^2$, which finishes the proof.

3 Proof of Theorem 1: coloring lower bound

We now prove Theorem 1 (restated below in the language of pseudo-expectations) from Theorem 2. In this section, we assume familiarity with the planted clique pseudo-expectation of [12].

► **Theorem** (Theorem 1, restated). *For sufficiently large n , for any $\varepsilon \in (\Omega(\sqrt{\frac{1}{\log n}}), \frac{1}{2})$, with probability $1 - 1/\text{poly}(n)$ over the draw of $G \sim G(n, 1/2)$, there is a degree $d = O(\varepsilon^2 \log n)$ coloring pseudo-expectation $\tilde{\mathbb{E}}$ using $k = n^{\frac{1}{2} + \varepsilon}$ colors.*

We begin by recalling the main theorem of [12].

► **Theorem 10** ([12]). *There is an absolute constant C such that for n sufficiently large, $C/\sqrt{\log n} \leq \varepsilon < \frac{1}{2}$, $\omega = n^{\frac{1}{2} - \varepsilon}$, and $d = (\varepsilon/C)^2 \log n$, with probability $1 - 1/\text{poly}(n)$ over $G \sim G(n, 1/2)$, the operator $\tilde{\mathbb{E}}_G$ defined in [12] satisfies:*

1. $\tilde{\mathbb{E}}_G[1] = 1 \pm n^{-\Omega(\varepsilon)}$,
2. $\tilde{\mathbb{E}}_G[\sum_i x_i] = \omega(1 \pm n^{-\Omega(\varepsilon)})$,
3. $\tilde{\mathbb{E}}_G[x_S] = 0$ for all $|S| \leq d$ that is not a clique in G ,
4. $\tilde{\mathbb{E}}_G[f^2] \geq \lambda \|\Pi'_G f\|_2^2$ where $\lambda = \Omega\left(\left(\frac{\omega}{n}\right)^{d+1}\right)$ and Π'_G is the projection onto x_S for S a clique in G .

We first observe that if $G \sim G(n, 1/2)$, then the complement graph $\bar{G} \sim G(n, 1/2)$ also, and moreover $\tilde{\mathbb{E}}_G$ will satisfy the independent set constraints as $\tilde{\mathbb{E}}_G[x_S] = 0$ for S that is not a clique in G , which is equivalent to S not being an independent set in \bar{G} . We also note that $\Pi'_G = \Pi_{\bar{G}}$, and that the final pseudo-expectation is obtained by setting $\tilde{\mathbb{E}}[x_S] := \tilde{\mathbb{E}}_G[x_S]/\tilde{\mathbb{E}}_G[1]$; this is done so that the normalization condition $\tilde{\mathbb{E}}[1] = 1$ is satisfied.

We thus see that $\tilde{\mathbb{E}}$ satisfies the second additional condition of Theorem 2. Hence, in order to apply Theorem 2 to conclude Theorem 1, it suffices to argue that with high probability over G , it holds that $\tilde{\mathbb{E}}_G[x_i] \geq \frac{\omega}{n}(1 - n^{-\Omega(\varepsilon)})$ for all i . Indeed, if this holds then we have $\tilde{\mathbb{E}}[x_i] \geq \frac{\omega}{n}(1 - n^{-\Omega(\varepsilon)})$ also, and then we can apply Theorem 2 with $k = \frac{n}{\omega}(1 + n^{-\Omega(\varepsilon)})$ which finishes the proof. Thus, it remains to prove the following claim.

▷ **Claim 11.** For each i , $\tilde{\mathbb{E}}_G[x_i] \geq \frac{\omega}{n} \cdot (1 - n^{-\Omega(\varepsilon)})$ with probability $1 - n^{-\log n}$.

Proof. We have that $\tilde{\mathbb{E}}_G[x_i] := \sum_{T \subseteq \binom{[n]}{2}; |V(T)| \leq \tau} \left(\frac{\omega}{n}\right)^{|V(T) \cup \{i\}|} \chi_T(G)$, where $\tau \leq (\varepsilon/C) \log n$. The $T = \emptyset$ term always contributes $\frac{\omega}{n}$. The other terms all have $|V(T)| \geq 2$. Let H_1 be the set of T such that $i \in V(T)$, and let H_2 be the set of T such that $i \notin V(T)$. Let $H_1^{(t)}$ be the set of $T \in H_1$ with $|V(T)| = t$, and similarly for $H_2^{(t)}$. Each set $H_1^{(t)}$ can be partitioned into families $\{\mathcal{T}_{1,r}^{(t)}\}_{r=1}^{p_{1,t}}$ where T and T' are in the same family if there is a permutation $\sigma: [n] \rightarrow [n]$ such that $T = \sigma(T')$, or equivalently if T and T' are isomorphic. Similarly, each set $H_2^{(t)}$ can be partitioned into families $\{\mathcal{T}_{2,r}^{(t)}\}_{r=1}^{p_{2,t}}$.

