Clique Is Hard on Average for Sherali-Adams with Bounded Coefficients*

Susanna F. de Rezende
Lund University

Aaron Potechin
University of Chicago

Kilian Risse
EPFL

April 26, 2024

Abstract

We prove that Sherali-Adams with polynomially bounded coefficients requires proofs of size $n^{\Omega(d)}$ to rule out the existence of an $n^{\Theta(1)}$-clique in Erdős-Rényi random graphs whose maximum clique is of size $d \leqslant 2 \log n$. This lower bound is tight up to the multiplicative constant in the exponent. We obtain this result by introducing a technique inspired by pseudo-calibration which may be of independent interest. The technique involves defining a measure on monomials that precisely captures the contribution of a monomial to a refutation. This measure intuitively captures progress and should have further applications in proof complexity.

*This is the full-length version of a paper with the title *"Clique Is Hard on Average for Unary Sherali-Adams"* that appeared in the *Proceedings of the 64th Annual IEEE Symposium on Foundations of Computer Science (FOCS '23)*.

Contents

1 Introduction

The problem of identifying a maximum clique in a given graph, that is, finding a fully connected subgraph of maximum size, is one of the fundamental problems of theoretical computer science. Already mentioned by Cook [Coo71] in his seminal paper introducing the theory of **NP**-complete problems, it was one of the first combinatorial problems proven **NP**-hard by Karp [Kar72]. Building on the PCP theorem this result was later strengthened to even rule out polynomial time algorithms that approximate the maximum clique size within a factor of $n^{1-\varepsilon}$ [Hås99, Zuc07], unless $\mathbf{P} = \mathbf{NP}$.

A related problem is k-clique: given an n-vertex graph, determine whether it contains a clique of size k. This problem can be solved in time $O(n^k)$ by iterating over all subsets of vertices of size k and checking whether one of them is a clique. This naïve algorithm is essentially the fastest known; a clever use of fast matrix multiplication [NP85] allows a slight improvement upon the constant in the exponent but no algorithms with a sublinear dependence on k in the exponent are known.

If we only assume $\mathbf{P} \neq \mathbf{NP}$, it is unknown whether there are faster algorithms for k-clique. However, improving upon the linear dependence on k in the exponent would disprove the exponential time hypothesis [CHKX04] and getting rid of the dependence on k altogether would imply that the class of fixed parameter tractable problems collapses to $\mathbf{W}[1]$ [DF95]. Hence, if one is willing to make the strong assumption that the exponential time hypothesis holds, then the naïve algorithm has essentially optimal running time in the worst-case.

Besides studying k-clique in the worst-case, one may consider it in the average-case setting. Suppose the given graph is an Erdős-Rényi graph with edge probability around the threshold of containing a k-clique. Does k-clique require time $n^{\Omega(k)}$ on such graphs? Or, even less ambitiously, is there an algorithm running in time $n^{o(k)}$ that decides the n^{ε}-clique problem on such graphs? It is unlikely that the hardness of such average-case questions can be based on worst-case hardness assumptions such as $\mathbf{P} \neq \mathbf{NP}$ or the exponential time hypothesis [BT06]. They are, in fact, being used as hardness assumptions themselves: the *planted clique conjecture* states that $n^{1/2-\varepsilon}$-clique requires time $n^{\Omega(\log n)}$ on Erdős-Rényi graphs with edge probability $1/2$.

In order to obtain *unconditional* lower bounds – that do not rely on hardness assumptions – for such average-case questions, we focus on limited models of computation. This approach has turned out to be quite fruitful and several results of this form have emerged over the past few decades. For Boolean circuits, Rossman [Ros08, Ros10] proved two remarkable results: he showed that monotone circuits, i.e., circuits consisting of $\vee$ and $\wedge$ gates only, as well as circuits of constant depth require size $\Omega(n^{k/4})$ to refute the existence of a k-clique in the average-case setting.

Instead of studying circuits, it is also possible to approach this problem through the lens of proof complexity. Very broadly, proof complexity studies certificates of unsatisfiability of propositional formulas. As we cannot argue about certificates of unsatisfiability in general we consider certificates of a certain form, or in terms of proof complexity, refutations in a given proof system. For instance, if we prove that any certificate in a proof system P that witnesses that a given n-vertex graph contains no k-clique requires length $n^{\Omega(k)}$ on average, then we immediately obtain average-case $n^{\Omega(k)}$ running time lower bounds for any algorithm whose trace can be interpreted as a proof in the system P. It is often the case that state-of-the-art algorithms can be captured by seemingly simple proof systems, as was shown to be the case for clique algorithms [ABdR+21].

It is often the case that weak proof systems are sensitive to the precise encoding of combinatorial principles. The k-clique formula is no exception: it is somewhat straightforward to prove almost optimal $n^{\Omega(k)}$ resolution size lower bounds for the less usual binary encoding of the k-clique formula [LPRT17] and these lower bounds can even be extended to an $n^{\Omega(k)}$ lower bound for the Res(s) proof system for constant s [DGGM20]. For the more natural unary encoding not much is known. There are essentially optimal $n^{\Omega(k)}$ average-case size lower bounds for regular resolution [ABdR+21, Pan21a] and tree-like resolution [BGLR12, Lau18]. For resolution, there are two average-case lower bounds that hold in different regimes: for $n^{5/6} \ll k \leqslant n/3$, Beame et al. [BIS07] proved an average-case $\exp(n^{\Omega(1)})$ size lower bound and for $k \leqslant n^{1/3}$, Pang [Pan21a] proved a $2^{k^{1-o(1)}}$ lower bound. It is a long standing open problem, mentioned, e.g., in [BGLR12], to prove an unconditional $n^{\Omega(k)}$ resolution size lower bound for the unary encoding – even in the worst case.

Little is known about the average-case hardness of the k-clique formula in the semi-algebraic setting. There are optimal degree lower bounds for $k \leqslant n^{1/2-\varepsilon}$ for the Sum-of-Squares proof system [MPW15, BHK+19, Pan21b], but there are no non-trivial lower bounds on size. For Nullstellensatz, however, if restricted to not use dual variables, then size lower bounds follow by a simple syntactic argument [Mar08]. Prior to our work no other size lower bounds were known for algebraic or semi-algebraic proof systems.

1.1 Our Result

In this work we establish that Sherali-Adams [SA94, DM13] with polynomially bounded coefficients requires size $n^{\Omega(D)}$ to refute the $n^{1/100}$-clique formula on random graphs whose maximum clique size is of size $D \leqslant 2\log n$. Qualitatively this establishes the planted clique conjecture for Sherali-Adams with polynomially bounded coefficients. This is the first size lower bound on the clique formula for a semi-algebraic proof system.

Theorem 1.1 (Informal). For all integers $n \in \mathbb{N}^+$ and $D \leqslant 2\log n$, if $G \sim \mathcal{G}(n, n^{-2/D})$ is an Erdős-Rényi random graph, then it holds asymptotically almost surely that Sherali-Adams with polynomially bounded coefficients requires size at least $n^{\Omega(D)}$ to refute the claim that G contains a clique of size k, for any $k \leqslant n^{1/67}$.

Note that Sherali-Adams with polynomially bounded coefficients is stronger than unary Sherali-Adams and is incomparable to resolution [GHJ+22]. Our result further applies to the SubCubeSums proof system [FMSV23] as our proof strategy gives a lower bound on the sum of the magnitude of the coefficients of a Sherali-Adams refutation, ignoring Boolean axioms.

Let us stress that the size lower bound holds regardless of the degree of the refutation. This is a somewhat unique feature of our technique – all other lower bound strategies for Sherali-Adams and Sum-of-Squares are tailored to proving degree lower bounds, which, if strong enough, imply size lower bounds by the size-degree relation [AH19]. Since the clique formula has refutations of degree D we cannot expect to obtain size lower bounds through this connection for $D \leqslant \sqrt{n}$. We therefore introduce a new technique, inspired by pseudo-calibration [BHK+19], that is more refined – for any monomial m, of arbitrary degree, we determine a lower bound on the size of the smallest bounded-coefficient Sherali-Adams derivation of m.

1.2 Organization

The rest of this paper is organized as follows. In Section 2 we introduce some basic terminology to then outline our proof strategy in Section 3 where we also attempt to convey some intuition.

With the motivation at hand from Section 3 we then go on to define the central combinatorial concept of a *core* of a graph in Section 4 and a notion of pseudorandomness in Section 5. We proceed in Section 6 to prove the main theorem for any graph satisfying our notion of pseudorandomness, but postpone the proof of one of the main lemmas to Section 7. Finally, in Section 8, we show that a random graph satisfies our pseudorandomness property and conclude with some open problems in Section 9.

2 Preliminaries

Natural logarithms (base e) are denoted by ln, whereas base 2 logarithms are denoted by log. For integers $n \geqslant 1$ we introduce the shorthand $[n] = \{1, 2, \ldots, n\}$ and sometimes identify singletons $\{u\}$ with the element u. Let $\binom{S}{\ell}$ denote the set of subsets of S of size ℓ and, for a given a random variable X and an event P, we denote by $\mathbb{1}_P(X)$ the indicator random variable that is 1 if P holds and 0 otherwise.

2.1 Semantic Sherali-Adams

Let $\mathcal{P} = \{p_1 = 0, \ldots, p_m = 0\}$ be a polynomial system of equations over Boolean variables $x_1, \ldots, x_n$ and their twin variables $\bar{x}_1, \ldots, \bar{x}_n$. Denote by

$$B_n = \{x_i(1 - x_i) \mid i \in [n]\} \cup \{\bar{x}_i(1 - \bar{x}_i) \mid i \in [n]\} \cup \{1 - \bar{x}_i - x_i \mid i \in [n]\} \tag{2.1}$$

the corresponding Boolean axioms and negation axioms and let I_{B_n} denote the ideal generated by B_n, that is, I_{B_n} consists of all polynomials of the form $\sum_j r_j q_j$ where the r_j are arbitrary polynomials and $q_j \in B_n$. For an ideal I and polynomials p and q we write $p \equiv q \mod I$ if $p - q \in I$.

A *semantic Sherali-Adams refutation of* $\mathcal{P}$ is a sequence of polynomials $(g_1, \ldots, g_m, f_0)$ such that f_0 is of the form

$$f_0 = \sum_{\substack{A,B \subseteq [n] \\ \alpha_{A,B} \geqslant 0}} \alpha_{A,B} \prod_{i \in A} x_i \prod_{i \in B} \bar{x}_i \tag{2.2}$$

and it holds that

$$\sum_{j \in [m]} g_j p_j + f_0 \equiv -1 \mod I_{B_n} . \tag{2.3}$$

The *size* of a refutation π, denoted by $\mathrm{Size}(\pi)$, is the sum of the size of the binary encodings of the non-zero coefficients on the left hand side of (2.3) when all polynomials are expanded as a sum of monomials (without cancellations), while the *coefficient size* of π is the sum of the magnitudes of the coefficients of the aforementioned monomials.

We note that this definition differs from that of the usual Sherali-Adams proof system [SA94, DM13] where Boolean axioms and negation axioms are written out explicitly and the size is measured taking also these axioms into account. We are not the first to disregard the size contribution of these axioms: most size lower bounds for Sherali-Adams also apply in this setting. The question of whether these two size measures are polynomially related was raised explicitly in [FHR+24], where the above system is referred to as *succinct* Sherali-Adams, and remains open.

To verify that semantic Sherali-Adams is a Cook-Reckhow proof system, we need to check that a semantic Sherali-Adams refutation is verifiable in time polynomial in the size of the refutation. This was originally shown in [FHR+24], but we provide a simpler and more direct proof of this fact.

Proposition 2.1. Let $\mathcal{P}$ be a polynomial system of equations over n Boolean variables and their twin variables. A semantic Sherali-Adams refutation π of $\mathcal{P}$ can be verified in time $O\big(\mathrm{Size}(\pi)^2 \cdot \mathrm{poly}(n)\big)$.

Proof. Denote the coefficient of a monomial m in a polynomial $p \in \mathcal{P}$ by $\gamma_p(m)$. Further, given a monomial m, let $\mathbb{1}_m$ be the vector of length 2^n, indexed by assignments $\rho \in \{0,1\}^n$, that corresponds to the truth table of m: the entry with index ρ is 1 if the monomial m evaluates to 1 under ρ and 0 otherwise. We denote the all-ones vector $\mathbb{1}_1$ by $\mathbb{1}$, and the all-zero vector $\mathbb{1}_0$ by $\mathbb{0}$.

With this notation at hand we may equivalently define a semantic Sherali-Adams refutation of $\mathcal{P}$ as a sequence of rational vectors $(\beta_1, \dots, \beta_{|\mathcal{P}|}, \alpha)$ indexed by multilinear monomials with $\alpha \geqslant 0$ and

$$\sum_{p \in \mathcal{P}} \sum_{m_1, m_2} \beta_p(m_1) \cdot \gamma_p(m_2) \cdot \mathbb{1}_{m_1 \cdot m_2} - \sum_{m} \alpha(m) \cdot \mathbb{1}_m - \mathbb{1} = \mathbb{0} \ . \tag{2.4}$$

View the left hand side as a vector π indexed by assignments $\rho \in \{0,1\}^n$. Since the vector π is rational, we have that π is the 0-vector if and only if the inner product is 0, that is, $\langle \pi, \pi \rangle = 0$. This can be efficiently checked by expanding the inner product

$$\begin{aligned} \langle \pi, \pi \rangle = &\Big(\sum_{p \in \mathbb{P}} \sum_{m_1, m_2} \beta_p(m_1) \cdot \gamma_p(m_2) \cdot \mathbb{1}_{m_1 \cdot m_2} \Big)^2 + \Big(\sum_{m} \alpha(m) \cdot \mathbb{1}_m \Big)^2 + \langle \mathbb{1}, \mathbb{1} \rangle \\ &- 2 \sum_{p \in \mathcal{P}} \sum_{m_1, m_2} \sum_{m} \alpha(m) \cdot \beta_p(m_1) \cdot \gamma_p(m_2) \cdot \langle \mathbb{1}_{m_1 \cdot m_2}, \mathbb{1}_m \rangle \\ &- 2 \sum_{p \in \mathcal{P}} \sum_{m_1, m_2} \beta_p(m_1) \cdot \gamma_p(m_2) \cdot \langle \mathbb{1}_{m_1 \cdot m_2}, \mathbb{1} \rangle + 2 \sum_{m} \alpha(m) \cdot \langle \mathbb{1}_m, \mathbb{1} \rangle \ . \end{aligned} \tag{2.5}$$

Observe that in the above expression all the inner products can be efficiently computed. Hence checking whether π is indeed a valid semantic Sherali-Adams refutation boils down to verifying that the above sum of bounded rationals is equal to 0. $\square$

As the distinction between semantic Sherali-Adams and Sherali-Adams is not essential for what follows, we refer to semantic Sherali-Adams simply as Sherali-Adams going forward. A Sherali-Adams refutation π of $\mathcal{P}$ is a *Sherali-Adams refutation with* $f(n)$*-bounded-coefficients* if the magnitude of all coefficients is bounded by $f(n)$ and we call π a *Sherali-Adams refutation with* poly*-bounded-coefficients* if for some constant $c > 0$ it holds that π is a Sherali-Adams refutation with n^c-bounded-coefficients. Let us record the following observation.

Proposition 2.2. If Sherali-Adams requires coefficient size s to refute $\mathcal{P}$, then Sherali-Adams with $f(n)$-bounded-coefficients requires size $s/f(n)$ to refute $\mathcal{P}$.

Proof. As every coefficient in an $f(n)$-bounded-coefficient refutation is bounded by $f(n)$, there need to be at least $s/f(n)$ monomials with a non-zero coefficient. $\square$

Unary Sherali-Adams is a subsystem of Sherali-Adams where all coefficients of monomials are either $+1$ or -1 and the right-hand-side of Equation (2.3) is any negative integer

$$\sum_{j\in[m]} g_j p_j + f_0 \equiv -M \mod I_{B_n} \ , \tag{2.6}$$

where f_0 is again a non-negative sum of monomials.

Proposition 2.3. If Sherali-Adams requires coefficient size s to refute $\mathcal{P}$, then unary Sherali-Adams requires size at least s to refute $\mathcal{P}$.

Proof. We can transform any unary Sherali-Adams refutation of size s, summing to an integer $-M$, to a Sherali-Adams refutation of coefficient size at most s by dividing the left hand side by $M \geqslant 1$. □

2.2 Graph Theory

Before defining the k-clique formula, we introduce some terminology and notation that we use throughout the paper.

Unless stated otherwise, G denotes a k-partite graph with partitions $V_1, \ldots, V_k$ of size n each. We call a partition V_i a *block* and, for $S \subseteq [k]$, denote by V_S the vertices in blocks in S, that is, $V_S = \bigcup_{i\in S} V_i$. For disjoint sets $W_1, \ldots, W_s$ we let a *tuple* $t = (w_1, \ldots, w_s)$ be a sequence of vertices satisfying $w_i \in W_i$ for all $i \in [s]$. All tuples we consider are defined with respect to the partition $V_1, \ldots, V_k$, though, at times, may only be defined over a subset of the blocks, that is, not all tuples are of size k. For a tuple $t = (v_1, \ldots, v_k)$ and a set $S \subseteq [k]$ we denote the projection of t onto S by $t_S = (v_i \mid i \in S)$. An *s-tuple* is a tuple of size s and sometimes it is convenient for us to think of a tuple as a set of vertices. We take the liberty to interchangeably identify a tuple as a sequence as well as a set and hope that this causes no confusion.

A set Q of tuples is a *rectangle* if for some $S \subseteq [k]$ the set Q can be written as the Cartesian product of sets $U_i \subseteq V_i$ for $i \in S$, i.e., $Q = \bigtimes_{i\in S} U_i$; in other words, Q contains all tuples $t = (u_i \mid i \in S)$ satisfying $u_i \in U_i$ for all $i \in S$. Rectangles, unless explicitly stated, consist of k-tuples only, that is, $Q = \bigtimes_{i\in[k]} U_i$. Given a rectangle Q and a set $S \subseteq [k]$ we let Q_S be the projection of Q onto the blocks in S: if $Q = \bigtimes_{i\in[k]} U_i$, then $Q_S = \bigtimes_{i\in S} U_i$ and, in particular, we have $Q_i = U_i$ for $i \in [k]$.

While G always denotes a large graph, the graphs H and F denote small graphs: throughout the paper H and F are graphs on k labeled vertices. Usually these graphs have a small vertex cover and graphs denoted by F furthermore have many isolated vertices. For a graph H we denote the minimum vertex cover by $\mathrm{vc}(H)$ and sometimes refer to H as a *pattern graph*, whereas F is usually a *core graph* (see Section 4). We denote by $\mathcal{H}$ the set of graphs on k labeled vertices and for a parameter $i \in \mathbb{N}^+$ let $\mathcal{H}_i \subseteq \mathcal{H}$ be the family of graphs with a minimum vertex cover of size at most i, that is, all graphs $H \in \mathcal{H}_i$ satisfy $\mathrm{vc}(H) \leqslant i$.

We record here two simple lemmas about graphs that will come in handy throughout the paper.

Lemma 2.4. There are at most $2^{c\log k + b(c-(b+1)/2)} \leqslant 2^{c(b+\log k)}$ graphs H over k vertices with a vertex cover of size b and $\left|V\big(E(H)\big)\right| \leqslant c$.

Proof. We first choose the b vertices from the k vertices that form the vertex cover. Then, from the remaining $k-b$ vertices, we choose $c-b$ vertices that may be incident to an edge. We can add edges that are incident to the vertex cover and the other $c-b$ vertices and thus get that there are at most

$$\binom{k}{b}\binom{k-b}{c-b}2^{\binom{b}{2}}2^{b(c-b)} \leqslant 2^{c\log k+b(c-(b+1)/2)} \tag{2.7}$$

many such graphs. □

Recall that a maximal matching of H is a matching that cannot be extended in H.

Proposition 2.5. Any maximal matching in a graph H is of size at least $\lceil \mathrm{vc}(H)/2 \rceil$.

Proof. Since M is maximal, all edges of H are incident to $V(M)$. Thus the set $V(M)$ is a vertex cover of H. □

The distribution of random graphs we consider in this paper is a k-partite version of the Erdős-Rényi distribution. For a fixed set V of n vertices and a real number $0 \leqslant p \leqslant 1$, the Erdős-Rényi distribution $\mathcal{G}(n,p)$ is the distribution of random graphs on vertex set V where every potential edge $e = \{u,v\}$, for vertices $u \neq v$, is sampled independently with probability p. As was done in [BIS07, ABdR+21], we work with the block model, which is defined as follows. Given k blocks $V_1,\dots,V_k$ of size n and a real number $0 \leqslant p \leqslant 1$, we denote by $\mathcal{G}(n,k,p)$ the distribution over graphs on the vertex set $V_{[k]}$ defined by sampling each edge $e = \{u,v\}$ independently with probability p if u and v are in distinct blocks. We sometimes refer to such pairs of vertices $\{u,v\}$ as *potential edges*. Edges within the same block are never included and hence $\mathcal{G}(n,k,p)$ is a distribution over k-partite graphs.

2.3 Clique Formula

Below we present an encoding of the k-clique formula on k-partite graphs as a system of polynomial equations.

Given a k-partite graph G with blocks $V_1,\dots,V_k$ of size n we define the *k-clique formula over* G stating that G has a k-clique (with one vertex per block) as follows. The formula is defined over $2kn$ variables: each vertex $v \in V_{[k]}$ is associated with two variables x_v and $\bar{x}_v$. The intended meaning is that x_v is 1 if v is in the identified k-clique and 0 otherwise, and $\bar{x}_v = 1 - x_v$. Since we define the Sherali-Adams proof system (see Section 2.1) to operate modulo the ideal generated by the Boolean axioms $y(1-y)$ and the negation axioms $1 - \bar{x}_v - x_v$, we do not need to explicitly include such axioms in the formula.

For each block V_i we introduce the *block axiom* $\sum_{v\in V_i} x_v - 1 = 0$ stating that precisely one vertex from each block is chosen and for each pair of vertices $\{u,v\} \notin E(G)$ in distinct blocks we introduce the *edge axiom* $x_u x_v = 0$ that ensures that non-neighbors are not simultaneously selected. We note that we could also include edge axioms for pairs of vertices in the same block but we choose not to since these are easily derivable from the block axioms.

It should be evident that this formula is satisfiable modulo the Boolean and the negation axioms if and only if there is a k-tuple t such that the vertex induced subgraph $G[t]$ is a clique.

We make two remarks about this choice of encoding. The first is that our lower bound strategy is completely agnostic to the encoding of the block axioms. We could equally well have considered the polynomial translation of the CNF encoding, or the binary encoding, or any encoding of the constraints that one vertex of each block should be chosen to be part of the clique.

The second remark is that, as argued in Beame, Impagliazzo and Sabharwal [BIS07], the reason for choosing to define the formula over k-partite graphs is that proving a lower bound for this encoding implies a lower bound for other natural encodings. In particular, if we consider the k-clique formula defined over a (not necessarily k-partite) graph $G = (V, E)$ on kn vertices, as the formula containing the axiom $\sum_{v \in V} x_v = k$ and edge axioms $x_u x_v = 0$ for each pair of vertices $\{u, v\} \notin E$, then for any equal-sized k-partition $V = V_1 \dot\cup \cdots \dot\cup V_k$ it holds that a lower bound for the k-clique formula on G with this partition implies the same lower bound for the non-k-partite formula.

Proposition 2.6 ([BIS07]). Let $k, n \in \mathbb{N}^+$ be integer and let G be a graph on kn vertices. Then the minimum Sherali-Adams coefficient size to refute the k-clique formula over G is bounded from below by the coefficient size required to refute the k-clique formula defined with respect to any equal-sized k-partition of G.

This proposition was proven in [BIS07] for resolution size, and it is straightforward to see that it holds for Sherali-Adams coefficient size. Indeed, it is enough to observe that the non-k-partite k-clique formula can be easily derived from the k-partite k-clique formula.

3 Main Theorem and Proof Overview

The main result of this paper is a tight, up to constants in the exponent, Sherali-Adams coefficient size lower bound for k-clique formulas over Erdős-Rényi random graphs.

Theorem 3.1 (Main theorem). Let k and D be functions of n such that $D \leqslant 2 \log n$ and $k \leqslant n^{1/66}$. If $G \sim \mathcal{G}(n, k, n^{-2/D})$, then asymptotically almost surely Sherali-Adams requires coefficient size $n^{\Omega(D)}$ to refute the k-clique formula over G.

Note that Theorem 1.1 follows directly from Theorem 3.1 along with Propositions 2.2 and 2.6, where we assume that the coefficient size is bounded by a polynomial n^c where c is a constant independent of D. The main result of the conference version of this paper [DRPR23] follows from Theorem 3.1 and Propositions 2.3 and 2.6.

In the rest of this section we outline our proof strategy. We intend to come up with a so-called *pseudo-measure* which lower bounds the coefficient size of a Sherali-Adams refutation. Before we get ahead of ourselves let us define what a pseudo-measure is.

Definition 3.2 (Pseudo-measure). Let $\delta > 0$ and $\mathcal{P}$ be a set of polynomials over the polynomial ring $\mathbb{R}[x_1, \ldots, x_n, \bar{x}_1, \ldots, \bar{x}_n]$. A linear function $\mu\colon \mathbb{R}[x_1, \ldots, x_n, \bar{x}_1, \ldots, \bar{x}_n] \to \mathbb{R}$, mapping polynomials to reals, is a *δ-pseudo-measure for $\mathcal{P}$* if for all monomials m and all polynomials $p \in \mathcal{P}$ it holds that

1. $\mu(1) = 1$,
2. $|\mu(m \cdot p)| \leqslant \delta$, and
3. $\mu(m) \geqslant -\delta$.

A concept related to the notion of a pseudo-measure has previously appeared in [PZ22] for the Nullstellensatz proof system over the reals. We also note that in a recent work Hubáček, Khaniki and Thapen [HKT24] defined a similar notion for Sherali-Adams with bounded degree. We have the following simple proposition.

Proposition 3.3. There is a δ-pseudo-measure for $\mathcal{P}$ if and only if any Sherali-Adams refutation of $\mathcal{P}$ requires coefficient size $1/\delta$.

Proof. Given a monomial m, let $\mathbb{1}_m$ be the vector of length 2^n, indexed by assignments $\rho \in \{0,1\}^n$, that corresponds to the truth table of m, that is, on index ρ the vector $\mathbb{1}_m$ is 1 if the monomial m evaluates to 1 on assignment ρ and 0 otherwise. We denote the all-ones vector $\mathbb{1}_1$ by $\mathbb{1}$.

Let $\mathcal{P}$ be a polynomial system of equations over Boolean variables $x_1, \dots, x_n$ and their twin variables $\bar{x}_1, \dots, \bar{x}_n$. We denote the coefficient of a monomial m in a polynomial $p \in \mathcal{P}$ by $\gamma_p(m)$. We may write a linear program that searches for a Sherali-Adams refutation of $\mathcal{P}$ of minimum coefficient size as

$$\begin{aligned} \text{minimize} \quad & \sum_m \alpha(m) + \sum_{p\in\mathcal{P}} \sum_m \left(\beta_p^+(m) + \beta_p^-(m)\right) \\ \text{subject to} \quad & \sum_{p\in\mathbb{P}} \sum_{m_1,m_2} \left(\beta_p^+(m_1) - \beta_p^-(m_1)\right) \cdot \gamma_p(m_2) \cdot \mathbb{1}_{m_1\cdot m_2} - \sum_m \alpha(m) \cdot \mathbb{1}_m = \mathbb{1} \\ & \alpha, \beta^+, \beta^- \geqslant 0 \ . \end{aligned} \tag{3.1}$$

We appeal to duality to obtain the following dual program to the above. Denote by μ the vector of dual variables of dimension 2^n and introduce for a monomial m the shorthand $\mu(m) = \mu^T \mathbb{1}_m$ and for a polynomial $q = \sum_m \gamma(m) \cdot m$ let $\mu(q) = \sum_m \gamma(m) \cdot \mu(m)$. The dual to the above linear program may be expressed as

$$\begin{aligned} \text{maximize} \quad & \mu(1) \\ \text{subject to} \quad & \mu(m) \geqslant -1 \qquad \forall m \\ & |\mu(m \cdot p)| \leqslant 1 \quad \forall m, \forall p \in \mathcal{P} \ . \end{aligned} \tag{3.2}$$

Thus the notion of a pseudo-measure for $\mathcal{P}$ is indeed the dual object of a Sherali-Adams refutation of $\mathcal{P}$ with small coefficient size. By appealing to strong duality and normalizing by $\mu(1)$ the claim follows. $\square$

3.1 Our Pseudo-Measure

In what follows we define our pseudo-measure μ for the k-clique formula. We may think of μ as a progress measure: it assigns to each monomial a real value which can be thought of as the contribution of this monomial towards the refutation of the k-clique formula. Thus, intuitively, we would like to associate each monomial with the fraction of potentially satisfying assignments that it rules out. In order to define this a bit more formally, let us introduce the set of potentially satisfying assignments.

We say that an assignment α is *potentially satisfying* for the k-clique formula if there is a graph G such that the k-clique formula defined over G is satisfied by α. This set of assignments can be easily characterized: if we associate each k-tuple t with the assignment ρ_t that sets all variables x_u to 1 if $u \in t$ and to 0 otherwise, then the set of potentially satisfying assignments of the k-clique formula is

$$\{\rho_t \mid t \in V_1 \times V_2 \times \cdots \times V_k\} \ . \tag{3.3}$$

We say that a monomial m *rules out* an assignment ρ if $\rho(m) = 1$. As there is a one-to-one correspondence between potentially satisfying assignments and tuples, it is convenient to think of the tuples that a monomial rules out. We thus associate each monomial m with the set

$$Q(m) = \{t \mid \rho_t(m) = 1\} \tag{3.4}$$

of ruled out k-tuples. Note that $Q(1)$ is the set of all tuples, that is, $Q(1) = V_1 \times V_2 \times \cdots \times V_k$ and the set $Q(x_u x_v)$ associated with an edge axiom $x_u x_v$ consists of all k-tuples that contain the vertices u and v.

More generally, it is not too hard to see that the set of ruled out tuples of a monomial is a rectangle and that for each rectangle Q there is at least one monomial m such that Q is the set of tuples ruled out by m. We thus often discuss rectangles and it is implicitly understood that if a statement holds for all rectangles, then it also holds for all monomials. Finally, observe that if a monomial m satisfies $m = m_1 \cdot m_2$, then $Q(m) \subseteq Q(m_1)$.

For intuition we will now discuss two naïve, and fatally flawed, attempts to define a pseudo-measure. For our first attempt, we simply associate each monomial with the fraction of ruled out tuples, that is we map a monomial m to

$$\frac{|Q(m)|}{|Q(1)|} = n^{-k} \cdot |Q(m)| \ . \tag{3.5}$$

This measure maps the monomial 1 to 1 and is clearly non-negative and hence satisfies Properties 1 and 3 of Definition 3.2 for any $\delta > 0$. Furthermore, again for any $\delta > 0$, it satisfies Property 2 of Definition 3.2 for the block axioms. Only the edge axioms cause trouble: the rectangle $Q(x_u x_v)$ associated with the edge axiom $x_u x_v$ is a n^{-2} fraction of all tuples. As such, according to Proposition 3.3, this pseudo-measure may only gives us an n^2 coefficient size lower bound—not quite what we are after.

We may try to remedy this by not associating a monomial m with *all* tuples in $Q(m)$ but rather only with a subset of $Q(m)$ that depends on the graph G. One very naïve attempt would be to associate m with the number of k-cliques that it rules out, that is, we may associate a monomial m with the normalized measure

$$\tilde{\mu}(m) = n^{-k} \sum_{t \in Q(m)} 2^{\binom{k}{2}} \mathbb{1}_{\{t \text{ is a clique}\}}(G) \ . \tag{3.6}$$

This definition, at least *in expectation* over $G \sim \mathcal{G}(n, k, 1/2)$, satisfies all properties of a pseudo-measure: the monomial 1 is mapped to 1, the axioms are all mapped to 0 and the measure is non-negative.

The obvious problem is that all graphs we consider do *not* contain a k-clique and hence everything (including the monomial 1) is mapped to 0. Put differently, the main problem is that the random variable $\tilde{\mu}(1)$ over $G \sim \mathcal{G}(n, k, 1/2)$ has too large variance. We reduce this variance as pioneered by Barak et al. [BHK+19]: we expand Equation (3.6) in the Fourier basis and truncate the resulting expression. A careful choice of the truncation, along with some significant effort, allows us to argue that, on the one hand, the measure associated with 1 is large (i.e., the variance is reduced) while, on the other hand, the measure associated with edge axioms

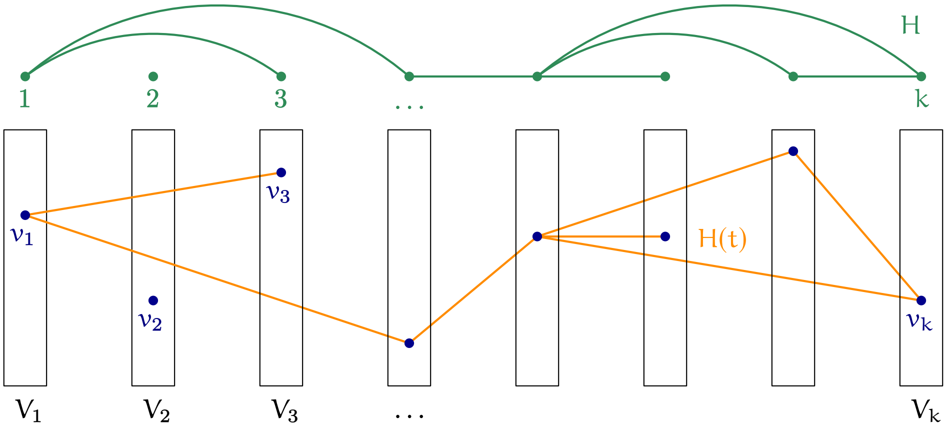

Figure 1: A pattern graph H mapped onto a tuple $t = (v_1, \dots, v_k)$

is still tightly concentrated around 0 and that the measure is essentially non-negative (i.e., the variance did not increase significantly). This measure thus constitutes a valid pseudo-measure as defined in Definition 3.2 up to normalization. In order to state the precise definition of our pseudo-measure μ we need some notation.

If p denotes the probability that a potential edge e is present in the graph, then the character $\chi_e(G)$ is defined by

$$\chi_e(G) = \begin{cases} \frac{1-p}{p} & \text{if } e \in E(G) \\ -1 & \text{otherwise;} \end{cases} \tag{3.7}$$

and for a set of potential edges E we let $\chi_E(G) = \prod_{e \in E} \chi_e(G)$. It is convenient for us to work with the above (non-standard) basis as it allows us to easily cancel characters in case an edge is missing. Observe that for $p = 1/2$ this is the usual ± 1 Fourier basis. First time readers are advised to keep this case in mind for the remainder of the article. Let us record some useful facts.

Proposition 3.4. Let $k, \ell, n \in \mathbb{N}^+$, let $p > 0$ and $G \sim \mathcal{G}(n, k, p)$. For any potential edge e, sampled with probability p, it holds that $\mathbb{E}_G[\chi_e(G)] = 0$ and $\mathbb{E}_G[\chi_e^{2\ell}(G)] = (1-p)(1 + (\frac{1-p}{p})^{2\ell-1})$. In particular, for $\ell = 1$, we have $\mathbb{E}_G[\chi_e^2(G)] = (1-p)/p$. A useful bound is $\mathbb{E}_G[\chi_e^{2\ell}(G)] \leqslant p^{-2\ell}$. Finally, observe that for any tuple t we have that $\sum_{E \subseteq \binom{t}{2}} \chi_E(G)$ is $p^{-\binom{k}{2}}$ if t is a clique and 0 otherwise.

To concisely state our pseudo-measure we need some further notation. We consider sums of tuples and want to treat edge sets that are equal up to the mapping onto a k-tuple as the same. More precisely, if we have two k-tuples $t = (v_1, \dots, v_k), t' = (v'_1, \dots, v'_k)$ and edge sets $E \subseteq \binom{t}{2}$ and $E' \subseteq \binom{t'}{2}$ such that $\{v_i, v_j\} \in E$ if and only if $\{v'_i, v'_j\} \in E'$, then we want to identify E and E' as the same edge set. To this end we consider *pattern graphs* H (similar to the *shape graphs* in the terminology of [BHK+19]) over the vertex set $[k]$. For a tuple $t = (v_1, \dots, v_k)$ and a pattern graph H we let $H(t)$ be the edge set that contains the edge $\{v_i, v_j\}$ if and only if the edge $\{i, j\}$ is present in H. See Figure 1 for an illustration.

With this notation at hand we define our pseudo-measure as

$$\mu(m) = \mu_d\big(Q(m)\big) = n^{-k} \sum_{t \in Q(m)} \sum_{\substack{H \\ vc(H) \leqslant d}} \chi_{H(t)}(G) \ , \tag{3.8}$$

where the second sum is over all graphs H over $[k]$ vertices with vertex cover at most d, and $d = \eta D$ is a small constant $\eta > 0$ times the maximum clique size of G.

Observe that Boolean axioms, the negation axioms and the block axioms multiplied by an arbitrary monomial are all mapped to 0 by μ. Hence it remains to prove that the measure μ maps the constant 1 monomial to a large value, that μ is small on subrectangles of edge axioms, i.e., any edge axiom multiplied by a monomial is mapped to a small value, and that all monomials are mapped to an approximately non-negative value.

By inspecting the second moment of $\mu(1)$ it is not too hard to see that there is quite a bit of freedom on how to choose the truncation in the definition of μ while maintaining the property that $\mu(1) = 1 \pm n^{-\Omega(1)}$ asymptotically almost surely. However, ensuring that the edge axioms are associated with small measure is more delicate. Here we heavily rely on our choice to truncate according to the minimum vertex cover. More specifically we rely on two crucial properties of graphs H satisfying $\mathrm{vc}(H) = d$: firstly, we use the fact that such graphs contain a matching of size $\lceil d/2 \rceil$ (see Proposition 2.5) and, secondly, that the family of these graphs satisfies a monotonicity property which leads to a useful partition of this family. For more details about this partition we refer to Section 4. Let us mention that it is conceivable that one could increase the bound on k for which our results hold by truncating according to the size of the maximum matching. As we do not know how to define the above mentioned partition with respect to the maximum matching we truncate according to the minimum vertex cover.

In the following sections we try to present some intuition as to why μ_d is a pseudo-measure up to normalization, that is, why it satisfies Definition 3.2 where Property 1 is relaxed to approximately 1. In Section 3.2 we verify that G sampled from $\mathcal{G}(n, k, 1/2)$ asymptotically almost surely satisfies $\mu(1) = \mu_d(\bigtimes_{i \in [k]} V_i) = 1 \pm n^{-\Omega(1)}$. As mentioned, this follows by a straightforward second moment argument. In Section 3.3 we outline why any subrectangle Q of an edge axiom satisfies $|\mu_d(Q)| \leqslant n^{-\Omega(d)}$. This proof motivates the definitions in Sections 4 and 5. Finally, in Section 3.4, we provide some high-level overview of how to prove that any rectangle Q is mapped to an approximately non-negative value, that is, it holds that $\mu_d(Q) \geqslant -n^{-\Omega(d)}$. This is the most technically challenging part of the paper.

3.2 Expected Behavior of Our Pseudo-Measure

The measure $\mu_d(Q)$ of any rectangle Q satisfies

$$\mathop{\mathbb{E}}_{G}[\mu_d(Q)] = n^{-k} \sum_{t \in Q} \sum_{H \in \mathcal{H}_d} \mathop{\mathbb{E}}_{G}[\chi_{H(t)}(G)] = n^{-k} \sum_{t \in Q} \mathop{\mathbb{E}}_{G}[\chi_{\emptyset(t)}(G)] = n^{-k}|Q| \ . \tag{3.9}$$

In particular, as $Q(1) = \bigtimes_{i \in [k]} V_i$, it holds that $\mathbb{E}_G[\mu(1)] = 1$. In what follows we show that, for $p = 1/2$, the measure is somewhat concentrated around the expected value. The concentration, though, is far from enough to perform a union bound over all rectangles to argue that the measure behaves as expected on all rectangles simultaneously.