We thus have

$$|\tilde{\mathbb{E}}_G[x_i] - \frac{\omega}{n}| \leq \sum_{t=2}^{\tau} \left[\left(\frac{\omega}{n}\right)^t \sum_{r=1}^{p_{1,t}} \left| \sum_{T \in \mathcal{T}_{1,r}^{(t)}} \chi_T(G) \right| + \left(\frac{\omega}{n}\right)^{t+1} \sum_{r=1}^{p_{2,t}} \left| \sum_{T \in \mathcal{T}_{2,r}^{(t)}} \chi_T(G) \right| \right].$$

► **Lemma 12.** Let \mathcal{T} be a family of subsets of $\binom{[n]}{2}$ such that $|V(T)| = t$ for every $T \in \mathcal{T}$, and for every $T, T' \in \mathcal{T}$, there exists $\sigma: [n] \rightarrow [n]$ such that $T = \sigma(T')$. Let $S = \bigcap_{T \in \mathcal{T}} V(T)$. Then for every $s \geq 0$ and even ℓ ,

$$\Pr_{G \sim G(n, 1/2)} \left[\left| \sum_{T \in \mathcal{T}} \chi_T(G) \right| \leq s \right] \geq 1 - \frac{n^{(t-|S|)\ell/2} \cdot (t\ell)^{t\ell}}{s^\ell}.$$

We postpone the proof of Lemma 12 to the end of the section, and now use it to finish the proof of Claim 11. Applying Lemma 12 with $\ell = (\log n)^2$, we get

$$\begin{aligned} \left| \sum_{T \in \mathcal{T}_{1,r}^{(t)}} \chi_T(G) \right| &\leq n^{(t-1)/2} (\log n)^{3t} \text{ with probability } \geq 1 - 2^{-t \log^2 n (\log \log n - \log t)} \\ \left| \sum_{T \in \mathcal{T}_{2,r}^{(t)}} \chi_T(G) \right| &\leq n^{t/2} (\log n)^{3t} \text{ with probability } \geq 1 - 2^{-t \log^2 n (\log \log n - \log t)}. \end{aligned}$$

We observe that $p_{1,t}$ and $p_{2,t}$ are both at most 2^{t^2} , as an equivalence class with t vertices is uniquely determined by a graph on t vertices. By union bound, we see that the above holds for all equivalence classes $\mathcal{T}_{1,r}^{(t)}$ and $\mathcal{T}_{2,r}^{(t)}$ with probability at least $1 - 2 \sum_{t=2}^{\tau} 2^{t^2 - t \log^2 n (\log \log n - \log t)}$. Since $t \leq \tau \leq (\varepsilon/C) \log n$, it follows that

$$\begin{aligned} \sum_{t=2}^{\tau} 2^{t^2 - t \log^2 n (\log \log n - \log t)} &= \sum_{t=2}^{\tau} 2^{t(t - \log^2 n (\log \log n - \log t))} \\ &\leq \sum_{t=2}^{\tau} 2^{t(\frac{\varepsilon}{C} \log n - (\log^2 n)(\log \log n - \log \frac{\varepsilon}{C} - \log \log n))} \\ &\leq \tau \cdot 2^{2 \log n \cdot (\frac{\varepsilon}{C} - \log \frac{C}{\varepsilon} \cdot \log n)} \leq n^{-\log n}, \end{aligned}$$

as $\frac{C}{\varepsilon} \geq C \geq 16$. Thus, with probability at least $1 - n^{-\log n}$, we have

$$\begin{aligned} |\tilde{\mathbb{E}}_G[x_i] - \frac{\omega}{n}| &\leq \sum_{t=2}^{\tau} \left[\left(\frac{\omega}{n}\right)^t 2^{t^2} n^{(t-1)/2} (\log n)^{3t} + \left(\frac{\omega}{n}\right)^{t+1} 2^{t^2} n^{t/2} (\log n)^{3t} \right] \\ &\leq \frac{2}{\sqrt{n}} \sum_{t=2}^{\tau} \left(\frac{\omega}{n}\right)^t 2^{t^2} (\log n)^{3t} n^{t/2} = 2 \left(\frac{\omega}{n}\right) \sum_{t=2}^{\tau} n^{(t-1)/2} n^{-(t-1)\varepsilon} 2^{t^2} (\log n)^{3t} n^{t/2} n^{-1/2} \end{aligned}$$