We show that the measure concentrates by an application of Chebyshev's inequality. To this end we analyze the second moment: for $p = 1/2$ we have

$$\mathop{\mathbb{E}}_{G}[\mu_d^2(Q)] = n^{-2k} \cdot \sum_{H \in \mathcal{H}_d} \sum_{t, t' \in Q} \mathop{\mathbb{E}}_{G}[\chi_{H(t)}(G)\chi_{H(t')}(G)] \tag{3.10}$$

$$= n^{-2k} \cdot \sum_{H \in \mathcal{H}_d} \sum_{\substack{t, t' \in Q: \\ t_{V(E(H))} = t'_{V(E(H))}}} \mathop{\mathbb{E}}_{G}[\chi_{H(t)}(G)\chi_{H(t')}(G)] \tag{3.11}$$

$$= n^{-2k} \cdot \sum_{H \in \mathcal{H}_d} \left|\left\{(t,t') : t, t' \in Q \text{ and } t_{V(E(H))} = t'_{V(E(H))}\right\}\right| \tag{3.12}$$

$$= n^{-2k} \cdot \sum_{H \in \mathcal{H}_d} |Q_{V(E(H))}| \cdot |Q_{[k]\setminus V(E(H))}|^2 \tag{3.13}$$

$$\leqslant n^{-k}|Q| \cdot \left(1 + \sum_{\substack{H \in \mathcal{H}_d \\ H \neq \emptyset}} n^{-|V(E(H))|}\right) . \tag{3.14}$$

A careful application of Lemma 2.4 allows us to bound the number of pattern graphs H we sum over in (3.14) to conclude that $\mathbb{E}[\mu_d^2(Q)] = |Q|n^{-k}(1 \pm n^{-\Omega(1)})$, as long as k and d are small. By virtue of Chebyshev's inequality we then conclude that $\mu(1) = 1 \pm n^{-\Omega(1)}$ asymptotically almost surely.

A natural attempt to prove that the measure is mostly non-negative is to analyze higher moments in the hope that these are closely concentrated around the (positive) expected value. The fundamental difficulty in analyzing the pseudo-measure μ_d is that we have to analyze exponentially many rectangles simultaneously. Since there is such a large number of rectangles, for each input graph G, there will be some rectangles where the value of μ_d differs considerably from the expected value.

For example, the measure on a rectangle Q with only a few vertices Q_i in some block V_i heavily depends on the behavior of the edges incident to the vertices in Q_i. Hence, if Q_i is small enough, we expect large deviations from the expected value. A slightly simplified, though more concrete, example of this phenomenon goes as follows: let $v_1 \in V_1$ and $v_2 \in V_2$, let Q be the rectangle that consists of all tuples that contain v_1 as well as v_2, and let H be the graph with the single edge $\{1, 2\}$. In this setting the sum $\sum_{t \in Q} \chi_{H(t)}(G)$ heavily depends on whether the edge $\{v_1, v_2\}$ is present in G: if the edge is present, then the sum is equal to $n^{k-2}\frac{1-p}{p}$ and, if the edge is not present, then it is equal to $-n^{k-2}$. This indicates that on some rectangles the measure heavily depends on a few edges and we can thus not hope to naïvely prove concentration of the measure over all rectangles.

This slightly simplified example can be generalized to show that for a fixed H there is always a small number of rectangles where the value contributed by H is much larger than expected. Part of the technical challenge of the proof is to identify these bad rectangles and to handle them separately.

3.3 Edge Axioms Should Have Small Measure

We now explain the main ideas for bounding the magnitude of the measure of edge axioms. Recall that all other axioms are mapped to 0 by μ and we are thus just left to show that the value of the edge axioms is closely concentrated around 0.

For every pair of vertices $\{u, v\} \notin E(G)$ in distinct blocks we have an edge axiom $p_{uv} = x_u x_v$ stating that at least one of x_u and x_v are set to 0. Let Q be a subrectangle of $Q(p_{uv})$. Note that for every such rectangle Q there is a monomial m such that $Q = Q(m \cdot p_{uv})$ and hence these are the correct rectangles to consider if we want to prove Property 2 of Definition 3.2. In other words, if we manage to show for all such Q that $|\mu_d(Q)| \leqslant n^{-\Omega(d)}$, then it follows that for all monomials m it holds that $|\mu_d(m \cdot p_{uv})| \leqslant n^{-\Omega(d)}$, as needed.

We first show that for a fixed pair of vertices $\{u, v\} \notin E(G)$, with good probability, all such subrectangles Q have small absolute measure. By a union bound over all missing edges we then conclude that all subrectangles Q of an edge axiom satisfy $|\mu_d(Q)| \leqslant n^{-\Omega(d)}$. Let us fix an edge $\{u, v\} \notin E(G)$.

If Q is empty, then there is nothing to prove as $\mu_d(Q)$ is trivially 0. Hence we may assume that Q is non-empty, that is, Q has at least one vertex per block and hence each tuple in Q contains both u and v. Let $i \neq j \in [k]$ such that $u \in V_i$ and $v \in V_j$. For $e = \{i, j\}$ we may write

$$\mu_d(Q) = n^{-k} \sum_{t \in Q} \sum_{H \in \mathcal{H}_d} \chi_{H(t)}(G) \tag{3.15}$$

$$= n^{-k} \sum_{t \in Q} \Big(\sum_{\substack{H \in \mathcal{H}_d \\ e \notin H}} \chi_{H(t)}(G) + \sum_{\substack{H \in \mathcal{H}_d \\ e \in H}} \chi_{H(t)}(G) \Big) \tag{3.16}$$

$$= n^{-k} \sum_{t \in Q} \sum_{\substack{H:\, vc(H)=d, \\ vc(H \cup \{e\})=d+1}} \chi_{H(t)}(G) \ , \tag{3.17}$$

where the last equality follows from the fact that every tuple $t \in Q$ contains u and v and thus, if $e \notin H$, then $\chi_{H(t)}(G) = -\chi_{H(t)}(G) \cdot \chi_{\{u,v\}}(G) = -\chi_{(H \cup \{e\})(t)}(G)$ as $\{u, v\} \notin E(G)$.

The naïve approach to bounding $|\mu_d(Q)|$ is to try to bound the magnitude of $\sum_{t \in Q} \chi_{H(t)}(G)$ for each H separately and to then multiply this bound by the number of graphs H we sum over. Recall from Lemma 2.4 that there are about 2^{dk} graphs with a minimum vertex cover of size d. As the magnitude of $\sum_{t \in Q} \chi_{H(t)}(G)$ typically has value $\Omega(n^{-d}|Q|)$, for large rectangles Q, even with the optimal bound $|\sum_{t \in Q} \chi_{H(t)}(G)| \leqslant O(n^{-d}|Q|)$, we can only show a bound of

$$|\mu_d(Q)| \leqslant n^{-k} \sum_{\substack{H:\, vc(H)=d, \\ vc(H \cup \{e\})=d+1}} O\big(|Q| n^{-d}\big) \leqslant O\big(2^{dk} n^{-d}\big) = O\big(\exp(d(k - \log n))\big) \ . \tag{3.18}$$

Note that for $k = \mathrm{poly}(n)$ much larger than both d and $\log n$ the bound is at best $\exp\big(O(k)\big)$. We require a bound of the form $|\mu_d(Q)| \leqslant n^{-\Omega(d)}$, which is much smaller than $\exp\big(O(k)\big)$.

Instead of bounding the magnitude of $\sum_{t \in Q} \chi_{H(t)}(G)$ for each H separately, we partition the relevant set of graphs into different families and proceed to bound the magnitude of $\sum_{H \in \mathcal{H}(F, E_F^\star)} \sum_{t \in Q} \chi_{H(t)}(G)$ for each such family $\mathcal{H}(F, E_F^\star)$. More precisely, we have families of graphs indexed by graphs F with at most $3d$ non-isolated vertices of the form

$$\mathcal{H}(F, E_F^\star) = \{H \mid E(H) = E(F) \cup E, \text{ where } E \subseteq E_F^\star\} \ , \tag{3.19}$$

that partition the set of graphs H satisfying $vc(H) = d$ and $vc(H \cup \{e\}) = d + 1$. Using these families we can bound the magnitude of $\mu_d(Q)$ by

$$|\mu_d(Q)| = n^{-k} \Big| \sum_{t \in Q} \sum_{\substack{H:\, vc(H)=d, \\ vc(H \cup \{e\})=d+1}} \chi_{H(t)}(G) \Big| \tag{3.20}$$

$$\leqslant n^{-k} \sum_{F} \Big| \sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star)} \chi_{H(t)}(G) \Big| \tag{3.21}$$

$$= n^{-k} \sum_{F} \Big| \sum_{t \in Q} \chi_{F(t)}(G) \sum_{E \subseteq E_F^\star} \chi_{E(t)}(G) \Big| \ . \tag{3.22}$$

Observe that the innermost sum is, up to normalization, the indicator function of whether the edge set $E_F^\star(t)$ is present in G. In fact the innermost sum, with the appropriate definition of $E_F^\star$, is simply a statement about the common neighborhood sizes of different subsets of t in G.

We will need to argue that for random graphs, with high probability, all such sets behave as expected and the innermost sums are therefore bounded.

Furthermore, since each graph F has at most 3d with incident edges, there are fewer such graphs: according to Lemma 2.4 at most $2^{3d(d+\log k)}$. Since $k \leqslant n^{1/66}$ and $d \leqslant 2\eta \log n$, for some small constant η, it holds that there are at most $2^{d(d+\log k)} \lesssim n^{d/50}$ many such graphs F. Thus, an upper bound of $n^{k-\Omega(d)}$ on the absolute value of two innermost sums in Equation (3.22) can now be used to obtain the claimed bound $|\mu_d(Q)| \leqslant n^{-\Omega(d)}$. This completes the proof sketch for bounding the measure on edge axioms.

In Section 4 we formally define these *core* graphs F and the families $\mathcal{H}(F, E_F^\star)$. In Section 5 we introduce the pseudorandomness property of graphs we rely on in order to bound the two innermost sums in Equation (3.22). In Section 6.1 we formally prove that the measure on subrectangles of axioms is bounded in absolute value and lastly, in Section 8, we verify that random graphs indeed satisfy the necessary pseudorandomness properties.

3.4 Rectangles Should Be Approximately Non-Negative

To show that all rectangles Q have essentially non-negative measure, the main idea is to decompose Q into a collection $\mathcal{Q}$ of rectangles satisfying the following properties.

1. The collection $\mathcal{Q}$ is small, that is, $|\mathcal{Q}| \leqslant n^{O(d)}$.
2. Each rectangle $Q \in \mathcal{Q}$ is either
   (a) very small: $|Q| \leqslant n^{(1-\varepsilon)k}$ and hence $|\mu_d(Q)|$ is negligible,
   (b) a subrectangle of an axiom and thus, as argued in Section 3.3, $|\mu_d(Q)|$ is bounded, or
   (c) all common neighborhoods in Q are of expected size and therefore

$$\mu_d(Q) \approx |Q|/n^k > 0 \ . \tag{3.23}$$

In other words, $\mathcal{Q}$ contains some rectangles that have negligible measure and a collection of larger rectangles on which the measure behaves as expected. As the latter rectangles have strictly positive measure we may conclude that our pseudo-measure is essentially non-negative on all rectangles.

We bound the measure on small rectangles by summing the maximum possible magnitude of any character appearing in the definition of our pseudo measure.

Lemma 3.5. Any rectangle Q satisfies $|\mu_d(Q)| \leqslant O\big(|Q|n^{-k}k^d p^{-dk}\big)$.

Proof. We bound $\mu_d(Q)$ by counting the number of pattern graphs H we sum over multiplied by the maximum magnitude of each such character. We have that

$$|\mu_d(Q)| \leqslant n^{-k} \sum_{i=0}^{d} \sum_{j=i}^{ik} \Bigg| \sum_{t \in Q} \sum_{\substack{H:\\ vc(H)=i\\ |E(H)|=j}} \chi_{H(t)}(G) \Bigg| \tag{3.24}$$

$$\leqslant |Q| \cdot n^{-k} \sum_{i=0}^{d} \binom{k}{i} \sum_{j=0}^{ik} \binom{ik}{j} \left(\frac{1-p}{p}\right)^j \tag{3.25}$$

$$= |Q| \cdot n^{-k} \sum_{i=0}^{d} \binom{k}{i} p^{-ik} \leqslant O\big(|Q|n^{-k}k^d p^{-dk}\big) \ , \tag{3.26}$$

as claimed. □

We implement the above proof outline in Section 6.2. Proving that our pseudo-measure concentrates around a positive value on rectangles as described in Item 2c is the most delicate part of our proof. In fact, the above proof outline is somewhat inaccurate in that the value the pseudo-expectation concentrates around is not simply $|Q|/n^k$ but further depends on the number of small blocks in the rectangle Q. We refer to Definition 6.6 for the precise definition of these rectangles and to Lemma 6.7 for the claimed concentration inequality. Section 7 is dedicated to the proof of Lemma 6.7.

4 Cores

In this section we introduce the notion of a core of a pattern graph, which will be used extensively throughout the rest of the paper. Our notion of a core seems to be loosely connected to the notion of a *vertex cover kernel* as used in parameterized complexity (see, e.g., the survey by Fellows et al. [FJK+18]).

4.1 Cores and Boundaries

Recall that when bounding the measure of subrectangles of axioms A_e, we were left with sums over graphs H such that $\mathrm{vc}(H) = d$ and $\mathrm{vc}(H \cup \{e\}) = d + 1$ (see Equation (3.17)). Such graphs motivate the following definition of sets of graphs in the *boundary* of an edge.

Definition 4.1 (Boundary). Let $i \in \mathbb{N}$, H be a graph and $e \in \binom{V(H)}{2}$ be an edge. The graph H is in the (i, e)-*boundary*, denoted by $\mathcal{H}_i(e)$, if and only if $\mathrm{vc}(H) = i$ and $\mathrm{vc}(H \cup \{e\}) = i + 1$. Furthermore, we say that H is in the e-*boundary* if and only if H is in an (i, e)-boundary for some $i \in \mathbb{N}$.

As mentioned in the proof sketch bounding the edge axioms, we cannot bound each H in the e-boundary separately (there are too many pattern graphs H) so we partition such graphs according to *cores* as explained below.

Definition 4.2 (Core). A vertex induced subgraph F of H is a *core* if any minimum vertex cover of F is also a vertex cover of H.

Ultimately we are interested in cores that are induced by small vertex sets. It turns out that, in general, we cannot hope for cores of a graph H that are induced by fewer than $3 \cdot \mathrm{vc}(H)$ many vertices: as the graph H that consists of $\mathrm{vc}(H)$ vertex disjoint paths of length 2 has only a single core $F = H$, the best we can hope for are cores of size $3 \cdot \mathrm{vc}(H)$.

The notions of cores and (i, e)-boundaries interact nicely in the following sense.

Proposition 4.3. A core of a graph H is in the (i, e)-boundary if and only if H is.

Proof. Let F be a core of H. We first argue that if a core F of the graph H is in the (i, e)-boundary, then so is H. Indeed, by definition it holds that $\mathrm{vc}(F) = \mathrm{vc}(H) = i$. Moreover, F being in the (i, e)-boundary implies that the minimum vertex cover of $F \cup \{e\}$ has size $i + 1$, and therefore the minimum vertex cover of $H \cup \{e\}$ must also be $i + 1$ since F is a subgraph of H.

It remains to argue that if H is in the (i, e)-boundary, then so is the core F. By definition of core, $\mathrm{vc}(F) = \mathrm{vc}(H) = i$. Suppose, for the sake of contradiction, that F is not in the (i, e)-boundary

Algorithm 1 Computes the core of H

procedure $\textsc{Core}(H)$
  $W \leftarrow$ lex first minimum vertex cover of H
  $U_1 \leftarrow$ lex first maximal set in $[k] \setminus W$ with matching $M_1 \subseteq H$ from U_1 to W of size $|U_1|$
  $U_2 \leftarrow$ lex first maximal set in $[k] \setminus (W \cup U_1)$ with matching $M_2 \subseteq H$ from U_2 to W of size $|U_2|$
  return $H[W \cup U_1 \cup U_2]$

and thus $\mathrm{vc}(F \cup \{e\}) = i$. Let W be a minimum-sized vertex cover of $F \cup \{e\}$. Since $|W| = i$, it holds that W is also a minimum-sized vertex cover of F and thus, by definition of core, W is also a vertex cover of H. But this contradicts the assumption that H is in the (i, e)-boundary since W also covers the edge e and hence is a vertex cover of size i of $H \cup \{e\}$. □

Recall that $\mathcal{H}$ is the set of graphs on k labeled vertices. We consider a map core from $\mathcal{H}$ to small cores that satisfies certain properties as described below.

Theorem 4.4. There is a map core that maps graphs $H \in \mathcal{H}$ to a core of H with the following properties. For every graph F in the image of core we have that $|V(E(F))| \leqslant 3 \cdot \mathrm{vc}(F)$ and that there exists an edge set $E_F^\star \subseteq V\big(E(F)\big) \times \big([k] \setminus V\big(E(F)\big)\big)$ such that $\mathrm{core}(H) = F$ if and only if $E(H) = E(F) \cup E$ for $E \subseteq E_F^\star$.

We prove Theorem 4.4 in Section 4.2 below. From now on we only consider the cores given by the map core as in Theorem 4.4. With a slight abuse of nomenclature we say that $\mathrm{core}(H)$ is *the core* of H. Note that for a graph F in the image of core we have that $\mathrm{core}^{-1}(F) = \mathcal{H}(F, E_F^\star) = \{H \mid E(H) = E(F) \cup E, \text{ for } E \subseteq E_F^\star\}$, as introduced in Section 3.3.

4.2 Proof of Theorem 4.4

We first explain how to construct a map core and then prove that it satisfies the required properties. In order to construct the mapping we require an order on subsets of vertices of H: consider every subset of vertices as a sequence of vertices sorted in ascending order and say that a set U is lexicographically smaller than a set V if the ascending sequence $(u_1, \ldots, u_s)$ of U is lexicographically smaller than the sequence $(v_1, \ldots, v_t)$ of V, that is, for $i = 1, \ldots, \min\{s, t\}$ compare u_i with v_i: if $u_i < v_i$ (respectively $u_i > v_i$), then U is lexicographically smaller (respectively larger) than V; if $u_i = v_i$ continue with $i = i + 1$; if the prefix of length $\min\{s, t\}$ is equal, then U is lexicographically smaller (larger) than V if $s < t$ (respectively $s > t$).

Equivalently U is lexicographically smaller than V if the smallest element in the symmetric difference of U and V is contained in U. From this alternate definition we immediately obtain the following property of the order.

Fact 4.5. Let W and V be sets such that W is lexicographically smaller than V and $W \not\subset V$. Then, for any element w, the set $W \cup \{w\}$ is lexicographically smaller than V.

We extend this notion in the natural manner to vertex covers and say that a minimum vertex cover W is the *lexicographically first minimum vertex cover of* H if W is the lexicographically smallest minimum vertex cover of H.

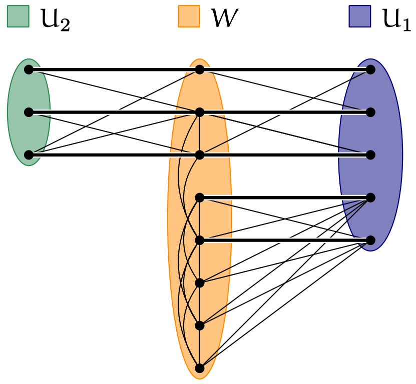

Figure 2: A core with edges in M_1 and M_2 highlighted

The map core is constructed as follows (see Algorithm 1 for an algorithmic description). Given a graph H with lexicographically first minimum vertex cover W define the following two sets U_1 and U_2 of vertices. Let $U_1 \subseteq [k] \setminus W$ be the lexicographically first maximal (with respect to set inclusion) set of vertices with a matching $M_1 \subseteq H$ from U_1 to W that covers all of U_1, that is, $U_1 \subseteq V(M_1)$ and $M_1 \subseteq U_1 \times W$. Similarly let $U_2 \subseteq [k] \setminus (W \cup U_1)$ be the lexicographically first maximal set of vertices with a matching M_2 from U_2 to W of size $|U_2|$. The core of H is defined to be $\mathrm{core}(H) = H[W \cup U_1 \cup U_2]$. An illustration is provided in Figure 2.

Let us record some simple observations

Claim 4.6. For H, W, U_1, U_2 and matchings M_1, M_2 defined as above, it holds that

1. $W \cap V(M_1) \supseteq W \cap V(M_2)$,
2. any edge $e \in H \setminus H[W \cup U_1]$ is incident to $W \cap V(M_1)$, and
3. any edge $e \in H \setminus H[W \cup U_1 \cup U_2]$ is incident to $W \cap V(M_2)$.

Proof. Item 1 follows by maximality of U_1. For Item 2, let $e = \{v, w\} \in H \setminus H[W \cup U_1]$. As W is a vertex cover of H we may assume that $w \in W$ and hence, as $e \notin H[W \cup U_1]$, the vertex v is not in $W \cup U_1$. Hence by maximality of U_1 it holds that $w \in W \cap V(M_1)$. The same argument establishes Item 3. ◻

Since $|U_2| \leqslant |U_1| \leqslant |W|$, it follows that $\left|V\big(E(F)\big)\right| \leqslant 3\,|W| \leqslant 3\,\mathrm{vc}(H)$. To argue that this map satisfies the other required properties, we first prove that the size of the minimum vertex cover of $H[W \cup U_1]$ is the same as that of H. Clearly $\mathrm{vc}(H) \geqslant \mathrm{vc}\big(H[W \cup U_1]\big)$ so it remains to prove the opposite inequality.

Lemma 4.7. For H, W and U_1 defined as above, it holds that $\mathrm{vc}(H) \leqslant \mathrm{vc}\big(H[W \cup U_1]\big)$.

Proof. Let M_1 be the matching from U_1 to W that covers U_1 and let R be the vertices in W that are not matched by M_1, that is, $R = W \setminus V(M_1)$. Towards contradiction suppose that the graph $H[W \cup U_1]$ has a vertex cover of size $\mathrm{vc}(H) - 1$.

By Claim 4.6, Item 2, the only edges in H which do not appear in the graph $H[W \cup U_1]$ are edges between a vertex in $W \cap V(M_1)$ and a vertex outside of $W \cup U_1$.

Let W' be a vertex cover of $H[W \cup U_1]$ of size $\mathrm{vc}(H) - 1$ that maximizes $|W' \cap V(M_1)|$. If W' is a vertex cover of H we have reached a contradiction. Otherwise, there exists a vertex $u \in W \cap V(M_1)$ that is not contained in W' and an edge $e_0 = \{u, w\} \in E(H)$ for $w \in [k] \setminus (W \cup U_1)$.

We now want to argue that either we can construct a vertex cover of $H[W \cup U_1]$ of size $vc(H) - 1$ that contradicts the fact that W' maximized $|W' \cap V(M_1)|$, or we can construct a strict superset of U_1 with a matching to W, contradicting the maximality of U_1.

We iteratively construct two sets, $A \subseteq W \setminus W'$ and $B \subseteq W' \cap U_1$ as follows. We start by including u in A. We then iteratively include in B all vertices v in U_1 that are matched by M_1 to some vertex in A, that is, vertices v such that $\{v, v'\} \in M_1$ for some $v' \in A$; and include in A all vertices of W not covered by $W' \setminus B$. Note that $|A| \geqslant |B|$ and that we keep the invariant that $B \subseteq W'$ since the edges in M_1 must be covered by some vertex. We consider two cases.

Case $|A| = |B|$: In this case, $(W' \cup A) \setminus B$ is a vertex cover of $H[W \cup U_1]$ of size $vc(H) - 1$ contradicting the fact that W' maximized $|W' \cap V(M_1)|$.

Case $|A| > |B|$: This can only happen if there is a vertex v in $R \cap A$. In this case, we can define an augmenting path from v to w, alternating between edges not in M_1 and edges in M_1. This implies we can define a matching that matches all of U_1 to vertices in W as well as the vertex w. This is in contradiction with the maximality of U_1. □

To prove that $H[W \cup U_1 \cup U_2]$ is a core of H we must show that any minimum vertex cover of $H[W \cup U_1 \cup U_2]$ is a vertex cover of H.

Corollary 4.8. Any minimum vertex cover of $H[W \cup U_1 \cup U_2]$ is also a vertex cover of H.

Proof. By Lemma 4.7 we have that

$$vc(H) = vc\big(H[W \cup U_1]\big) = vc\big(H[W \cup U_1 \cup U_2]\big) \ . \tag{4.1}$$

Therefore, any minimum vertex cover W' of $H[W \cup U_1 \cup U_2]$ has no vertex in U_2 (otherwise, $W' \setminus U_2$ would be a smaller vertex cover of $H[W \cup U_1]$). As W' needs to cover the edges of the matching M_2 from U_2 to W, the vertex cover W' contains all vertices in $V(M_2) \cap W$.

According to Claim 4.6, Item 3, only vertices in $V(M_2) \cap W$ may have edges incident in H that are not present in $H[W \cup U_1 \cup U_2]$. Hence W' is a vertex cover of H; the statement follows. □

It remains to argue that for every $F \in \mathrm{img}(\mathrm{core})$ there is a set $E_F^\star$ such that $\mathrm{core}(H) = F$ if and only if $H = F \cup E$ for $E \subseteq E_F^\star$. Let $E_F^\star = \bigcup_{H \in \mathrm{core}^{-1}(F)} H \setminus F$. By definition, if $\mathrm{core}(H) = F$, then there is an $E \subseteq E_F^\star$ such that $H = F \cup E$. We establish the reverse direction by the following two claims.

In Appendix A we provide an alternate proof that characterizes the set $E_F^\star$ explicitly.

Claim 4.9. Let $\mathrm{core}(H) = F$ and let W, U_1, U_2 (respectively W', U_1', U_2') denote the sets as in Algorithm 1 when run on H (respectively on F). It holds that $W = W'$, $U_1 = U_1'$ and $U_2 = U_2'$.

Proof. By definition of the map core given by Algorithm 1, we have that $F = H[W \cup U_1 \cup U_2]$. Corollary 4.8 implies that W is the lexicographically first minimum vertex cover of F (otherwise W would not be the lexicographically first minimum vertex cover of H) and hence $W = W'$.

Clearly any matching M_1 in H from W to U_1 is also in $F = H[W \cup U_1 \cup U_2]$, and any matching M_1' in F from W to U_1' is also in $H \supseteq F$, and thus it must hold $U_1 = U_1'$. Similarly, any matching M_2 in $H[[k] \setminus U_1]$ from W to U_2 is also in $F[[k] \setminus U_1] = H[W \cup U_2]$, and any matching M_2' in $F[[k] \setminus U_1]$ from W to U_2' is also in $H[[k] \setminus U_1] \supseteq F[[k] \setminus U_1]$, and thus it must hold $U_2 = U_2'$. □

Note that Claim 4.9 implies that $\mathrm{core}(F) = F$ and that the sets W, U_1 and U_2 as in Algorithm 1 run on any two graphs H and H' such that $\mathrm{core}(H) = \mathrm{core}(H')$ are identical.

Lemma 4.10. For $F \in \mathrm{img}(\mathrm{core})$ and for all $E \subseteq E^\star_F$ it holds that $\mathrm{core}(F \cup E) = F$.

Proof. Let E', E'' be subsets of $E^\star_F$ such that $\mathrm{core}(F \cup E') = \mathrm{core}(F \cup E'') = F$. It suffices to show that if $E \subseteq E' \cup E''$, then it holds that $\mathrm{core}(F \cup E) = F$. Let W, U_1, U_2 (W', U'_1, U'_2) be the sets as in Algorithm 1 when run on $F \cup E$ (respectively on $F \cup E'$). By Claim 4.9 the sets W', U'_1 and U'_2 could be equivalently defined as the sets from Algorithm 1 when run on F or $F \cup E''$.

We first show that $W = W'$. By Claim 4.9 we have that W' is the lexicographically first minimum vertex cover of F. It suffices to argue that W' is a vertex cover of $F \cup E$ since any vertex cover of $F \cup E$ is a vertex cover of F. But this is easy to see since W' is a vertex cover of $F \cup E'$ and of $F \cup E''$ and thus it is a vertex cover of $F \cup E' \cup E''$, and hence also of $F \cup E \subseteq F \cup E' \cup E''$.

Suppose for the sake of contradiction that $U_1 \neq U'_1$ and let u be the smallest vertex in the symmetric difference of U_1 and U'_1. Note that any matching M'_1 in $F \cup E'$ from W to U'_1 is also in F. Hence either $U'_1 \subseteq U_1$ (since U_1 is a maximal set with a matching $M_1 \subseteq F \cup E$ into W) or U'_1 is lexicographically larger than U_1 (as U_1 is the lexicographically first such set). In both cases it holds that $u \in U_1$.

Let $U = (U_1 \cap U'_1) \cup \{u\}$ and observe that since $u \notin U'_1$, it holds that either $U'_1 \subsetneq U$ or U'_1 is lexicographically larger than U. Let M be a matching in $F \cup E$ from U to W that covers U (which exists as $U \subseteq U_1$) and denote by e the edge in M adjacent to u. Since $e \in F \cup E \subseteq F \cup E' \cup E''$ it must be the case that $e \in F \cup E'$ or $e \in F \cup E''$. Without loss of generality suppose that $e \in F \cup E'$ and thus $M \subseteq F \cup E'$. Note that this contradicts the choice of U'_1: if $U'_1 \subsetneq U$, then U'_1 is not maximal and, if U'_1 is lexicographically larger than U, then by Fact 4.5, it is not the lexicographically first maximal set that can be matched to W. We conclude that $U_1 = U'_1$.

A similar argument can be used to show that $U_2 = U'_2$ and thus $\mathrm{core}(F \cup E) = F$ as claimed. □

5 Well-Behaved Graphs

In this section, we define the notion of *well-behaved* graphs, which is based on two combinatorial properties of graphs related to common neighborhoods of small tuples, and two analytic properties that bound certain character sums. In Section 8 we then show that random graphs are asymptotically almost surely well-behaved. In the following sections we prove that our measure satisfies the required conditions to obtain our Sherali-Adams coefficient size lower bound for any well-behaved graph.

Let us start by introducing the concepts needed to define well-behaved graphs. We say a rectangle Q is *s-small* if $|Q_i| \leqslant s$ for all $i \in [k]$ and, given a set $A \subseteq [k]$, a rectangle Q is said to be *(s, A)-large* if $|Q_i| > s$ for all $i \in A$. For any set $\mathcal{D}$ we say that a function $f : \mathcal{D} \to \mathbb{R}^+$ is *r-bounded* if $f(x) \leqslant r$ for all $x \in \mathcal{D}$.

We require some terminology and notation from graph theory. The neighborhood of a vertex $v \in V$ in a graph $G = (V, E)$ is $N(v) = N_G(v) = \{u \mid \{u, v\} \in E\}$ and the neighborhood of a set of vertices $U \subseteq V$ is $N(U) = N_G(U) = \{v \notin U \mid \exists u \in U : \{u, v\} \in E\}$. For a set $W \subseteq V$ the neighborhood of a vertex v in W is $N(v, W) = N(v) \cap W$ and similarly for a set U we let the neighborhood of U in W be $N(U, W) = N(U) \cap W$. The common neighborhood of U is $N^\cap(U) = \bigcap_{u \in U} N(u)$ and the common neighborhood of U in W is $N^\cap(U, W) = N^\cap(U) \cap W$. This notation is naturally extended to a tuple t by considering t as a set of vertices.

The next two definitions are purely combinatorial. They are similar to definitions that have appeared in previous papers on k-clique [BIS07, BGLR12, ABdR[+]21]. Recall that throughout the paper graphs denoted by G are k-partite with partitions $V_1, \ldots, V_k$ of size n each.

Definition 5.1 (Bounded common neighborhoods). A graph G has (β, p)-*bounded common neighborhoods from* $Q = \bigtimes_{i \in A} Q_i$ *to* $R \subseteq V(G)$ if it holds that for all $B \subseteq A$ and all $t \in Q_B$

$$|N^\cap(t, R)| \in (1 \pm \beta) p^{|t|} |R| \ .$$

A graph G has (β, p, d)-*bounded common neighborhoods in every block* if for all $A \subseteq [k]$ of size at most d and all $i \in [k] \setminus A$, G has (β, p)-bounded common neighborhoods from V_A to V_i.

While it turns out that random graphs do have bounded common neighborhoods, the graph induced by a rectangle may certainly have tuples with ill-behaved common neighborhoods: we may for example have an isolated vertex in a rectangle. The following definition roughly states that, while there may be tuples with ill-behaved neighborhoods in a rectangle, there is a large sub-rectangle which has bounded common neighborhoods.

Definition 5.2 (Bounded error sets). A graph G has (s, w, β, p, d)-*bounded error sets* if for all rectangles $Q = \bigtimes_{i \in [k]} Q_i$ satisfying $|Q_i| \geqslant s$ or $|Q_i| = 0$ it holds that there exists a small set of vertices $W \subseteq V(G)$, $|W| \leqslant w$, such that for all $S \subseteq [k]$ of size at most d it holds that all tuples $t \in \bigtimes_{i \in S}(Q_i \setminus W)$ satisfy

$$\left|N^\cap(t, Q_j \setminus W)\right| \in (1 \pm \beta) p^{|t|} \left|Q_j \setminus W\right|$$

for all $j \in [k] \setminus S$. We refer to W as the *error set of* Q.

Recall from the edge axiom proof sketch in Section 3.3 that we require bounds of the form $n^{k - \Omega(\mathrm{vc}(F))}$ on the absolute value of certain character sums. It turns out that, in order to prove that monomials are mapped to an essentially non-negative value, we need tighter (depending on $|Q|$) as well as "localized" versions of these bounds. For conciseness we introduce the following terminology.

Definition 5.3 (Bounded character sums). Let $s \in \mathbb{N}^+$, $B \subseteq [k]$, $Q_B = \bigtimes_{i \in B} Q_i$ and F be a core graph. A graph G has s-*bounded character sums over* Q_B *for* F if it holds that

$$\left| \sum_{t \in Q_B} \sum_{H \in \mathcal{H}(F, E_F^\star[B])} \chi_{H[B](t)}(G) \right| \leqslant s \ .$$

We are now ready to state the pseudorandomness property of graphs that allows us to prove average-case Sherali-Adams coefficient size lower bounds for the k-clique formula. As Properties 3 and 4 are somewhat difficult to parse we give an informal description upfront.

Property 3 states that all character sums over the families $\mathcal{H}(F, E_F^\star)$ are of bounded magnitude if the rectangle considered has large minimum block size. Smaller rectangles are unfortunately not as well-behaved. However, for certain rectangles, we can guarantee something similar: Property 4 states that if the common neighborhood of small tuples in a rectangle are bounded, then the mentioned character sums can still be bounded.

First time readers may, for now, choose to skip the formal definition of Property 4. It might be more insightful to first read Sections 6 and 7 and return to Property 4 once it is used.

Definition 5.4 (Well-behaved graphs). We say that a k-partite graph G with partitions of size n is D-*well-behaved* if, for $p = n^{-2/D}$, the following properties hold:

1. G has $(1/k, p, D/4)$-bounded common neighborhoods in every block.
2. There exists a constant $C \in \mathbb{R}^+$ such that G has $(2s, s, 1/k, p, \ell)$-bounded error sets for all $\ell \leqslant D/4$ and $s \geqslant Ck^4 \ell \ln n / p^{2\ell}$.
3. For any core F satisfying $vc(F) \leqslant D/4$, and any $\big(n/2, V\big(E(F)\big)\big)$-large rectangle Q it holds that G has s-bounded character sums over $Q_{[n]}$ for F, where
$$s = 6 \cdot p^{-|E(F)|} \cdot n^{k - \lambda\, vc(F)/4} ,$$
for any $\lambda < 1 - \log(k)/\log(n)$.
4. Let F be a core satisfying $vc(F) \leqslant D/4$, let $\Lambda \geqslant 20\, k \log n$, denote by $B \subseteq [k]$ a set of vertices and let $A = V\big(E(F)\big) \cap B$. Then for any rectangle Q which is (4Λ)-small and (Λ, B)-large the following holds. If G has $(3/k, p)$-bounded common neighborhoods from Q_A to Q_i for every $i \in B \setminus A$, then G has s-bounded character sums over Q_B for F, where
$$s = O\big(p^{-|E(F[B])|} \cdot (\Lambda/10\, k \log n)^{-\, vc(F[B])/4} \cdot |Q_B|\big) .$$

In what follows we often state that a graph G is D-well-behaved in which case it is implicitly understood that G is k-partite with partitions of size n. In Section 8 we prove that a graph G sampled from the distribution $\mathcal{G}(n, k, n^{-2/D})$ is asymptotically almost surely D-well-behaved.

Theorem 5.5. If n is a large enough integer, $k \in \mathbb{N}^+$ and $D \in \mathbb{R}^+$ satisfy $4 \leqslant D \leqslant 2 \log n$ and $k \leqslant n^{1/5}$, then $G \sim \mathcal{G}(n, k, n^{-2/D})$ is asymptotically almost surely D-well-behaved.

6 Clique Is Hard on Well-Behaved Graphs

In this section we prove that our measure μ_d is an $n^{-\Omega(D)}$-pseudo-measure for the k-clique formula, if the formula is defined over a D-well-behaved graph G.

Theorem 6.1. There are constants $\eta, c > 0$ and $D_0 \in \mathbb{N}$ such that the following holds for large enough $n \in \mathbb{N}$ and all D satisfying $D_0 < D \leqslant 2 \log n$. If $D \leqslant k \leqslant n^{1/66}$, $d = \eta D$ and G is a D-well-behaved k-partite graph with n vertices per block, then the normalized measure $\mu(m) = \mu_d(m)/\mu_d(1)$ is an n^{-cD}-pseudo-measure for the k-clique formula over G.

From Theorem 6.1 along with Theorem 5.5 and Proposition 3.3 we obtain Theorem 3.1.

In order to prove that the measure μ satisfies the properties of a pseudo-measure as listed in Definition 3.2, we show that μ_d maps any axiom multiplied by a monomial to approximately 0 and that all monomials are associated with an essentially non-negative value. Finally, we argue that $\mu_d(1) \geqslant 1 - n^{-\Omega(1)}$.

Recall that the clique formula consists of block axioms $\sum_{v \in V_i} x_v - 1 = 0$ for each block V_i and edge axioms $x_u x_v = 0$ for every non-edge in the graph. By linearity of μ_d over the tuples it holds for any monomial m that $\mu_d\big(m(\sum_{v \in V_i} x_v - 1)\big) = 0$. The lemma below implies that for edge axioms $p_{uv} = x_u x_v$ it holds that $|\mu_d(m p_{uv})| \leqslant n^{-cD}$, for any monomial m. As mentioned in Section 3.4, we also rely on this lemma to prove that the measure is essentially non-negative. Since that proof requires a careful choice of parameters we need to state the lemma with some precision.

Lemma 6.2. Let G be a D-well-behaved graph, let n be a large enough integer and let $d = \eta D \leqslant 2\eta \log n$ for some constant $\eta > 0$. It holds that all edge axioms p_{uv} and all rectangles $Q \subseteq Q(p_{uv})$ satisfy $|\mu_d(Q)| \leqslant O\big((n^{\lambda/4-12\eta}/(2k)^3)^{-d}\big)$ for any $\lambda < 1-\log(k)/\log(n)$.

Note that by choosing $\lambda = 1/2$, and considering $k \leqslant n^{1/66}$ and $\eta > 0$ small enough, Lemma 6.2 implies that any subrectangle of an edge axiom satisfies $|\mu_d(Q)| \leqslant n^{-cD}$ for some small enough constant c. We postpone the proof of Lemma 6.2 to Section 6.1.

In addition to the bound on the magnitude of the measure on the axioms we also need to prove that the measure is essentially non-negative. We state this formally below and defer the proof to Section 6.2.

Lemma 6.3. There are constants $\eta, c > 0$ such that if G is a D-well-behaved graph, n is large enough, $d = \eta D \leqslant 2\eta \log n$ and $D \leqslant k \leqslant n^{1/66}$, then any rectangle Q satisfies $\mu_d(Q) \geqslant -n^{-cD}$.

In Section 3.2 we argued that with high probability $\mu_d(1)$ is approximately 1 if G is a random graph and $p = 1/2$. We now show that this holds for any D-well-behaved graph.