$$\begin{aligned}
&\leq 2 \left(\frac{\omega}{n}\right) \sum_{t=2}^{\tau} n^{-(t-1)\varepsilon} 2^{t^2} (\log n)^{3t} \leq \left(\frac{\omega}{n}\right) \cdot \max_{2 \leq t \leq \tau} 2n^{-(t-1)\varepsilon} 2^{t^2} (\log n)^{3t+1} \\
&\leq \left(\frac{\omega}{n}\right) \cdot 2n^\varepsilon (\log n) \max_{2 \leq t \leq \tau} (n^{-\varepsilon} 2^\tau (\log n)^3)^t \leq \left(\frac{\omega}{n}\right) \cdot n^\varepsilon n^{2\varepsilon/K} \max_{2 \leq t \leq \tau} (n^{-\varepsilon} n^{\varepsilon/C} \cdot n^{3\varepsilon/K})^t \\
&\leq \left(\frac{\omega}{n}\right) \cdot n^\varepsilon n^{2\varepsilon/K} n^{-2\varepsilon(1-1/C-3/K)} = \left(\frac{\omega}{n}\right) \cdot n^{-\varepsilon(1-2/C-8/K)} \leq \left(\frac{\omega}{n}\right) \cdot n^{-\varepsilon/2},
\end{aligned}$$

as $\varepsilon \geq C/\sqrt{\log n} \geq K \log \log n / \log n$ for $K \geq 32$ and $\tau \leq (\varepsilon/C) \log n$. Hence, with probability $1 - 1/n^{\log n}$, we have that $\mathbb{E}_G[x_i] = \frac{\omega}{n}(1 \pm n^{-\varepsilon/2})$, which completes the proof. \triangleleft

Proof of Lemma 12. Let $\ell \in \mathbb{N}$ be even. We have that

$$\mathbb{E}_{G \sim G(n, 1/2)} \left| \sum_{T \in \mathcal{T}} \chi_T(G) \right|^\ell = \mathbb{E}_{G \sim G(n, 1/2)} \left(\sum_{T \in \mathcal{T}} \chi_T(G) \right)^\ell = \sum_{T_1, \dots, T_\ell \in \mathcal{T}} \mathbb{E}_{G \sim G(n, 1/2)} \prod_{i=1}^{\ell} \chi_{T_i}(G).$$

We have that $\mathbb{E}_{G \sim G(n, 1/2)} \prod_{i=1}^{\ell} \chi_{T_i}(G) = 1$ iff $\bigoplus_{i=1}^{\ell} T_i = \emptyset$, that is, every edge in the multiset $\cup_{i=1}^{\ell} T_i$ appears an even number of times, and otherwise the term is 0. Since every edge in the multiset appears an even number of times, every vertex also appears an even number of times in $\cup_{i=1}^{\ell} V(T_i)$, and hence every vertex appears at least twice. Since $S \subseteq V(T_i)$ for all i , every vertex in S appears exactly ℓ times. So, the number of distinct vertices in $\cup_{i=1}^{\ell} (V(T_i) \setminus S)$ is at most $(t - |S|) \cdot \ell/2$. Each tuple (T_1, \dots, T_ℓ) with this property can thus be chosen by (1) selecting $(t - |S|) \cdot \ell/2$ distinct vertices S' (at most $n^{(t-|S|)\ell/2}$ choices), and then (2) choosing injections $\sigma_i: V(T) \rightarrow S'$ and setting $T_i = \sigma_i(T)$, where $T \in \mathcal{T}$ is an arbitrary fixed element (at most $(|S'|^t)^\ell \leq (t\ell)^{t\ell}$ choices). Thus, we get $\mathbb{E}_{G \sim G(n, 1/2)} \left| \sum_{T \in \mathcal{T}} \chi_T(G) \right|^\ell \leq n^{(t-|S|)\ell/2} (t\ell)^{t\ell}$. By Markov's inequality, it follows that $\Pr_{G \sim G(n, 1/2)} \left[\left| \sum_{T \in \mathcal{T}} \chi_T(G) \right| > s \right] = \Pr_{G \sim G(n, 1/2)} \left[\left| \sum_{T \in \mathcal{T}} \chi_T(G) \right|^\ell > s^\ell \right] \leq \frac{n^{(t-|S|)\ell/2} (t\ell)^{t\ell}}{s^\ell}$, which completes the proof. \blacktriangleleft