Lemma 6.4. There are constants $\eta, c > 0$ such that for n large enough, $k \leqslant n^{1/20}$, $D \leqslant 2\log n$ and $d = \eta D$ it holds that if G is a D-well-behaved graph, then $\mu_d(1) \geqslant 1 - n^{-c}$.

Proof. This is a direct consequence of the definition of a D-well-behaved graph and Theorem 4.4. Recall the map core as defined in Theorem 4.4 and the families of graphs

$$\mathcal{H}(F, E^\star_F) = \{H \mid E(H) = E(F) \cup E,\ \text{for } E \subseteq E^\star_F\} \ , \tag{6.1}$$

defined for core graphs $F \in \mathrm{img(core)}$ such that $\mathrm{vc}(F) \leqslant d$. Recall that each such core graph F satisfies that $|V(E(F))| \leqslant 3\,\mathrm{vc}(F)$ and hence $|E(F)| \leqslant 3(\mathrm{vc}(F))^2 \leqslant 3d\,\mathrm{vc}(F)$.

By appealing to Property 3 of a D-well-behaved graph, that is, Property 3 of Definition 5.4, with $\lambda = 4/5$ we can conclude that for every $F \in \mathrm{img(core)}$ it holds that

$$n^{-k}\Big|\sum_{t\in Q(1)}\ \sum_{H\in\mathcal{H}(F,E^\star_F)} \chi_{H(t)}(G)\Big| \leqslant 6\,p^{-|E(F)|} n^{-\,\mathrm{vc}(F)/5} \leqslant n^{-\,\mathrm{vc}(F)/6} \ , \tag{6.2}$$

where we used the bound $p^{-|E(F)|} \leqslant p^{-3d\,\mathrm{vc}(F)} \leqslant n^{6\eta\,\mathrm{vc}(F)}$, which holds since $p = n^{-2/D}$ and $d = \eta D \leqslant 2\eta \log n$, and, furthermore, relied on the assumption that η is a small enough constant.

Recall that the families $\{\mathcal{H}(F, E^\star_F) \mid F \in \mathrm{img(core)}\}$ as defined in Equation (6.1) partition the set $\mathcal{H}_d$ of graphs of vertex cover at most d. We may hence write

$$\mu_d(1) = n^{-k}\sum_{H\in\mathcal{H}_d}\ \sum_{t\in Q(1)} \chi_{H(t)}(G) \tag{6.3}$$

$$= 1 + n^{-k}\sum_{\substack{H\in\mathcal{H}_d\\ H\neq\emptyset}}\ \sum_{t\in Q(1)} \chi_{H(t)}(G) \tag{6.4}$$

$$\geqslant 1 - n^{-k}\sum_{i=1}^{d}\ \sum_{\substack{F\in\mathrm{img(core)}\\ \mathrm{vc}(F)=i}} \Big|\sum_{t\in Q(1)}\ \sum_{H\in\mathcal{H}(F,E^\star_F)} \chi_{H(t)}(G)\Big| \ . \tag{6.5}$$

Since each graph $F \in \operatorname{img}(\text{core})$ satisfies that $|V(E(F))| \leqslant 3\,\text{vc}(F)$, by appealing to Lemma 2.4, we obtain the bound $|\{F \in \operatorname{img}(\text{core}) \mid \text{vc}(F) = i\}| \leqslant 2^{3i(d+\log k)}$. Combining this bound along with the bound from Equation (6.2) we may conclude that

$$\mu_d(1) \geqslant 1 - n^{-k} \sum_{i=1}^{d} \sum_{\substack{F\in \operatorname{img}(\text{core}) \\ \text{vc}(F)=i}} \Big| \sum_{t\in Q(1)} \sum_{H\in \mathcal{H}(F,E_F^\star)} \chi_{H(t)}(G) \Big| \tag{6.6}$$

$$\geqslant 1 - \sum_{i=1}^{d} 2^{3i(d+\log k)} n^{-i/6} \tag{6.7}$$

$$\geqslant 1 - n^{-c} \ , \tag{6.8}$$

for some small constant $c > 0$. The final inequality relies on the assumptions $d \leqslant 2\eta \log n$, that η is a small enough constant and that $k \leqslant n^{1/20}$. This concludes the proof. □

This completes the proof of Theorem 6.1 modulo Lemma 6.2 and Lemma 6.3. In Section 6.1 we prove Lemma 6.2 and the proof of Lemma 6.3 is provided in Section 6.2.

6.1 Axioms Have Small Measure

In this section we show that any subrectangle of an edge axiom has small measure in absolute value. We rely on the following technical lemma.

Lemma 6.5. If G is a D-well-behaved graph, then for any core graph F and any rectangle Q it holds that

$$\Big| \sum_{t\in Q} \sum_{H\in \mathcal{H}(F,E_F^\star)} \chi_{H(t)}(G) \Big| \leqslant 6 \cdot 2^{|A|} \cdot p^{-|E(F)|} \cdot n^{k-\lambda\,\text{vc}(F)/4} \ ,$$

where $A = V(E(F))$ and $\lambda < 1 - \log(k)/\log(n)$.

Proof. Let F be a core graph, let $A = V(E(F))$ and $s = 6 \cdot p^{-|E(F)|} \cdot n^{k-\lambda\,\text{vc}(F)/4}$. By Property 3 of Definition 5.4 we have that if Q is $(n/2, A)$-large , that is, if Q satisfies $|Q_i| \geqslant n/2$ for all $i \in A$, then $\big| \sum_{t\in Q} \sum_{H\in \mathcal{H}(F,E_F^\star)} \chi_{H(t)}(G) \big| \leqslant s$.

Given any rectangle Q (not necessarily $(n/2, A)$-large), let $T \subseteq A$ be the set of blocks of Q such that $|Q_i| < n/2$. By a simple inclusion-exclusion argument, we obtain that

$$Q = \sum_{S\subseteq T} (-1)^{|S|} \Big(\bigtimes_{i\in S} (V_i \setminus Q_i) \Big) \times \Big(\bigtimes_{i\in T\setminus S} V_i \Big) \times \Big(\bigtimes_{i\in [k]\setminus T} Q_i \Big) \ . \tag{6.9}$$

For $S \subseteq T$, denote by Q^S the rectangle $\big(\bigtimes_{i\in S} (V_i \setminus Q_i) \big) \times \big(\bigtimes_{i\in T\setminus S} V_i \big) \times \big(\bigtimes_{i\in[k]\setminus T} Q_i \big)$. Note that Q^S is $(n/2, A)$-large and therefore by Property 3 of Definition 5.4, G has s-bounded character sums over Q^S for F. This implies that

$$\Big| \sum_{t\in Q} \sum_{H\in \mathcal{H}(F,E_F^\star)} \chi_{H(t)}(G) \Big| \leqslant \sum_{S\subseteq T} \Big| \sum_{t\in Q^S} \sum_{H\in \mathcal{H}(F,E_F^\star)} \chi_{H(t)}(G) \Big| \leqslant 2^{|A|} \cdot s \ , \tag{6.10}$$

as claimed. □

We are now ready to prove Lemma 6.2 restated for convenience.

Lemma 6.2 (Restated). Let G be a D-well-behaved graph, let n be a large enough integer and let $d = \eta D \leqslant 2\eta \log n$ for some constant $\eta > 0$. It holds that all edge axioms p_{uv} and all rectangles $Q \subseteq Q(p_{uv})$ satisfy $|\mu_d(Q)| \leqslant O\big(\big(n^{\lambda/4-12\eta}/(2k)^3\big)^{-d}\big)$ for any $\lambda < 1 - \log(k)/\log(n)$.

Proof. Fix an edge $\{u, v\} \notin E(G)$, let $i, j \in [k]$ such that $u \in V_i$ and $v \in V_j$, consider the edge axiom $p_{uv} = x_u x_v$ and let $Q \subseteq Q(p_{uv})$ be an arbitrary subrectangle of this edge axiom. Recall that $\mathcal{H}_d$ denotes the set of graphs H satisfying $\mathrm{vc}(H) \leqslant d$, and as explained in Section 3.3, recall that every tuple $t \in Q$ contains the vertices u and v. Thus for $e = \{i, j\}$ we may cancel a character χ_H satisfying $e \notin H$ with the character $\chi_{H \cup \{e\}}$ to obtain that

$$\mu_d(Q) = n^{-k} \sum_{t \in Q} \sum_{H \in \mathcal{H}_d} \chi_{H(t)}(G) \tag{6.11}$$

$$= n^{-k} \sum_{t \in Q} \sum_{H \in \mathcal{H}_d(e)} \chi_{H(t)}(G) \ , \tag{6.12}$$

where $\mathcal{H}_d(e)$, as defined in Definition 4.1, denotes the set of graphs in the (d, e)-boundary, that is, all graphs H such that $\mathrm{vc}(H) = d$ and if we add the edge e to H, then the size of the minimum vertex cover increases. Let the map core be as guaranteed by Theorem 4.4. Recall that according to Proposition 4.3 the graph $\mathrm{core}(H)$ is in the (d, e)-boundary $\mathcal{H}_d(e)$ if and only if H is. Hence the sets

$$\big\{\mathrm{core}^{-1}(F) = \mathcal{H}(F, E_F^\star) \mid F \in \mathcal{H}_d(e) \text{ and } F \in \mathrm{img}(\mathrm{core})\big\} \tag{6.13}$$

partition the (d, e)-boundary and we may thus write

$$|\mu_d(Q)| = \Big| n^{-k} \sum_{t \in Q} \sum_{H \in \mathcal{H}_d(e)} \chi_{H(t)}(G) \Big| \tag{6.14}$$

$$= \Big| n^{-k} \sum_{\substack{F \in \mathcal{H}_d(e) \\ F \in \mathrm{img}(\mathrm{core})}} \sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star)} \chi_{H(t)}(G) \Big| \tag{6.15}$$

$$\leqslant n^{-k} \sum_{\substack{F \in \mathcal{H}_d(e) \\ F \in \mathrm{img}(\mathrm{core})}} \Big| \sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star)} \chi_{H(t)}(G) \Big| \ . \tag{6.16}$$

By Lemma 6.5 each inner part can be bounded by $6 \cdot 2^{|V(E(F))|} \cdot p^{-|E(F)|} \cdot n^{k - \lambda\, \mathrm{vc}(F)/4}$. Note that $\mathrm{vc}(F) = d$ and, according to Theorem 4.4, it holds that $|V\big(E(F)\big)| \leqslant 3d$. Hence since $d \leqslant 2\eta \log n$ it holds that $p^{-|E(F)|} \leqslant p^{-3d^2} = n^{6\eta d}$ and we may thus conclude that

$$|\mu_d(Q)| \leqslant \sum_{\substack{F \in \mathcal{H}_d(e) \\ F \in \mathrm{img}(\mathrm{core})}} 6 \cdot 2^{3d} \cdot n^{-d(\lambda/4 - 6\eta)} \tag{6.17}$$

$$\leqslant 2^{3d(d + \log k)} \cdot 6 \cdot 2^{3d} \cdot n^{-d(\lambda/4 - 6\eta)} \tag{6.18}$$

$$\leqslant 6 \cdot \big(n^{\lambda/4 - 12\eta}/(2k)^3\big)^{-d} \ , \tag{6.19}$$

where we used Lemma 2.4 to bound the number of core graphs and relied, again, on the assumption $d \leqslant 2\eta \log n$. This concludes the proof of Lemma 6.2. ∎

6.2 All Rectangles Are Approximately Non-Negative

In this section we prove that the measure is essentially non-negative on all monomials modulo a concentration inequality whose proof we postpone to Section 7. For convenience we restate the precise claim.

Lemma 6.3 (Restated). There are constants $\eta, c > 0$ such that if G is a D-well-behaved graph, n is large enough, $d = \eta D \leqslant 2\eta \log n$ and $D \leqslant k \leqslant n^{1/66}$, then any rectangle Q satisfies $\mu_d(Q) \geqslant -n^{-cD}$.

Recall from the proof sketch given in Section 3.4 that we intend to decompose any given rectangle into a small family $\mathcal{Q}$ of rectangles such that each rectangle $Q \in \mathcal{Q}$ either

1. contains few tuples,
2. is a subrectangle of an edge axioms, or
3. is a so-called *good rectangle*.

Since rectangles as described in Items 1 and 2 have negligible measure (see Lemma 3.5 and Lemma 6.2) essentially all the measure is concentrated on these good rectangles. Hence if we can show that the measure concentrates around a strictly positive value on such good rectangles, then the statement follows. Let us introduce these *good rectangles*.

Before defining these rectangles formally, let us give an informal description. A good rectangle Q consists of two parts. The first part is very small: on a few blocks the rectangle Q only consists of single vertices. Each vertex in this small part is adjacent to *all* other vertices in Q. Equivalently, on this small part we have a clique and the remaining vertices in Q are in the common neighborhood of this clique.

On the other blocks, where Q does not consist of a single vertex, we require that these blocks are large, of size at least $s = \mathrm{poly}(n)$. In addition we also require that all common neighborhoods are bounded on this large part. An illustration of a good rectangle can be found in Figure 3. The formal definition follows.

Definition 6.6 (Good rectangle). Let G be a k-partite graph and let $s, \beta, p, d \in \mathbb{R}^+$ and $R \subseteq [k]$. A rectangle $Q = \bigtimes_{i \in [k]} Q_i$ is (s, β, p, d, R)-*good for* G if it satisfies the following properties.

1. If $i \in R$, then $Q_i = \{v_i\}$; otherwise $|Q_i| \geqslant s$.
2. For all $i \in R$ it holds that $N(v_i) \supseteq \bigcup_{j \neq i} Q_j$.
3. For all $S \subseteq [k] \setminus R$ of size at most d and for all $i \notin R \cup S$, G has (β, p)-bounded common neighborhoods from Q_S to Q_i.

On good rectangles the measure is tightly concentrated around the expected value. In Section 7 we prove the following concentration bound.

Lemma 6.7. For constants $\varepsilon > 0$ and $\eta < 1/50$, for $n, k, d \in \mathbb{N}$ and $p = n^{-2/D} \leqslant 1/2$ satisfying $d \leqslant \eta D$ and $D \leqslant k \leqslant n$ the following holds. If $s \geqslant k^{13} n^{48\eta+\varepsilon} \log n$ and G is a D-well-behaved graph, then any $(s, 1/k, p, d, R)$-good rectangle Q for G with $|R| = \ell < d$ satisfies

$$\mu_d(Q) = p^{-\ell(k-(\ell+1)/2)} \, |Q| \, n^{-k} \big(1 \pm O(n^{-\varepsilon/8})\big) \ .$$

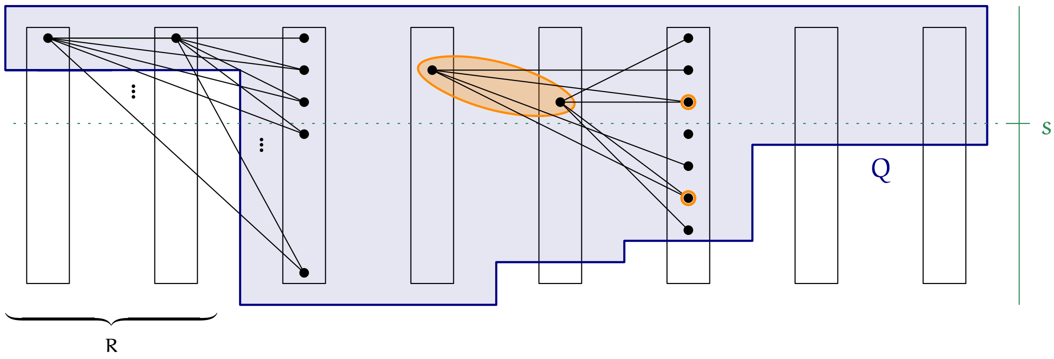

Figure 3: The rectangle Q is a good rectangles as the vertices in R have all vertices as neighbors, the blocks outside R are large and small tuples on these blocks have common neighborhoods of expected size

In the remainder of this section we prove Lemma 6.3, assuming Lemma 6.7. As outlined in Section 3.4, we intend to decompose any rectangle Q into a small family $\mathcal{Q}$ of rectangles such that each rectangle in $\mathcal{Q}$ either contains few tuples, is a subrectangle of an edge axiom or is a good rectangle. The following lemma summarizes our claim.

Lemma 6.8. Let G be a D-well-behaved graph, let $p = n^{-2/D}$, $d \leqslant D/4$ and $s \geqslant Ck^4 d \ln n/p^{2d}$ for some large enough constant C. Then any rectangle Q_0 can be partitioned into a set of rectangles $\mathcal{Q}$ of size $|\mathcal{Q}| \leqslant 2kn(2s)^d$ such that each $Q \in \mathcal{Q}$ satisfies that either

1. Q is small: $|Q| \leqslant O\big((n \cdot p^d)^{k-d}\big)$,
2. Q is a subrectangle of an edge axiom, or
3. Q is $(s, 1/k, p, d, R)$-good for G, where $R \subseteq [k]$ satisfies $|R| < d$.

Before proving Lemma 6.8, let us first show how Lemma 6.3 follows. The idea of the proof is to apply Lemma 6.8 to a given rectangle Q_0 to obtain a collection $\mathcal{Q}$ of rectangles. It holds that $\mu_d(Q_0) = \sum_{Q\in\mathcal{Q}} \mu_d(Q)$. By Lemma 3.5 there is a $\delta > 0$ such that all small rectangles $Q \in \mathcal{Q}$ satisfy $|\mu_d(Q)| \leqslant n^{-\delta D}$ and similarly by Lemma 6.2 the same holds for $Q \in \mathcal{Q}$ that are a subrectangle of an edge axiom. Further, by our choice of parameters, the size of $\mathcal{Q}$ is small—we may think of it as $n^{\delta D/2}$. We can thus lower bound

$$\mu_d(Q_0) = \sum_{Q\in\mathcal{Q}} \mu_d(Q) \geqslant -n^{-\delta D/2} + \sum_{\substack{Q\in\mathcal{Q} \\ Q \text{ is good}}} \mu_d(Q) \ . \tag{6.20}$$

Lemma 6.7 states that the remaining good rectangles in the above sum have strictly positive value. Thus $\mu_d(Q_0) \geqslant -n^{-\delta D/2}$ as claimed. In what follows we verify that this indeed holds for our choice of parameters.

Proof of Lemma 6.3. Let Q_0 be any rectangle. Our goal is to show that $\mu_d(Q_0) \geqslant -n^{-cD}$, for a sufficiently small constant c. Let $D \leqslant k \leqslant n^{1/66}$ be as in the statement of the lemma and choose $\lambda = 1 - \varepsilon - \log(k)/\log(n)$ for sufficiently small constants $\varepsilon > 0$ and $\eta > 0$ such that for $s = k^{13} n^{48\eta+\varepsilon} \log n$ it holds that $s \leqslant n^{\lambda/4-12\eta-\varepsilon}/k^3$. Let $d = \eta D \leqslant 2\eta \log n$ and $p = n^{-2/D}$. Note that for our choice of parameters it holds that $s = \omega(k^4 n^{4\eta} \log^2 n)$, hence $s = \omega(k^4 d \ln n/p^{2d})$, and we may thus apply Lemma 6.8 with $d = \eta D$ to the rectangle Q_0 to obtain a family $\mathcal{Q}$ of size at most $|\mathcal{Q}| \leqslant 2kn(2s)^d$.

By Lemma 6.2, any subrectangle of an axiom has measure bounded by $O\big((n^{\lambda/4-12\eta}/(2k)^3)^{-d}\big)$. Moreover, according to Lemma 3.5 each small rectangle $Q \in \mathcal{Q}$ has measure of magnitude at most

$$|\mu_d(Q)| \leqslant O\big(|Q|n^{-k}k^d p^{-dk}\big) = O(n^{-d/2}) \ , \tag{6.21}$$

which is even smaller than the bound on axioms. Since $s \leqslant n^{\lambda/4-12\eta-\varepsilon}/k^3$, we conclude that the measure of all small rectangles and all subrectangles of axioms in $\mathcal{Q}$ add up, in magnitude, to at most $|\mathcal{Q}| \cdot O\big((n^{\lambda/4-12\eta}/(2k)^3)^{-d}\big) \leqslant n^{-cD}$, for a small enough constant c.