References

- 1 Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. *J. ACM*, 62(5):Art. 42, 25, 2015. doi:10.1145/2775105.
- 2 Sanjeev Arora, Béla Bollobás, László Lovász, and Iannis Tourlakis. Proving integrality gaps without knowing the linear program. *Theory Comput.*, 2:19–51, 2006. doi:10.4086/toc.2006.v002a002.
- 3 Sanjeev Arora, Eden Chlamtac, and Moses Charikar. New approximation guarantee for chromatic number. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 215–224. ACM, New York, 2006. doi:10.1145/1132516.1132548.
- 4 Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2):Art. 5, 37, 2009. doi:10.1145/1502793.1502794.
- 5 Mitali Bafna, Boaz Barak, Pravesh Kothari, Tselil Schramm, and David Steurer. Playing unique games on certified small-set expanders. *CoRR*, abs/2006.09969, 2020. arXiv:2006.09969.
- 6 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. arXiv:2012.02119.
- 7 Ainesh Bakshi and Pravesh Kothari. List-decodable subspace recovery via sum-of-squares. *CoRR*, abs/2002.05139, 2020. arXiv:2002.05139.
- 8 Ainesh Bakshi and Pravesh Kothari. Outlier-robust clustering of non-spherical mixtures, 2020. arXiv:2005.02970.
- 9 Jess Banks, Robert Kleinberg, and Cristopher Moore. The lovász theta function for random regular graphs and community detection in the hard regime. In *APPROX-RANDOM*, volume 81 of *LIPICs*, pages 28:1–28:22. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017.

- 10 Jess Banks, Robert Kleinberg, and Cristopher Moore. The lovász theta function for random regular graphs and community detection in the hard regime. *SIAM J. Comput.*, 48(3):1098–1119, 2019. doi:10.1137/18M1180396.
- 11 Boaz Barak, Siu On Chan, and Pravesh K. Kothari. Sum of squares lower bounds from pairwise independence [extended abstract]. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 97–106. ACM, New York, 2015.
- 12 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.
- 13 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.
- 14 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>.
- 15 Siavosh Benabbas, Siu On Chan, Konstantinos Georgiou, and Avner Magen. Tight gaps for vertex cover in the sherali-adams SDP hierarchy. In *FSTTCS*, volume 13 of *LIPICs*, pages 41–54. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2011.
- 16 Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8(1):269–289, 2012.
- 17 Siavosh Benabbas and Avner Magen. Extending SDP integrality gaps to sherali-adams with applications to quadratic programming and maxcutgain. In *IPCO*, volume 6080 of *Lecture Notes in Computer Science*, pages 299–312. Springer, 2010.
- 18 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. SIAM, 2012.
- 19 B. Bollobás. The chromatic number of random graphs. *Combinatorica*, 8(1):49–55, 1988. doi:10.1007/BF02122551.
- 20 Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for Sherali-Adams relaxations. In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, pages 283–292. ACM, New York, 2009.
- 21 Eden Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *FOCS*, pages 691–701. IEEE Computer Society, 2007.
- 22 Eden Chlamtac and Madhur Tulsiani. Convex relaxations and integrality gaps. In *Handbook on semidefinite, conic and polynomial optimization*, volume 166 of *Internat. Ser. Oper. Res. Management Sci.*, pages 139–169. Springer, New York, 2012. doi:10.1007/978-1-4614-0769-0_6.
- 23 Amin Coja-Oghlan. The lovász number of random graphs. *Comb. Probab. Comput.*, 14(4):439–465, 2005. doi:10.1017/S0963548305006826.
- 24 Ilias Diakonikolas, Samuel Hopkins, Daniel Kane, and Sushrut Karmalkar. Robustly learning any clusterable mixture of gaussians. *Personal Communication*, 2020.
- 25 Tommaso d’Orsi, Pravesh K. Kothari, Gleb Novikov, and David Steurer. Sparse PCA: algorithms, adversarial perturbations and certificates. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 553–564. IEEE, 2020. doi:10.1109/FOCS46700.2020.00058.
- 26 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. doi:10.1561/04000000086.
- 27 Mrinal Kanti Ghosh and Madhur Tulsiani. From weak to strong LP gaps for all csps. In *Computational Complexity Conference*, volume 79 of *LIPICs*, pages 11:1–11:27. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017.
- 28 Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-squares lower bounds for sherrington-kirkpatrick via planted affine planes, 2020. arXiv:2009.01874.

- 29 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. doi:10.1145/227683.227684.
- 30 D. Grigoriev. Complexity of Positivstellensatz proofs for the knapsack. *Comput. Complexity*, 10(2):139–154, 2001. doi:10.1007/s00037-001-8192-0.
- 31 Dima Grigoriev and Nicolai Vorobjov. Complexity of Null- and Positivstellensatz proofs. *Ann. Pure Appl. Logic*, 113(1-3):153–160, 2002. First St. Petersburg Conference on Days of Logic and Computability (1999). doi:10.1016/S0168-0072(01)00055-0.
- 32 Max Hopkins, Tali Kaufman, and Shachar Lovett. High dimensional expanders: Random walks, pseudorandomness, and unique games. *CoRR*, abs/2011.04658, 2020. arXiv:2011.04658.
- 33 S. B. Hopkins, P. K. Kothari, A. Potechin, P. Raghavendra, T. Schramm, and D. Steurer. The power of sum-of-squares for detecting hidden structures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731, 2017. doi:10.1109/FOCS.2017.72.
- 34 Sam B. Hopkins and Jerry Li. Mixture models, robustness, and sum of squares proofs, 2017.
- 35 Samuel B. Hopkins, Pravesh K. Kothari, and Aaron Potechin. Sos and planted clique: Tight analysis of MPW moments at all degrees and an optimal lower bound at degree four. *CoRR*, abs/1507.05230, 2015. arXiv:1507.05230.
- 36 Samuel B. Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *COLT*, volume 40 of *JMLR Workshop and Conference Proceedings*, pages 956–1006. JMLR.org, 2015.
- 37 Samuel B. Hopkins and David Steurer. Efficient bayesian estimation from few samples: Community detection and related problems. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 379–390. IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.42.
- 38 David R. Karger, Rajeev Motwani, and Madhu Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, 1998.
- 39 Sushrut Karmalkar, Adam R. Klivans, and Pravesh K. Kothari. List-decodable linear regression. *CoRR*, abs/1905.05679, 2019. arXiv:1905.05679.
- 40 Pravesh Kothari, Ryan O’Donnell, and Tselil Schramm. SOS lower bounds with hard constraints: think global, act local. *CoRR*, abs/1809.01207, 2018. arXiv:1809.01207.
- 41 Pravesh K. Kothari and Ruta Mehta. Sum-of-squares meets nash: lower bounds for finding any equilibrium. 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 1241–1248. ACM, 2018. doi:10.1145/3188745.3188892.
- 42 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.
- 43 Pravesh K. Kothari and Jacob Steinhardt. Better agnostic clustering via relaxed tensor norms, 2017.
- 44 Pravesh K. Kothari and David Steurer. Outlier-robust moment-estimation via sum-of-squares. *CoRR*, abs/1711.11581, 2017. arXiv:1711.11581.
- 45 Dmitriy Kunisky and Afonso S. Bandeira. A tight degree 4 sum-of-squares lower bound for the sherrington-kirkpatrick hamiltonian. *CoRR*, abs/1907.11686, 2019. arXiv:1907.11686.
- 46 Dmitriy Kunisky, Alexander S. Wein, and Afonso S. Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. *CoRR*, abs/1907.11636, 2019. arXiv:1907.11636.
- 47 Jean Bernard Lasserre. Optimisation globale et théorie des moments. *C. R. Acad. Sci. Paris Sér. I Math.*, 331(11):929–934, 2000. doi:10.1016/S0764-4442(00)01750-X.
- 48 Allen Liu and Ankur Moitra. Settling the robust learnability of mixtures of gaussians. *CoRR*, abs/2011.03622, 2020. arXiv:2011.03622.
- 49 László Lovász and Alexander Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