Hence the measure of Q_0 is mostly on the good rectangles of $\mathcal{Q}$ and on these rectangles we know that it is closely concentrated around a strictly positive value. Indeed, we can apply Lemma 6.7 to any rectangle Q which is $(s, 1/k, p, d, R)$-good for G to conclude that

$$\mu_d(Q) = p^{-\ell(k-(\ell+1)/2)}\,|Q|\,n^{-k}\big(1 \pm O(n^{-\varepsilon/8})\big) > 0 \ .$$

This concludes the proof of Lemma 6.3 modulo Lemma 6.7 and Lemma 6.8. $\square$

Let us proceed to prove Lemma 6.8.

Proof of Lemma 6.8. Let us describe a recursive decomposition procedure that can be applied to any rectangle $Q = \bigtimes_{i\in[k]} Q_i$.

If either Q is small, a subrectangle of an axiom or $(s, 1/k, p, d, R)$-good for some $R \subseteq [k]$, then return Q. Otherwise decompose in the following recursive fashion.

1. If there is a singleton $Q_i = \{v_i\}$ such that $N(v_i) \not\supseteq \bigcup_{j\neq i} Q_j$, then we decompose Q into $|Q \setminus N(v_i)| + 1$ many rectangles as follows. Denote by $u_1, u_2, \ldots, u_m$ the vertices in Q that are not a neighbor of v_i and assume that they are in blocks $j_1, j_2, \ldots, j_m$. For $v = 1, \ldots, m$ we remove all tuples that contain the vertex u_v: let $R^0 = Q$ so we can write

$$Q^v = \{u_v\} \times \bigtimes_{j\neq j_v} R_j^{v-1} \qquad \text{and} \qquad R^v = \big(R_{j_v}^{v-1} \setminus u_v\big) \times \bigtimes_{j\neq j_v} R_j^{v-1} \ . \tag{6.22}$$

   Note that the rectangles $Q^1, \ldots, Q^m, R^m$ partition Q. Add the Q^v to the partition as these are subrectangles of edge axioms and recursively decompose R^m.
2. If there is a block $i \in [k]$ of size $1 < |Q_i| \leqslant 2s$, then split Q into the $|Q_i|$ rectangles

$$\Big\{\{v_i\} \times \bigtimes_{j\neq i} Q_j : v_i \in Q_i\Big\} \tag{6.23}$$

   and recursively decompose each of these rectangles.
3. Let A be the set of blocks of size greater than $2s$. Because G is D-well-behaved, by Property 2 of Definition 5.4, it holds that G has $(2s, s, 1/k, p, d)$-bounded error sets. In particular Q_A has an error set $U = \{u_1, \ldots, u_m\}$ of size at most s. Decompose Q into $Q^1, \ldots, Q^m$ and R^m as in Case 1. By definition the rectangle R^m is $(s, 1/k, p, d, [k] \setminus A)$-good and we may thus add it to the partition. Recursively decompose the rectangles $Q^1, \ldots, Q^m$.

This completes the description of the decomposition procedure. We need to argue that the decomposition $\mathcal{Q}$ created by above procedure is not too large, that is, of size $|\mathcal{Q}| \leqslant 2kn \cdot (2s)^d$. Let us start with a few observations.

Because G is D-well-behaved it holds that G has $(1/k, p, D/4)$-bounded common neighborhoods in every block (see Property 1 of Definition 5.4). Let Q be a rectangle with d blocks of size

1 and with the remaining vertices contained in the common neighborhood of these singletons. All such rectangles Q are small. Thus the decomposition procedure does not need to decompose such rectangles Q any further.

Whenever we decompose a rectangle in Cases 2 and 3 all rectangles that we need to recursively decompose have one more singleton. Because we can stop decomposing after identifying d singletons and in Cases 2 and 3 we create at most $2s$ many rectangles that require further decomposition we end up with at most $2(2s)^d$ many rectangles. We ignored the rectangles from Case 1 so far. But each rectangle that requires further decomposition from Cases 2 and 3 results in at most another kn many rectangles from Case 1. Thus the size of the family of rectangles is bounded by $2kn \cdot (2s)^d$. □

7 The Measure Is Concentrated on Good Rectangles

This section is devoted to proving Lemma 6.7 that states that the measure on good rectangles is very well concentrated. We restate it here for convenience.

Lemma 6.7 (Restated). For constants $\varepsilon > 0$ and $\eta < 1/50$, for $n, k, d \in \mathbb{N}$ and $p = n^{-2/D} \leqslant 1/2$ satisfying $d \leqslant \eta D$ and $D \leqslant k \leqslant n$ the following holds. If $s \geqslant k^{13} n^{48\eta+\varepsilon} \log n$ and G is a D-well-behaved graph, then any $(s, 1/k, p, d, R)$-good rectangle Q for G with $|R| = \ell < d$ satisfies

$$\mu_d(Q) = p^{-\ell(k-(\ell+1)/2)} |Q| \, n^{-k} \big(1 \pm O(n^{-\varepsilon/8})\big) \ .$$

Let us introduce some notation and state Lemma 6.7 once more in a more convenient form for what follows. Let $S_i^{\leqslant j}$ be the star with center i and leaves $[j] \setminus i$, that is, $S_i^{\leqslant j}$ consists of the vertices $[k]$ and edges $\big\{\{i, j'\} : j' \in [j], j' \neq i\big\}$. For simplicity we denote by S_i the star $S_i^{\leqslant k}$ and for $I \subseteq [k]$ let $S_I = \cup_{i \in I} S_i$.

Lemma 7.1. For constants $\varepsilon > 0$ and $\eta < 1/50$, for $n, k, d \in \mathbb{N}$ and $p = n^{-2/D} \leqslant 1/2$ satisfying $d \leqslant \eta D$, $D \leqslant k \leqslant n$ the following holds. If $s \geqslant k^{13} n^{48\eta+\varepsilon} \log n$ and G is a D-well-behaved graph, then any $(s, 1/k, p, d, R)$-good rectangle Q for G with $|R| < d$ satisfies

$$\Big| p^{-|S_R|} |Q| - \sum_{t \in Q} \sum_{H \in \mathcal{H}_d} \chi_{H(t)}(G) \Big| \leqslant O\big(p^{-|S_R|} |Q| \, n^{-\varepsilon/8}\big) \ .$$

Before proving Lemma 7.1 let us verify that Lemma 6.7 indeed follows.

Proof of Lemma 6.7. Recall that we defined our measure μ_d for a rectangle Q as

$$\mu_d(Q) = n^{-k} \sum_{t \in Q} \sum_{H \in \mathcal{H}_d} \chi_{H(t)}(G) \ . \tag{7.1}$$

Hence Lemma 7.1 implies, for the appropriate parameters, that all $(s, 1/k, p, d, R)$-good rectangles Q for G satisfy

$$\mu_d(Q) = p^{-|S_R|} |Q| \, n^{-k} \big(1 \pm O(n^{-\varepsilon/8})\big) \ . \qquad \square$$

Let us give some intuition for the statement of Lemma 7.1. Clearly, if $R = \emptyset$, then we are showing that the measure of a rectangle is tightly concentrated around the expected value. Let us explain where the $p^{-|S_{[\ell]}|}$ factor comes from.

Consider two blocks, say blocks 1 and 2. For vertices $v_1 \in V_1$ and $v_2 \in V_2$ let us denote by $Q_{v_1 v_2}$ the rectangle consisting of all tuples that contain v_1 as well as v_2. According to Lemma 6.2, if there is no edge between v_1 and v_2, then $\mu_d(Q_{v_1 v_2}) \approx 0$. Recall that the measure of the whole space is roughly 1—hence if there is an edge between vertices v_1 and v_2 as above, we expect that $\mu(Q_{v_1 v_2})$ "compensates" for the 0 value rectangles, that is, we expect that $\mu(Q_{v_1 v_2}) \approx (|Q_{v_1 v_2}|/n^k)) \cdot (1/p) = 1/pn^2$ if the edge $v_1 v_2$ is present. More generally, for $R \subseteq [k]$ we expect to pick up a factor of $1/p$ for each edge that we condition on being present between a vertex in R and another block in Q. Lemma 7.1 establishes that this is indeed how the measure behaves.

Let us consider an $(s, 1/k, p, d, R)$-good rectangle Q. For ease of exposition let us assume that $R = [\ell]$, the other cases are analogous. In other words, we assume, for all $i \in [\ell]$, that it holds that $|Q_i| = 1$, these ℓ vertices form a clique and all edges from the first ℓ vertices to any other vertex in Q are present.

Recall that $\mathcal{H}_d$ is the family of graphs on k labeled vertices with a vertex cover of size at most d, and that $\mathcal{H}_d(e) \subseteq \mathcal{H}_d$ denotes the set of graphs in the (d, e)-boundary, as defined in Section 4. Finally, for two graphs H and H' defined over the same vertex set V we denote by $H \setminus H'$ the graph over V containing all edges of H that are not present in H'.

We prove Lemma 7.1 in three steps. First we split the sum of Fourier characters into two parts: the *main sum* and some *boundary sums*. In a second step we show that the boundary sums are negligible, i.e., that they add up to very little. In the final step we then show that the main sum is tightly concentrated around the expected value.

The following claim splits the sum of Fourier characters into the main sum and several boundary sums. We postpone the straightforward proof until after we prove Lemma 7.1.

Claim 7.2. For Q as in Lemma 7.1 and any tuple $t \in Q$ it holds that

$$\sum_{H \in \mathcal{H}_d} \chi_{H(t)}(G) = p^{-|S_{[\ell]}|} \sum_{\substack{H \in \mathcal{H}_d \\ S_{[\ell]} \subseteq H}} \chi_{(H \setminus S_{[\ell]})(t)}(G)$$

$$+ \sum_{i \in [\ell]} \sum_{j=i+1}^{k} p^{-|S_{[i-1]} \cup S_i^{\leqslant j-1}|} \sum_{\substack{H \in \mathcal{H}_d(\{i,j\}) \\ S_{[i-1]} \cup S_i^{\leqslant j-1} \subseteq H}} \chi_{(H \setminus (S_{[i-1]} \cup S_i^{\leqslant j-1}))(t)}(G) \ .$$

The sum with the factor $p^{-|S_{[\ell]}|}$ in Claim 7.2 is the so-called *main sum*. Intuitively most of the measure is in the main sum and it adds up to approximately $p^{-|S_{[\ell]}|}$ times the size of Q, that is,

$$p^{-|S_{[\ell]}|} \sum_{t \in Q} \sum_{\substack{H \in \mathcal{H}_d \\ S_{[\ell]} \subseteq H}} \chi_{(H \setminus S_{[\ell]})(t)}(G) \approx p^{-|S_{[\ell]}|} \cdot |Q| \ . \tag{7.2}$$

The latter sums in Claim 7.2 with coefficients $p^{-|S_{[i-1]} \cup S_i^{\leqslant j-1}|}$ are the so-called *boundary sums*. All of these sums turn out to be tiny, that is, for $i \in [\ell]$ and $j + 1 \leqslant i \leqslant k$ it holds that

$$p^{-|S_{[i-1]} \cup S_i^{\leqslant j-1}|} \left| \sum_{t \in Q} \sum_{\substack{H \in \mathcal{H}_d(\{i,j\}) \\ S_{[i-1]} \cup S_i^{\leqslant j-1} \subseteq H}} \chi_{(H \setminus (S_{[i-1]} \cup S_i^{\leqslant j-1}))(t)}(G) \right| \lesssim p^{-|S_{[\ell]}|} \cdot |Q| \cdot n^{-(d-\ell)} \ . \tag{7.3}$$

Both of the bounds corresponding to (7.2) and (7.3), and stated formally below, are proven in Section 7.1. Claim 7.2 together with these bounds will allow us to conclude that the measure $\mu(Q)$ is tightly concentrated around $p^{-|S_{[\ell]}|} \cdot |Q|$.

Let us first state the bound for the main sum. Intuitively we may think of the below lemma as a version of Lemma 6.4 that holds on a local part of the graph.

Lemma 7.3. For G as in Lemma 7.1 the following holds for $p = n^{-2/D}$ and $\ell < d \leqslant \eta D \leqslant 2\eta \log n$. If $s \geqslant 10\, k^{13} n^{48\eta+\varepsilon} \log n$, then all $(s, 1/k, p, d, [\ell])$-good rectangles Q satisfy

$$\left| |Q| - \sum_{t \in Q} \sum_{\substack{H \in \mathcal{H}_d \\ S_{[\ell]} \subseteq H}} \chi_{(H \setminus S_{[\ell]})(t)}(G) \right| \leqslant O\left(|Q|\, n^{-\varepsilon/4}\right) .$$

On the other hand, reminiscent of the edge axioms, we have that the boundary sums are quite small. Note that, in contrast to Lemma 6.5, the bound below depends on the size of the rectangle Q.

Lemma 7.4. For G as in Lemma 7.1 the following holds for any constant $0 < \eta < 1/50$ and $\varepsilon > 0$, for $p = n^{-2/D}$ and $\ell < d \leqslant \eta D \leqslant 2\eta \log n$. Let $D \leqslant k \leqslant n$, $s \geqslant 10\, k^2 n^{48\eta+\varepsilon} \log n$ and $i \in [\ell]$. Then all $(s, 1/k, p, d, [\ell])$-good rectangles Q satisfy that if $j \leqslant d+2$, then

$$p^{-|S_{[i-1]} \cup S_i^{\leqslant j-1}|} \cdot \left| \sum_{t \in Q} \sum_{\substack{H \in \mathcal{H}_d(\{i,j\}) \\ S_{[i-1]} \cup S_i^{\leqslant j-1} \subseteq H}} \chi_{(H \setminus (S_{[i-1]} \cup S_i^{\leqslant j-1}))(t)}(G) \right| \leqslant O\left(p^{-|S_{[\ell]}|}\, |Q|\, n^{-\varepsilon(d-\ell)/4}\right) ,$$

and, on the other hand, if $j \geqslant d+3$, then the above sum is empty.

Assuming these statements the proof of Lemma 7.1 boils down to a sequence of syntactic manipulations.

Proof of Lemma 7.1. According to Claim 7.2 the expression

$$\left| p^{-|S_{[\ell]}|} \cdot |Q| - \sum_{t \in Q} \sum_{H \in \mathcal{H}_d} \chi_{H(t)}(G) \right| \tag{7.4}$$

is equal to

$$\begin{aligned} \Bigg| p^{-|S_{[\ell]}|} \cdot |Q| - p^{-|S_{[\ell]}|} \sum_{t \in Q} \sum_{\substack{H \in \mathcal{H}_d \\ S_{[\ell]} \subseteq H}} \chi_{(H \setminus S_{[\ell]})(t)}(G) \\ - \sum_{i \in [\ell]} \sum_{j=i+1}^{k} p^{-|S_{[i-1]} \cup S_i^{\leqslant j-1}|} \sum_{t \in Q} \sum_{\substack{H \in \mathcal{H}_d(\{i,j\}) \\ S_{[i-1]} \cup S_i^{\leqslant j-1} \subseteq H}} \chi_{(H \setminus (S_{[i-1]} \cup S_i^{\leqslant j-1}))(t)}(G) \Bigg| . \end{aligned} \tag{7.5}$$

Appealing to the triangle inequality, Lemma 7.3 and Lemma 7.4, we may upper bound (7.5) by

$$
\begin{aligned}
p^{-|S_{[\ell]}|} \cdot & \left| |Q| - \sum_{t\in Q} \sum_{\substack{H\in\mathcal{H}_d \\ S_{[\ell]}\subseteq H}} \chi_{(H\setminus S_{[\ell]})(t)}(G) \right| \\
& + \sum_{i\in[\ell]} \sum_{j=i+1}^{d+2} p^{-|S_{[i-1]}\cup S_i^{\leqslant j-1}|} \cdot \left| \sum_{t\in Q} \sum_{\substack{H\in\mathcal{H}_d(\{i,j\}) \\ S_{[i-1]}\cup S_i^{\leqslant j-1}\subseteq H}} \chi_{(H\setminus(S_{[i-1]}\cup S_i^{\leqslant j-1}))(t)}(G) \right| \qquad (7.6)\\
& \leqslant O\Big(p^{-|S_{[\ell]}|} \, |Q| \, \big(n^{-\varepsilon/4} + \sum_{i\in[\ell]} \sum_{j=i+1}^{d+2} n^{-\varepsilon(d-\ell)/4}\big)\Big) \\
& \leqslant O\big(p^{-|S_{[\ell]}|} \, |Q| \, n^{-\varepsilon/8}\big) \ . \qquad (7.7)
\end{aligned}
$$

Putting everything together we may conclude that (7.4) is upper bounded by the expression in (7.7), as claimed. □

We now proceed to prove Claim 7.2.

Proof of Claim 7.2. Suppose $\ell \geqslant 1$ and let us start by considering the edge $\{1,2\}$. We first observe that for any H such that $\{1,2\} \in H$, it holds that

$$
\chi_{H(t)}(G) = \frac{1-p}{p} \chi_{(H\setminus\{1,2\})(t)}(G) \qquad (7.8)
$$

by definition of a good rectangle as every tuple $t \in Q$ contains the edge $\{1,2\}$. Let us write

$$
\sum_{H\in\mathcal{H}_d} \chi_{H(t)}(G) = \frac{1-p}{p} \cdot \sum_{\substack{H\in\mathcal{H}_d \\ \{1,2\}\in H}} \chi_{(H\setminus\{1,2\})(t)}(G) + \sum_{\substack{H\in\mathcal{H}_d \\ \{1,2\}\notin H}} \chi_{H(t)}(G) \ . \qquad (7.9)
$$

Note that we can partition the set $\{H \in \mathcal{H}_d : \{1,2\} \notin H\}$ into two parts: the first part contains all the graphs in the boundary of $\{1,2\}$, that is, graphs $H \in \mathcal{H}_d(\{1,2\})$, and the second set contains all the remaining graphs. Note that graphs H contained in the latter set satisfy that $H \cup \{1,2\} \in \mathcal{H}_d$. Hence, for every $t \in Q$, it holds that

$$
\begin{aligned}
\sum_{H\in\mathcal{H}_d} \chi_{H(t)}(G) = \frac{1-p}{p} \cdot & \sum_{\substack{H\in\mathcal{H}_d \\ \{1,2\}\in H}} \chi_{(H\setminus\{1,2\})(t)}(G) \\
& + \sum_{\substack{H\in\mathcal{H}_d \\ \{1,2\}\in H}} \chi_{(H\setminus\{1,2\})(t)}(G) + \sum_{H\in\mathcal{H}_d(\{1,2\})} \chi_{H(t)}(G) \qquad (7.10)\\
= \frac{1}{p} \cdot & \sum_{\substack{H\in\mathcal{H}_d \\ \{1,2\}\in H}} \chi_{(H\setminus\{1,2\})(t)}(G) + \sum_{H\in\mathcal{H}_d(\{1,2\})} \chi_{H(t)}(G) \ . \qquad (7.11)
\end{aligned}
$$

If we continue to rewrite the first sum in the above fashion for edges $\{1,3\},\ldots,\{1,k\}$ we obtain

$$
\begin{aligned}
\sum_{H\in\mathcal{H}_d} \chi_{H(t)}(G) = p^{-(k-1)} \cdot & \sum_{\substack{H\in\mathcal{H}_d \\ S_1\subseteq H}} \chi_{(H\setminus S_1)(t)}(G) \\
& + \sum_{j=2}^{k} p^{-(j-2)} \cdot \sum_{\substack{H\in\mathcal{H}_d(\{1,j\}) \\ S_1^{\leqslant j-1}\subseteq H}} \chi_{(H\setminus S_1^{\leqslant j-1})(t)}(G) \ . \qquad (7.12)
\end{aligned}
$$

By iteratively applying the above arguments to vertices $i = 2, \dots, \ell$ we conclude that

$$\sum_{H \in \mathcal{H}_d} \chi_{H(t)}(G) = p^{-|S_{[\ell]}|} \cdot \sum_{\substack{H \in \mathcal{H}_d \\ S_{[\ell]} \subseteq H}} \chi_{(H \setminus S_{[\ell]})(t)}(G) \\ + \sum_{i \in [\ell]} \sum_{j=i+1}^{k} p^{-|S_{[i-1]} \cup S_i^{\leqslant j-1}|} \sum_{\substack{H \in \mathcal{H}_d(\{i,j\}) \\ S_{[i-1]} \cup S_i^{\leqslant j-1} \subseteq H}} \chi_{(H \setminus (S_{[i-1]} \cup S_i^{\leqslant j-1}))(t)}(G) \ , \tag{7.13}$$

as claimed. □

7.1 Concentration of the Main Sum and Bounding the Boundary Sums

In this section we prove Lemma 7.3 and Lemma 7.4. We rely on the following lemma which guarantees bounded character sum on arbitrary good rectangles, regardless of size. This is in contrast to Property 4 of Definition 5.4 which only holds for small rectangles. The more general statement follows from the latter by splitting any large rectangle into many small rectangles while maintaining goodness so that Property 4 of Definition 5.4 holds for these small rectangles.