- 50 Tengyu Ma, Jonathan Shi, and David Steurer. Polynomial-time tensor decompositions with sum-of-squares. In *FOCS*, pages 438–446. IEEE Computer Society, 2016.
- 51 Sidhanth Mohanty, Prasad Raghavendra, and Jeff Xu. Lifting sum-of-squares lower bounds: Degree-2 to degree-4. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, page 840–853, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3357713.3384319.
- 52 Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.
- 53 Prasad Raghavendra and Tselil Schramm. Tight lower bounds for planted clique in the degree-4 SOS program. *CoRR*, abs/1507.05136, 2015. arXiv:1507.05136.
- 54 Prasad Raghavendra and David Steurer. Integrality gaps for strong SDP relaxations of Unique Games. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2009*, pages 575–585. IEEE Computer Soc., Los Alamitos, CA, 2009. doi:10.1109/FOCS.2009.73.
- 55 Prasad Raghavendra and Morris Yau. List decodable learning via sum of squares. *CoRR*, abs/1905.04660, 2019. arXiv:1905.04660.
- 56 Prasad Raghavendra and Morris Yau. List decodable subspace recovery, 2020. arXiv:2002.03004.
- 57 Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *FOCS*, pages 593–602. IEEE Computer Society, 2008.
- 58 Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovasz-Schrijver LP relaxations of vertex cover and max cut. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 302–310. ACM, New York, 2007. doi:10.1145/1250790.1250836.
- 59 Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3(3):411–430, 1990. doi:10.1137/0403036.
- 60 Madhur Tulsiani. CSP gaps and reductions in the lasserre hierarchy. In *STOC*, pages 303–312. ACM, 2009.
- 61 Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy [extended abstract]. In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, pages 303–312. ACM, New York, 2009.

A Satisfying the booleanity, edge and positivity constraints

We prove the following three simple claims.

- ▷ Claim 13. $\tilde{\mathbb{E}}^{\otimes k}$ and $\tilde{\mathbb{E}}'$ satisfy the booleanity constraints $\{x_{i,c}^2 = x_{i,c} : i \in [n], c \in [k]\}$.
- ▷ Claim 14. $\tilde{\mathbb{E}}^{\otimes k}$ and $\tilde{\mathbb{E}}'$ satisfy the edge constraints $\{x_{i,c}x_{j,c} = 0 : (i, j) \in E(G), c \in [k]\}$.
- ▷ Claim 15. $\tilde{\mathbb{E}}^{\otimes k}$ and $\tilde{\mathbb{E}}'$ satisfy the positivity constraint.

Proof of Claim 13. Since $\tilde{\mathbb{E}}'$ is obtained by restricting $\tilde{\mathbb{E}}^{\otimes k}$ to a smaller domain, it suffices to show that $\tilde{\mathbb{E}}^{\otimes k}$ satisfies the constraints. We observe that $\tilde{\mathbb{E}}^{\otimes k}$ satisfies the above constraints if and only if for all monomials $\prod_{c=1}^k \prod_{i \in S_c} x_{i,c}^{\alpha_{i,c}}$ (where each $\alpha_{i,c} \geq 1$), it holds that $\tilde{\mathbb{E}}^{\otimes k}[\prod_{c=1}^k \prod_{i \in S_c} x_{i,c}^{\alpha_{i,c}}] = \tilde{\mathbb{E}}^{\otimes k}[\prod_{c=1}^k \prod_{i \in S_c} x_{i,c}]$. We have that $\tilde{\mathbb{E}}^{\otimes k}[\prod_{c=1}^k \prod_{i \in S_c} x_{i,c}^{\alpha_{i,c}}] = \prod_{c=1}^k \tilde{\mathbb{E}}[\prod_{i \in S_c} x_i^{\alpha_{i,c}}] = \prod_{c=1}^k \tilde{\mathbb{E}}[\prod_{i \in S_c} x_i] = \tilde{\mathbb{E}}^{\otimes k}[\prod_{c=1}^k \prod_{i \in S_c} x_{i,c}]$, as $\tilde{\mathbb{E}}$ satisfies the constraints $x_i^2 = x_i$, and so we are done. ◁

Proof of Claim 14. Since $\tilde{\mathbb{E}}'$ is obtained by restricting $\tilde{\mathbb{E}}^{\otimes k}$ to a smaller domain, it suffices to show that $\tilde{\mathbb{E}}^{\otimes k}$ satisfies the constraints. We observe that $\tilde{\mathbb{E}}^{\otimes k}$ satisfies the above constraints if and only if for all multilinear monomials $\prod_{c=1}^k x_{S_c,c}$ of coloring degree at most $d-2$, it holds

that $\tilde{\mathbb{E}}^{\otimes k}[x_{i,c}x_{j,c}\prod_{c'=1}^k x_{S_{c'},c'}] = 0$. This is because by Claim 13, we can reduce any polynomial modulo the booleanity constraints to make it multilinear. Using the tensor product structure, we have $\tilde{\mathbb{E}}^{\otimes k}[x_{i,c}x_{j,c}\prod_{c'=1}^k x_{S_{c'},c'}] = \prod_{c' \neq c} \tilde{\mathbb{E}}[x_{S_{c'}}] \cdot \tilde{\mathbb{E}}[x_{S_c}x_{i,j}] = \prod_{c' \neq c} \tilde{\mathbb{E}}[x_{S_{c'}}] \cdot 0 = 0$, since $\tilde{\mathbb{E}}$ satisfies the edge constraints. This completes the proof. \triangleleft

Proof of Claim 15. Since $\tilde{\mathbb{E}}'$ is obtained by restricting $\tilde{\mathbb{E}}^{\otimes k}$ to a smaller domain, it suffices to prove the claim only for $\tilde{\mathbb{E}}^{\otimes k}$. Let \mathcal{M} be the moment matrix of $\tilde{\mathbb{E}}$. That is, \mathcal{M} is the matrix indexed by sets (S, T) with $|S|, |T| \leq d/2$ and $\mathcal{M}(S, T) := \tilde{\mathbb{E}}[x_S x_T]$. We note that for any $f \in \mathcal{P}_{d/2}^n$, $\tilde{\mathbb{E}}[f^2] = f^\top \mathcal{M} f$, where we interpret f as a vector of coefficients in the second expression. The moment matrix of $\tilde{\mathbb{E}}^{\otimes k}$ is indexed by tuples of sets $((S_1, \dots, S_k), (T_1, \dots, T_k))$ where $|S_c|, |T_c| \leq d/2$ for all $c \in [k]$. We observe that the moment matrix of $\tilde{\mathbb{E}}^{\otimes k}$ is $\mathcal{M}^{\otimes k}$, as the $((S_1, \dots, S_k), (T_1, \dots, T_k))$ -th entry is $\tilde{\mathbb{E}}^{\otimes k}[\prod_{c=1}^k x_{S_c} x_{T_c}] = \prod_{c=1}^k \tilde{\mathbb{E}}[x_{S_c} x_{T_c}] = \prod_{c=1}^k \mathcal{M}(S_c, T_c)$. We also note that for any f with $\text{cdeg}(f) \leq d/2$, it holds that $\tilde{\mathbb{E}}^{\otimes k}[f^2] = f^\top \mathcal{M}^{\otimes k} f \geq 0$, as the tensor product of a positive semidefinite matrix is also positive semidefinite. This shows that $\tilde{\mathbb{E}}^{\otimes k}[f^2] \geq 0$ for all f with $\text{cdeg}(f) \leq d/2$, which finishes the proof. \triangleleft

B Tightness of degree in Theorem 1

In this section, we prove the following lemma, showing that the upper bound on d in Theorem 1 is tight up to constant factors.

► **Lemma 16.** *With high probability over $G \sim G(n, 1/2)$, there is no degree $8(1 + o(1)) \log_2 n$ coloring pseudo-expectation for G using $k \leq \frac{n}{e \cdot 2^{(1+o(1)) \log_2 n}}$ colors.*

Let $t = 2(1 + o(1)) \log_2 n$. We show that with high probability over $G \sim G(n, 1/2)$, there is no degree $4t$ coloring pseudo-expectation for G using $k \leq \frac{n}{et}$ colors. We first observe that with high probability, the maximum independent set in G has size at most t . Suppose that we draw $G \sim G(n, 1/2)$ such that this holds, and suppose that such a pseudo-expectation $\tilde{\mathbb{E}}'$ exists. We observe that there is a natural action of permutations $\sigma: [k] \rightarrow [k]$ on $\tilde{\mathbb{E}}'$, given by $\tilde{\mathbb{E}}'^{(\sigma)}[\prod_{c=1}^k \prod_{i \in S_c} x_{i,c}] := \tilde{\mathbb{E}}'[\prod_{c=1}^k \prod_{i \in S_{\sigma(c)}} x_{i,\sigma(c)}]$. Let $\tilde{\mathbb{E}}'' := \mathbb{E}_\sigma \tilde{\mathbb{E}}'^{(\sigma)}$ be the pseudo-expectation obtained by averaging over all σ . We then have that $\tilde{\mathbb{E}}''$ satisfies the coloring constraints and is symmetric with respect to the color classes, e.g. that $\tilde{\mathbb{E}}''[x_{i,c}] = \tilde{\mathbb{E}}''[x_{i,c'}]$ for all $c, c' \in [k]$. This implies that $\tilde{\mathbb{E}}''[x_{i,1}] = \frac{1}{k} \sum_{c=1}^k \tilde{\mathbb{E}}''[x_{i,c}] \geq \frac{1}{k} \cdot 1$. Let $\tilde{\mathbb{E}}$ be the projection of $\tilde{\mathbb{E}}''$ onto the first color, so that $\tilde{\mathbb{E}}[\prod_{i \in S} x_i] := \tilde{\mathbb{E}}''[\prod_{i \in S} x_{i,1}]$. We then see that $\tilde{\mathbb{E}}$ is a degree $4t$ independent set pseudo-expectation with $\tilde{\mathbb{E}}[\sum_i x_i] \geq \omega$, where $\omega := \frac{n}{k} \geq et$.