Lemma 7.5. Let $n, k \in \mathbb{N}$ and $s, \eta, D, d \in \mathbb{R}^+$ be such that $d \leqslant \eta D$, $k \leqslant n$ and $s \geqslant 9\,k^2 n^{2\eta} d \ln n$. If $p = n^{-2/D}$, G is a D-well-behaved graph and F is a core graph, then all $(s, 1/k, p, d, R)$-good rectangles Q satisfy, for $B = [k] \setminus R$, that

$$\left|\sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star[B])} \chi_{H[B](t)}(G)\right| = O\left(|Q| \cdot p^{-|E(F[B])|} \cdot (s/10\,k \log n)^{-\mathrm{vc}(F[B])/4}\right) \ .$$

We defer the proof of this lemma to Section 7.2. Before we proceed to prove Lemma 7.3 and Lemma 7.4 let us record two claims regarding vertex covers.

Claim 7.6. Any minimum vertex cover of a graph H contains all vertices of degree at least $\mathrm{vc}(H) + 1$.

Proof. Any vertex cover of H that does not contain a vertex u of degree at least $\mathrm{vc}(H) + 1$ must contain all neighbors of u and hence is of size at least $\mathrm{vc}(H) + 1$. □

Claim 7.7. For $i \in [k-1]$ and a graph H on k vertices it holds that if $S_{[i]} \subseteq H$, then $\mathrm{vc}(H \setminus S_{[i]}) = \mathrm{vc}(H) - i$.

Proof. If $i = k - 1$, then $H = S_{[i]}$ is the complete graph on k vertices and the statement readily follows. Otherwise, as every vertex in $[i]$ has degree $k - 1 \geqslant \mathrm{vc}(H) + 1$, it follows by Claim 7.6 that the set $[i]$ is contained in any minimum vertex cover W of H. Since the vertices in $W \setminus [i]$ have to cover the edges in $H \setminus S_{[i]}$, and since any vertex cover W' of $H \setminus S_{[i]}$ is such that $W' \cup [i]$ is a vertex cover of H, we may conclude $\mathrm{vc}(H \setminus S_{[i]}) = \left|W \setminus [i]\right| = \mathrm{vc}(H) - i$. □

Finally we also need to revisit our map core as we need a good bound on the number of cores that the graphs containing $S_{[\ell]}$ are mapped to. Before stating the claim let us recall the construction of the map core, as done in Section 4.2: given a graph H with lexicographically first vertex cover W we let U_1 be the lex first maximal set of vertices with a matching from U_1 to W that covers all vertices in U_1. Similarly we let U_2 be the lex first maximal set of vertices in $H \setminus U_1$ with a matching from U_2 to W covering all vertices in U_2. Finally we define $\mathrm{core}(H) = H[W \cup U_1 \cup U_2]$.

Lemma 7.8. Let $k, \ell, w \in \mathbb{N}$. If $\mathcal{F}_\ell(w)$ denotes the set of cores $F \in \mathrm{img}(\mathrm{core})$ such that $S_{[\ell]} \subseteq F \cup E_F^\star$ and $\mathrm{vc}(F) = \ell + w$, then it holds that

$$|\mathcal{F}_\ell(w)| \leqslant 2^{3w(\ell+w+\log k)} \ .$$

Proof. We argue that we can encode elements of the set of cores $\mathcal{F}_\ell(w)$ by using few bits. Fix a core $F \in \mathcal{F}_\ell(w)$ and let W, U_1, U_2 as in the construction of F (see Section 4.2 and the discussion above). By definition we have that $|W| = \ell + w$ and $|U_2| \leqslant |U_1| \leqslant \ell + w$.

According to Claim 7.6 every vertex in $[\ell]$ is in any minimum vertex cover of F and thus contained in W; the lex first one. Hence there are only w vertices in $W \cap \{\ell+1, \ldots, k\}$ to be specified. We spend $\log \binom{k}{w}$ bits to specify this set. Once we have specified this set we know that the lex smallest ℓ vertices outside W are contained in U_1 by construction. With another $\log \binom{k}{w}$ bits we may thus encode the remaining vertices of U_1. Similarly, for U_2, we know that the lex smallest ℓ vertices outside $W \cup U_1$ are in U_2. Hence $\log \binom{k}{w}$ bits suffice to specify U_2.

At this point we know the relevant vertices of the core. It remains to encode the edges. Some edges are given: all edges from $[\ell]$ to the rest of the core are present. As such there at most $w \cdot 3(\ell + w)$ many edges left to be specified. The claim follows. □

With these claims at hand we are ready to prove Lemma 7.3 and Lemma 7.4. We start with Lemma 7.3 restated here for convenience.

Lemma 7.3 (Restated). For G as in Lemma 7.1 the following holds for $p = n^{-2/D}$ and $\ell < d \leqslant \eta D \leqslant 2\eta \log n$. If $s \geqslant 10\, k^{13} n^{48\eta+\varepsilon} \log n$, then all $(s, 1/k, p, d, [\ell])$-good rectangles Q satisfy

$$\Big| |Q| - \sum_{t \in Q} \sum_{\substack{H \in \mathcal{H}_d \\ S_{[\ell]} \subseteq H}} \chi_{(H \setminus S_{[\ell]})(t)}(G) \Big| \leqslant O\big(|Q|\, n^{-\varepsilon/4}\big) \ .$$

Proof. Let $B = \{\ell+1, \ldots, k\}$ and observe that $H \setminus S_{[\ell]} = H[B]$. Further using that for $H = S_{[\ell]}$ it holds that $\sum_{t \in Q} \chi_{(H \setminus S_{[\ell]})(t)}(G) = |Q|$ we may bound

$$\Big| |Q| - \sum_{t \in Q} \sum_{\substack{H \in \mathcal{H}_d \\ S_{[\ell]} \subseteq H}} \chi_{(H \setminus S_{[\ell]})(t)}(G) \Big| = \Big| \sum_{t \in Q} \sum_{w=\ell+1}^{d} \sum_{\substack{F \in \mathrm{img}(\mathrm{core}) \\ \mathrm{vc}(F) = w}} \sum_{\substack{H \in \mathcal{H}(F, E_F^\star) \\ S_{[\ell]} \subseteq H}} \chi_{H[B](t)}(G) \Big| \tag{7.14}$$

$$= \Big| \sum_{w_B=1}^{d-\ell} \sum_{\substack{F \in \mathrm{img}(\mathrm{core}) \\ S_{[\ell]} \subseteq F \cup E_F^\star \\ \mathrm{vc}(F[B]) = w_B}} \sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star[B])} \chi_{H[B](t)}(G) \Big| \tag{7.15}$$

$$\leqslant \sum_{w_B=1}^{d-\ell} \sum_{\substack{F \in \mathrm{img}(\mathrm{core}) \\ S_{[\ell]} \subseteq F \cup E_F^\star \\ \mathrm{vc}(F[B]) = w_B}} \Big| \sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star[B])} \chi_{H[B](t)}(G) \Big| \ , \tag{7.16}$$

where we appeal to Claim 7.7 to obtain (7.15) and to the triangle inequality for (7.16). Since by Theorem 4.4 it holds that $\left|V\big(E(F)\big)\right| \leqslant 3 \cdot \mathrm{vc}(F)$ we may apply Lemma 7.5 with $\left|E\big(F[B]\big)\right| \leqslant 3 \cdot \mathrm{vc}(F) \cdot \mathrm{vc}\big(F[B]\big) \leqslant 3d \cdot \mathrm{vc}\big(F[B]\big)$ to the inner expressions to obtain the bound

$$\left||Q| - \sum_{t\in Q} \sum_{\substack{H\in\mathcal{H}_d\\ S_{[\ell]}\subseteq H}} \chi_{H[B](t)}(G)\right| \leqslant |Q| \cdot \sum_{w_B=1}^{d-\ell} \sum_{\substack{F\in\mathrm{img}(\mathrm{core})\\ S_{[\ell]}\subseteq F\cup E_F^\star\\ \mathrm{vc}(F[B])=w_B}} O\big(p^{-3d\cdot w_B} \cdot (s/10\,k\log n)^{-w_B/4}\big) \ . \tag{7.17}$$

Applying Lemma 7.8 along with the fact that $\ell + w_B \leqslant d \leqslant \eta D \leqslant 2\eta \log n$ allows us to conclude

$$\left||Q| - \sum_{t\in Q} \sum_{\substack{H\in\mathcal{H}_d\\ S_{[\ell]}\subseteq H}} \chi_{H[B](t)}(G)\right| \leqslant O\Big(|Q| \cdot \sum_{w_B=1}^{d-\ell} 2^{3w_B(d+\log k)} \cdot (s/10\,n^{24\eta}k\log n)^{-w_B/4}\Big) \tag{7.18}$$

$$\leqslant O\Big(|Q| \cdot \sum_{w_B=1}^{d-\ell} (s/10\,k^{13}n^{48\eta}\log n)^{-w_B/4}\Big) \tag{7.19}$$

$$\leqslant O\big(|Q| \cdot n^{-\varepsilon/4}\big) \ , \tag{7.20}$$

where we used that $\ell < d$, the bound $s \geqslant 10\,k^{13}n^{48\eta+\varepsilon}\log n$ and the fact that the final sum is a geometric series. This completes the proof of Lemma 7.3 modulo Lemma 7.5. □

In order to show that the magnitude of the boundary sums is small, we need to bound the number of core graphs that contain the edge set $S_{i-1} \,\dot\cup\, S_i^{\leqslant j-1}$. The bound from Lemma 7.8 turns out to be insufficient as it is with respect to $d-(i-1)$, that is, the vertex cover size outside the set $[i-1]$; as we can apply Lemma 7.5 with $B = \{\ell+1, \dots, k\}$ only, we get concentration with respect to the size of the vertex cover in the set $\{\ell+1,\dots,k\}$. This is a problem as the vertex cover in the set $\{\ell+1,\dots,k\}$ may be considerably smaller than in the set $\{i,\dots,k\}$. We overcome this difference in parameters by leveraging the difference in the exponential factors $p^{|S_{[i-1]}\cup S_i^{\leqslant j-1}|}$ and $p^{|S_{[\ell]}|}$ as follows.

Claim 7.9. Let $w_B, d, i, j, \ell \in [k]$, $i \leqslant \ell$, $B = \{\ell+1,\dots,k\}$, $E_0 = S_{[i-1]} \,\dot\cup\, S_i^{\leqslant j-1}$ and $E_1 = S_{[\ell]} \setminus E_0$. It holds that

$$\sum_{\substack{F\in\mathrm{img}(\mathrm{core})\\ F\in\mathcal{H}_d(\{i,j\})\\ E_0\subseteq F\cup E_F^\star\\ \mathrm{vc}(F[B])=w_B}} 2^{-|E_1\setminus E_F^\star|} \leqslant 3d^2 \cdot \binom{k}{d}^3 \cdot 2^{-k+3dw_B+4d+\ell} \ .$$

Proof. We first argue that the set over which we sum is not so large and then provide a lower bound on $|E_1 \setminus E_F^\star|$. These two bounds will together imply the claim.

Let $C = \{i,\dots,k\}$. Given parameters $w_C \leqslant d$ and $c \leqslant 3d$, we consider cores $F \in \mathrm{img}(\mathrm{core})$ such that $F \in \mathcal{H}_d(\{i,j\})$, $E_0 \subseteq F \cup E_F^\star$, $\mathrm{vc}(F[B]) = w_B$, $\mathrm{vc}(F[C]) = w_C$ and $\big|V\big(E(F)\big) \cap C\big| = c$. We claim that there are at most

$$\binom{k}{d}^3 \cdot 2^{\binom{c}{2}-\binom{c-w_C}{2}} \tag{7.21}$$

such core graphs. Indeed, it is enough to specify the $w_C \leqslant d$ vertices in a minimum vertex cover of $F[C]$, the other at most $2d$ vertices in $V(E(F)) \cap C$, and the edges between vertices in the identified minimum vertex cover of $F[C]$ and $V(E(F)) \cap C$.

In order to obtain a lower bound on $|E_1 \setminus E_F^\star|$, note that the edge set $E_F^\star$ consists only of edges with exactly one endpoint in $V(E(F))$. Hence all edges that are in E_1 and that have both endpoints in $V(E(F))$ are also in $E_1 \setminus E_F^\star$. These include edges between $V(E(F)) \cap \{i+1, \ldots, \ell\}$ and $V(E(F)) \cap C$. Note that all edges in $F[C] \setminus F[B]$ have one endpoint in $V(E(F)) \cap \{i, \ldots, \ell\}$ and thus

$$|V(E(F)) \cap \{i, \ldots, \ell\}| \geqslant \mathrm{vc}(F[C] \setminus F[B]) \geqslant \mathrm{vc}(F[C]) - \mathrm{vc}(F[B]) = w_C - w_B \ . \tag{7.22}$$

This implies that $|V(E(F)) \cap \{i+1, \ldots, \ell\}| \geqslant w_C - w_B - 1$ and therefore

$$|E_1 \setminus E_F^\star| \geqslant \binom{c}{2} - \binom{c - (w_C - w_B - 1)}{2} \ . \tag{7.23}$$

This bound, however, is not good enough for us. We need to also count edges in $E_1 \setminus E_F^\star$ that are adjacent to i and that we have not considered yet. Note that these include all the edges in $(E_1 \setminus E_F^\star) \cap (\{i\} \times B)$. Let $\widetilde{N}(i)$ be the neighbors of i in the graph $F \cup E_F^\star$. In particular, we have that $[j-1] \subseteq \widetilde{N}(i)$ since $S_i^{\leqslant j-1} \subseteq E_0 \subseteq F \cup E_F^\star$. This implies that $\{i\} \times (B \setminus \widetilde{N}(i)) \subseteq E_1 \setminus E_F^\star$ and therefore $|(E_1 \setminus E_F^\star) \cap (\{i\} \times B)| \geqslant |\{i\} \times (B \setminus \widetilde{N}(i))| \geqslant |B| - |\widetilde{N}(i)|$. Now note that it must be the case that $\widetilde{N}(i)$ is in every minimum vertex cover of F; otherwise we would have a minimum vertex cover of F that either contains i (and then F would not be in the $(d, \{i,j\})$-boundary) or is not a vertex cover of the graph $F \cup E_F^\star$. This implies that $|\widetilde{N}(i)| \leqslant d$. We may thus conclude that

$$|E_1 \setminus E_F^\star| \geqslant \binom{c}{2} - \binom{c - (w_C - w_B - 1)}{2} + k - \ell - d \ . \tag{7.24}$$

Using the bounds in Equation (7.21) and Equation (7.24) we may write

$$\sum_{\substack{F \in \mathrm{img}(\mathrm{core}) \\ F \in \mathcal{H}_d(\{i,j\}) \\ E_0 \subseteq F \cup E_F^\star \\ \mathrm{vc}(F[B]) = w_B \\ \mathrm{vc}(F[C]) = w_C \\ |V(E(F)) \cap C| = c}} 2^{-|E_1 \setminus E_F^\star|} \leqslant \binom{k}{d}^3 \cdot 2^{\binom{c}{2} - \binom{c - w_C}{2}} \cdot 2^{-\binom{c}{2} + \binom{c - (w_C - w_B - 1)}{2} - k + \ell + d} \tag{7.25}$$

$$= \binom{k}{d}^3 \cdot 2^{(c - w_C)(w_B + 1) - \binom{w_B + 1}{2} - k + \ell + d} \ . \tag{7.26}$$

Since $c \leqslant 3d$, we have that $(c - w_C)(w_B + 1) - \binom{w_B+1}{2} \leqslant 3d(w_B + 1)$. Finally, we need to sum over all possible values of w_C and c. Since $w_C \leqslant d$ and $c \leqslant 3d$, the statement follows. $\square$

In the remainder of this section we prove that boundary sums are small. For convenience we restate the claim.

Lemma 7.4 (Restated). For G as in Lemma 7.1 the following holds for any constant $0 < \eta < 1/50$ and $\varepsilon > 0$, for $p = n^{-2/D}$ and $\ell < d \leqslant \eta D \leqslant 2\eta \log n$. Let $D \leqslant k \leqslant n$, $s \geqslant 10\, k^2 n^{48\eta+\varepsilon} \log n$ and $i \in [\ell]$. Then all $(s, 1/k, p, d, [\ell])$-good rectangles Q satisfy that if $j \leqslant d+2$, then

$$p^{-|S_{[i-1]} \cup S_i^{\leqslant j-1}|} \cdot \left| \sum_{t \in Q} \sum_{\substack{H \in \mathcal{H}_d(\{i,j\}) \\ S_{[i-1]} \cup S_i^{\leqslant j-1} \subseteq H}} \chi_{(H \setminus (S_{[i-1]} \cup S_i^{\leqslant j-1}))(t)}(G) \right| \leqslant O\left(p^{-|S_{[\ell]}|} |Q| n^{-\varepsilon(d-\ell)/4}\right) ,$$

and, on the other hand, if $j \geqslant d+3$, then the above sum is empty.

Proof. Let us first consider the case $j \geqslant d+3$. For all such j it holds that all graphs H we sum over contain the edges $S_{[i-1]} \cup S_i^{\leqslant d+2}$. As vertex i has degree at least $d+1$ and $\mathrm{vc}(H) = d$, according to Claim 7.6, it holds that i is contained in any minimum vertex cover of H. This implies that there are no graphs H satisfying $S_{[i-1]} \cup S_i^{\leqslant d+2} \subseteq H$ that are also in the $(d, \{i, j\})$-boundary. Thus the considered sum is empty for $j \geqslant d+3$.