To complete the proof, we show the following lemma.

► **Lemma 17.** *Suppose that the maximum independent set in G has size $\leq t$. Then there is no degree $4t$ independent set pseudo-expectation $\tilde{\mathbb{E}}$ for G with $\tilde{\mathbb{E}}[\sum_i x_i] = \omega \geq et$.*

Proof. Suppose that such a pseudo-expectation $\tilde{\mathbb{E}}$ exists. Let $f = \sum_i x_i$, and let $\ell \in \mathbb{N}$ be the smallest integer so that $2^\ell \geq 2t$. Note that $2^\ell \leq 4t$ must hold also. By Cauchy-Schwarz, we have

$$\begin{aligned} \tilde{\mathbb{E}}[f^{2^\ell}] &\geq (\tilde{\mathbb{E}}[f^{2^{\ell-1}}])^2, \\ \tilde{\mathbb{E}}[f^{2^{\ell-1}}] &\geq \tilde{\mathbb{E}}[f^{2^{\ell-2}}]^2 \geq \dots \geq \tilde{\mathbb{E}}[f]^{2^{\ell-1}}, \\ \implies \tilde{\mathbb{E}}[f^{2^\ell}] &\geq (\tilde{\mathbb{E}}[f^{2^{\ell-1}}]) \cdot (\tilde{\mathbb{E}}[f])^{2^{\ell-1}} = \tilde{\mathbb{E}}[f^{2^{\ell-1}}] \cdot \omega^{2^{\ell-1}}. \end{aligned}$$

Note that each polynomial above has degree at most $2^\ell \leq 4t$, so the above pseudo-expectations are all well-defined. Now, we observe that

$$\tilde{\mathbb{E}}[f^{2^{\ell-1}}] = \tilde{\mathbb{E}}\left[\sum_{S \subseteq [n]: |S| \leq 2^{\ell-1}} c_S x_S\right] = \sum_{S: |S| \leq t, S \text{ indep set in } G} c_S \tilde{\mathbb{E}}[x_S],$$

$$\tilde{\mathbb{E}}[f^{2^\ell}] = \sum_{S: |S| \leq t, S \text{ indep set in } G} c'_S \tilde{\mathbb{E}}[x_S]$$

where the coefficients c_S and c'_S are each nonnegative integers. Notice that $c'_S \leq |S|^{2^\ell}$, as every contribution to x_S is made by choosing an $i \in S$ from each of the $\sum_i x_i$ factors. We also observe that $c_S \geq |S|^{2^{\ell-1}-|S|} \cdot (|S|!)$, as we can choose each $i \in S$ exactly once from the first $|S|$ factors, and then select an arbitrary $i \in S$ from the remaining $2^{\ell-1} - |S|$ factors. Note that here we use the fact that $|S| \leq t \leq 2^{\ell-1}$ always holds. Fix S , and let $s = |S|$. We observe that

$$\frac{c'_S}{c_S} \leq \frac{s^{2^\ell}}{s^{2^{\ell-1}-s} \cdot s!} \leq s^{2^{\ell-1}} \cdot s^s \cdot \frac{1}{\sqrt{2\pi} \cdot s^{s+\frac{1}{2}} e^{-s}} < s^{2^{\ell-1}} \cdot e^s \leq (e \cdot s)^{2^{\ell-1}} \leq \omega^{2^{\ell-1}},$$

using Stirling's approximation and the fact that $\omega \geq et \geq es$. It therefore follows that $(c'_S - c_S \omega^{2^{\ell-1}}) \tilde{\mathbb{E}}[x_S] < 0$. Hence,

$$\tilde{\mathbb{E}}[f^{2^\ell}] - \tilde{\mathbb{E}}[f^{2^{\ell-1}}] \cdot \omega^{2^{\ell-1}} = \sum_{S: |S| \leq t, S \text{ indep set in } G} (c'_S - c_S \omega^{2^{\ell-1}}) \tilde{\mathbb{E}}[x_S] < 0,$$

which is a contradiction. ◀