We now assume $j \leqslant d+2$. We can also assume $j > i$ as otherwise the edge $\{i, j\}$ is in any graph we consider and hence cannot be a boundary edge. Recall that $i \leqslant \ell$, and let $E_0 = S_{[i-1]} \cup S_i^{\leqslant j-1}$, $E_1 = S_{[\ell]} \setminus E_0$ and $B = \{\ell+1, \dots k\}$. Similar to the proof of Lemma 7.3, though this time we rely on Proposition 4.3, we may rewrite the boundary sum as

$$\sum_{t \in Q} \sum_{\substack{H \in \mathcal{H}_d(\{i,j\}) \\ E_0 \subseteq H}} \chi_{(H \setminus E_0)(t)}(G) = \sum_{t \in Q} \sum_{\substack{F \in \mathrm{img(core)} \\ F \in \mathcal{H}_d(\{i,j\})}} \sum_{\substack{H \in \mathcal{H}(F, E_F^\star) \\ E_0 \subseteq H}} \chi_{(H \setminus E_0)(t)}(G) \tag{7.27}$$

$$= \sum_{\substack{F \in \mathrm{img(core)} \\ F \in \mathcal{H}_d(\{i,j\}) \\ E_0 \subseteq F \cup E_F^\star}} \sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star \setminus E_0)} \chi_{(H \setminus E_0)(t)}(G) \ . \tag{7.28}$$

We now focus on the inner two sums. Since $S_{[\ell]} = E_0 \mathbin{\dot\cup} E_1$ we have that $\chi_{(H \setminus E_0)(t)}(G) = \chi_{(H \cap E_1)(t)}(G) \cdot \chi_{(H \setminus S_{[\ell]})(t)}(G)$. By definition of a good rectangle, for all $t \in Q$, the edges $S_{[\ell]}(t)$ are present in G thus $\chi_{(H \cap E_1)(t)}(G) = ((1-p)/p)^{|H \cap E_1|}$. Let $T = |E_F^\star \cap E_1|$. Using that $H \setminus S_{[\ell]} = H[B]$ and $E_F^\star \setminus S_{[\ell]} = E_F^\star[B]$ we can derive that

$$\sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star \setminus E_0)} \chi_{(H \setminus E_0)(t)}(G) = \sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star \setminus E_0)} \chi_{(H \cap E_1)(t)}(G) \cdot \chi_{(H \setminus S_{[\ell]})(t)}(G) \tag{7.29}$$

$$= \sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star \setminus S_{[\ell]})} \sum_{\nu=0}^{T} \binom{T}{\nu} \left(\frac{1-p}{p}\right)^{\nu} \chi_{H \setminus S_{[\ell]}(t)}(G) \tag{7.30}$$

$$= \sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star[B])} \chi_{H[B](t)}(G) \sum_{\nu=0}^{T} \binom{T}{\nu} \left(\frac{1-p}{p}\right)^{\nu} \tag{7.31}$$

$$= p^{-T} \sum_{t \in Q} \sum_{H \in \mathcal{H}(F, E_F^\star[B])} \chi_{H[B](t)}(G) \ . \tag{7.32}$$

Appealing to the triangle inequality we may thus bound

$$\Big|\sum_{t\in Q}\sum_{\substack{H\in\mathcal{H}_d(\{i,j\})\\ E_0\subseteq H}}\chi_{(H\setminus E_0)(t)}(G)\Big| = \Big|\sum_{\substack{F\in \mathrm{img}(\mathrm{core})\\ F\in\mathcal{H}_d(\{i,j\})\\ E_0\subseteq F\cup E_F^\star}} p^{-|E_F^\star\cap E_1|}\sum_{t\in Q}\sum_{H\in\mathcal{H}(F,E_F^\star[B])}\chi_{H[B](t)}(G)\Big| \tag{7.33}$$

$$\leqslant \sum_{\substack{F\in \mathrm{img}(\mathrm{core})\\ F\in\mathcal{H}_d(\{i,j\})\\ E_0\subseteq F\cup E_F^\star}} p^{-|E_F^\star\cap E_1|}\Big|\sum_{t\in Q}\sum_{H\in\mathcal{H}(F,E_F^\star[B])}\chi_{H[B](t)}(G)\Big| \,. \tag{7.34}$$

Fix F and let $w_B = \mathrm{vc}(F[B])$. Note that the characters we sum in Equation (7.34) solely depend on edges between blocks of size at least s. We may thus apply Lemma 7.5 to the inner expressions and, using that $p^{-|E(F[B])|} \leqslant p^{-3\cdot \mathrm{vc}(F)\cdot \mathrm{vc}(F[B])} = p^{-3dw_B} \leqslant n^{6\eta w_B}$ and $s \geqslant 10\,k^2 n^{48\eta+\varepsilon}\log n$, we obtain

$$\Big|\sum_{t\in Q}\sum_{H\in\mathcal{H}(F,E_F^\star[B])}\chi_{H[B](t)}(G)\Big| \leqslant h(w_B) = O\big(|Q|\cdot(n^{24\eta+\varepsilon}k)^{-w_B/4}\big)\ . \tag{7.35}$$

From Lemma 7.8 we know that there are at most $2^{3(d-(i-1))(d+\log k)}$ many cores we sum over. The problem is that this bound does not depend on w_B and hence applying Lemma 7.8 to Equation (7.34) does not readily result in the desired bound.

In order to bound the number of cores, we partition them according to the size w_B of their minimum vertex cover. Note that by Claim 7.6, if $S_{[i-1]} \subseteq E_0 \subseteq F \cup E_F^\star$ it must be the case $S_{[i-1]} \subseteq F$ by the definition of core graphs. Therefore, for such graphs F, we have that $\mathrm{vc}(F[B]) = \mathrm{vc}(F\setminus S_{[\ell]}) \leqslant \mathrm{vc}(F\setminus S_{[i-1]}) = d-(i-1)$, where the last equality follows from Claim 7.7. Using that $|S_{[\ell]}| = |E_0| + |E_1\cap E_F^\star| + |E_1\setminus E_F^\star|$ since $S_{[\ell]} = E_0 \mathbin{\dot\cup} E_1$, we can apply Claim 7.9 to obtain that

$$p^{-|E_0|}\cdot\Big|\sum_{t\in Q}\sum_{\substack{H\in\mathcal{H}_d(\{i,j\})\\ E_0\subseteq H}}\chi_{(H\setminus E_0)(t)}(G)\Big| \leqslant p^{-|E_0|}\cdot\sum_{\substack{F\in \mathrm{img}(\mathrm{core})\\ F\in\mathcal{H}_d(\{i,j\})\\ E_0\subseteq F\cup E_F^\star}} p^{-|E_F^\star\cap E_1|}\cdot h(\mathrm{vc}(F[B])) \tag{7.36}$$

$$= p^{-|S_{[\ell]}|}\sum_{w_B=d-\ell}^{d-(i-1)} h(w_B)\cdot \sum_{\substack{F\in \mathrm{img}(\mathrm{core})\\ F\in\mathcal{H}_d(\{i,j\})\\ E_0\subseteq F\cup E_F^\star\\ \mathrm{vc}(F[B])=w_B}} p^{|E_1\setminus E_F^\star|} \tag{7.37}$$

$$\leqslant p^{-|S_{[\ell]}|}3d^2\binom{k}{d}^3 2^{-k+5d}\sum_{w_B=d-\ell}^{d-(i-1)} 2^{3dw_B}\cdot h(w_B)\ , \tag{7.38}$$

where $h(w_B) = O\big(|Q|\cdot(n^{24\eta+\varepsilon}k)^{-w_B/4}\big)$, we use that $\ell < d$ and, in order to apply Claim 7.9, we use that $p \leqslant 1/2$. Note that, as $d \leqslant \eta D \leqslant \eta k < k/50$, we have that $\binom{k}{d}^3 \leqslant \left(\frac{4k}{d}\right)^{3d} < 200^{3k/50} < 2^{k/2}$. This allows us to bound $3d^2\binom{k}{d}^3 2^{-k+5d} = O(1)$.

Finally, because $2^{3d} \leqslant n^{24\eta/4}$ and the sum in (7.38) is a geometric series with common ratio $2^{3d}\cdot(n^{24\eta+\varepsilon}k)^{-1/4} \leqslant (n^\varepsilon k)^{-1/4}$ and coefficient $O\big(|Q|\cdot(n^\varepsilon k)^{(d-\ell)/4}\big)$ we may conclude that

$$p^{-|E_0|}\cdot\Big|\sum_{t\in Q}\sum_{\substack{H\in\mathcal{H}_d(\{i,j\})\\ E_0\subseteq H}}\chi_{(H\setminus E_0)(t)}(G)\Big| \leqslant O\big(p^{-|S_{[\ell]}|}\,|Q|\,n^{-\varepsilon(d-\ell)/4}\big)\ , \tag{7.39}$$

as claimed. □

7.2 Bounds for All Good Rectangles

This section is devoted to the proof of Lemma 7.5. We rely on the following lemma.

Lemma 7.10. Let $a, b, c, f \geqslant 1$ be integers such that $b > 12\ln(4a)$ and let γ be a positive real number such that $(3a\ln(4af)/c)^{1/2} < \gamma < 1$. Let U be a set satisfying $|U| = a \cdot b$ and $\mathcal{F} \subseteq 2^U$ be a family of subsets over U of size $|\mathcal{F}| = f$, where each $F \in \mathcal{F}$ is of size at least c. Then there is a partition of $U = \bigcup_{i=1}^{a} U_i$ such that

1. $b/2 \leqslant |U_i| \leqslant 3b/2$ for all $1 \leqslant i \leqslant a$, and
2. for each set $F \in \mathcal{F}$ and $1 \leqslant i \leqslant a$ it holds that $(1-\gamma)|F|/a \leqslant |F \cap U_i| \leqslant (1+\gamma)|F|/a$.

Proof. We use the probabilistic method to show that a partition as claimed exists. Independently color each element in U by a color in $\{1, 2, \ldots, a\}$. Let U_i denote color class i.

Let us argue Item 1. By the multiplicative Chernoff bound we have that for a single $i \in [a]$, Item 1 holds except with probability $2 \cdot \exp(-b/12)$. A union bound over the a sets establishes that except with probability $2a \cdot \exp(-b/12) < 1/2$ Item 1 holds.

To argue Item 2 we again appeal to the multiplicative Chernoff bound to see that for a fixed F and i the property holds except with probability $2 \cdot \exp(-\gamma^2|F|/3a) \leqslant 2 \cdot \exp(-\gamma^2 c/3a) < 1/2af$. A union bound over the family $\mathcal{F}$ and $i \in [a]$ shows that Item 2 holds except with probability less than $1/2$.

A final union bound shows the existence of a partition as claimed. □

As a corollary we can show that any good rectangle can be partitioned into many small good rectangles with only a slight loss in parameters.

Corollary 7.11. Let $n, k \in \mathbb{N}$ and $s, \eta, D, d \in \mathbb{R}^+$ be such that $d \leqslant \eta D$, $k \leqslant n$ and $s = 9k^2n^{2\eta}d\ln n$, let $p = n^{-2/D}$ and $R \subseteq [k]$. If Q is a $(s, 1/k, p, d, R)$-good rectangle and $T \subseteq [k] \setminus R$, then there is a partition of Q into a set $\mathcal{Q}$ of at most $(n/s)^{|T|}$ rectangles such that each $Q' \in \mathcal{Q}$ is $(s, 3/k, p, d, R)$-good and satisfies $|Q'_i| \leqslant 4s$ for all $i \in T$.

Proof. This is a direct consequence of Lemma 7.10: fix $i \in T$ such that $|Q_i| > 4s$. Apply Lemma 7.10 with $U = Q_i$, $b = 2s$, $a = |Q_i|/b$, $\gamma = 1/k$ and $\mathcal{F}$ being the common neighborhoods in Q_i of all tuples of size at most d. Thus we get the bound $f \leqslant \binom{nk}{\leqslant d} \leqslant (nk)^d$ and that every set in $\mathcal{F}$ is of size at least $c = (1 - 1/k)p^d|Q_i|$. The bound on s in the lemma statement has been chosen so that the conditions of Lemma 7.10 are satisfied.

Thus we obtain a partition $\mathcal{Q}_i$ of Q_i such that each set $Q'_i \in \mathcal{Q}_i$ satisfies $s \leqslant |Q'_i| \leqslant 3s$ as required. Furthermore, each small tuple t has a common neighborhood of size $(1 \pm 1/k)^2 p^{|t|}|Q'_i|$ in the set Q'_i. Here we used Property 3 of Definition 6.6 which states that Q has $(1/k, p)$-bounded common neighborhoods into Q_i. Because the interval $[(1-1/k)^2, (1+1/k)^2]$ is contained in $[1 - 3/k, 1 + 3/k]$ we have that each tuple t has a $(3/k, p)$-bounded common neighborhood in Q_i. Applying Lemma 7.10 iteratively in the above manner to each large block $i \in T$ gives us the family $\mathcal{Q}$ in the statement. □

We are now ready to prove Lemma 7.5 restated here for convenience.

Lemma 7.5 (Restated). Let $n, k \in \mathbb{N}$ and $s, \eta, D, d \in \mathbb{R}^+$ be such that $d \leqslant \eta D$, $k \leqslant n$ and $s \geqslant 9k^2 n^{2\eta} d \ln n$. If $p = n^{-2/D}$, G is a D-well-behaved graph and F is a core graph, then all $(s, 1/k, p, d, R)$-good rectangles Q satisfy, for $B = [k] \setminus R$, that

$$\Big|\sum_{t\in Q}\sum_{H\in\mathcal{H}(F,E_F^\star[B])}\chi_{H[B](t)}(G)\Big| = O\big(|Q|\cdot p^{-|E(F[B])|}\cdot(s/10\,k\log n)^{-\operatorname{vc}(F[B])/4}\big)\ .$$

Proof. We first apply Corollary 7.11 to an $(s, 1/k, p, d, R)$-good rectangle Q with $T = [k] \setminus R$ to obtain a family $\mathcal{Q}$ of $(s, 3/k, p, d, R)$-good rectangles with the additional property that each block is bounded in size by $4s$. We can therefore write

$$\Big|\sum_{t\in Q}\sum_{H\in\mathcal{H}(F,E_F^\star[B])}\chi_{H[B](t)}(G)\Big| \leqslant \sum_{\tilde{Q}\in\mathcal{Q}}\Big|\sum_{t\in\tilde{Q}}\sum_{H\in\mathcal{H}(F,E_F^\star[B])}\chi_{H[B](t)}(G)\Big| \tag{7.40}$$

$$\leqslant \sum_{\tilde{Q}\in\mathcal{Q}} O\big(|\tilde{Q}|\cdot p^{-|E(F[B])|}\cdot(s/10\,k\log n)^{-\operatorname{vc}(F[B])/4}\big) \tag{7.41}$$

$$= O\big(|Q|\cdot p^{-|E(F[B])|}\cdot(s/10\,k\log n)^{-\operatorname{vc}(F[B])/4}\big)\ , \tag{7.42}$$

where (7.41) follows directly from Property 4 of Definition 5.4 by noting that every $\tilde{Q} \in \mathcal{Q}$ is an $(s, 3/k, p, d, R)$-good rectangles such that $|\tilde{Q}_i| \leqslant 4s$ for all $i \in [k]$, and that every $|B|$-tuple $t_B \in \tilde{Q}_B$ has a unique k-tuple extension $t \supseteq t_B$ in $\tilde{Q}$. □

8 Random Graphs Are Well-Behaved

In this section we prove that graphs sampled from $\mathcal{G}(n, k, n^{-2/D})$ are asymptotically almost surely D-well-behaved. For convenience we recall the precise statement below.

Theorem 5.5 (Restated). If n is a large enough integer, $k \in \mathbb{N}^+$ and $D \in \mathbb{R}^+$ satisfy $4 \leqslant D \leqslant 2\log n$ and $k \leqslant n^{1/5}$, then $G \sim \mathcal{G}(n, k, n^{-2/D})$ is asymptotically almost surely D-well-behaved.

Recall that Definition 5.4 consists of four properties: Property 1 states that the common neighborhood of small tuples behave as expected. When we focus our attention on a subgraph induced by a rectangle, we cannot expect that the common neighborhoods in this rectangle still behave as expected—we may for example have an isolated vertex in a rectangle. Property 2 of Definition 5.4 guarantees thus a slightly weaker property: it roughly states that every rectangle has only few small tuples whose common neighborhood is ill-behaved.

While the first two properties of Definition 5.4 are of combinatorial nature, the final two properties are more of analytical nature. Property 3 states that character sums over families $\mathcal{H}(F, E_F^\star)$ are bounded for rectangles with large enough minimum block size. Once we consider smaller rectangles, things are not so well-behaved anymore. However, for certain rectangles, we can still guarantee a similar property: Property 4 of Definition 5.4 states that if the common neighborhoods of small tuples in a rectangle are well-behaved, then we can still guarantee that the mentioned character sums are bounded. In fact, this bound is in some sense tighter than the one in Property 3 as it actually depends on the size of the considered rectangle.

In Section 8.1 we start by proving that Properties 1 and 2 of Definition 5.4 hold asymptotically almost surely. As previously mentioned, similar notions have been used in previous papers on the clique formula [BIS07, BGLR12, ABdR+21]. Our arguments towards Properties 1 and 2 are heavily influenced by these papers, and some are essentially identical.

Once we established the combinatorial properties, in Section 8.2, we then prove the analytical Properties 3 and 4 of Definition 5.4, modulo a probabilistic bound on the weighted sums of Fourier characters. Section 8.3 is dedicated to the proof of the probabilistic bound which, inspired by [AMP21], follows via an encoding argument applied to a high moment version of the Markov inequality.

8.1 Random Graphs Have Bounded Common Neighborhoods

In this section we prove that asymptotically almost surely $G \sim \mathcal{G}(n, k, n^{-2/D})$ satisfy Properties 1 and 2 of Definition 5.4.

Lemma 8.1. For any constant $\delta > 0$, any integer $D, k, n \in \mathbb{N}$ satisfying $k \leqslant n$ and $4 \leqslant D \leqslant n^{\delta/2}$ the following holds for $\beta \geqslant n^{-1/4+\delta/2}$. Except with probability $\exp(-\Omega(n^{\delta}))$ a graph $G \sim \mathcal{G}(n, k, p)$ has $(\beta, p, D/4)$-bounded common neighborhoods in every block, for $p = n^{-2/D}$.

Proof. Given a set of vertices S, let X_v be the indicator random variable of whether a vertex v is in the common neighborhood of S. Note that $\Pr[X_v = 1] = p^{|S|}$. Thus in expectation we have $np^{|S|}$ many common neighbors per block. Applying the multiplicative Chernoff bound we get that

$$\Pr\Big[\Big|\sum_{i=1}^{n} X_i - n \cdot p^{|S|}\Big| \geqslant \beta n \cdot p^{|S|}\Big] \leqslant 2 \cdot \exp\left(-\frac{\beta^2 n}{3 \cdot p^{-|S|}}\right). \tag{8.1}$$

Taking a union bound over all blocks and all sets S of size less than or equal to $D/4$ we get

$$\Pr[\exists S, V_i \text{ such that } N(S) \cap V_i \text{ has uncommon size}] \leqslant k \cdot \sum_{i=1}^{D/4} \binom{kn}{i} 2 \cdot \exp\left(-\frac{\beta^2 n}{3 \cdot p^{-i}}\right) \tag{8.2}$$

$$\leqslant 2k \cdot \sum_{i=1}^{D/4} (kn)^i \cdot \exp\left(-\frac{n^{1/2+\delta}}{3 \cdot n^{2i/D}}\right) \quad , \tag{8.3}$$

which is exponentially small in $\Omega(n^{\delta})$. ☐

While all small sets have common neighborhoods of expected size in any block, this is not necessarily true for subsets of a block. For example, if $S \subseteq V_i$ consists of all non-neighbors of a vertex v, then clearly v does not have a neighborhood in S of expected size. We can, nonetheless, show that for any large enough set S there is a small set of vertices W, which we refer to as the *error set* of S, such that all small sets of vertices not intersection W have a common neighborhoods in S of expected size. The following lemma works out the precise dependency between the size of W and the size of S.

Lemma 8.2. For all $k, n \in \mathbb{N}^+$, $k \leqslant n$, $\gamma > 0$, and $0 < p \leqslant 1/2$ the following holds asymptotically almost surely for $G \sim \mathcal{G}(n, k, p)$. For all $\ell \in [k]$, for all $i \in [k]$ and for every subset $S \subseteq V_i$ of size $s \geqslant 2w$, for $w = 12\ell \ln n / p^{\ell}\gamma^2$, there is a set $W_{S,\ell}$ of size at most w such that every tuple t, $|t| \leqslant \ell$, disjoint from V_i and $W_{S,\ell}$ has a common neighborhood size in S of expected size, that is, it holds that

$$|N^{\cap}(t,S)| = (1 \pm \gamma) \mathop{\mathbb{E}}_{G}[|N^{\cap}(t,S)|] = (1 \pm \gamma) p^{|t|} s \ .$$

Proof. Given $\ell \in [k]$ and $S \subseteq V_i$, for some $i \in [k]$, we define a set $W_{S,\ell}$ such that for every tuple t with $|t| \leqslant \ell$ and that is disjoint from V_i and $W_{S,\ell}$ it holds that

$$|N^{\cap}(t,S)| = (1 \pm \gamma) p^{|t|} s \ . \tag{8.4}$$

Our goal is then to show that the probability of the event that there exists an $\ell \in [k]$, $i \in [k]$ and a set $S \subseteq V_i$ of size at least $2w$ such that $|W_{S,\ell}| > w$ is at most $n^{-\Omega(1)}$.

The set $W_{S,\ell}$ is constructed by including all tuples of that do not satisfy Equation (8.4). More formally, we first consider each possible tuple size $a = 1, 2, \ldots, \ell$ and separately construct sets W_a as follows. Let $t_1, t_2, \ldots, t_{w_a}$ be an arbitrary but fixed maximal sequence of pairwise disjoint tuples, each of size a, where each tuple either has too many or too few common neighbors in S, that is, $|N^{\cap}(t_j, S)| > (1+\gamma)p^a s$ or $|N^{\cap}(t_j, S)| < (1-\gamma)p^a s$ for all $j \in [w_a]$. We define $W_a = \bigcup_{j \in [w_a]} t_j$ and $W_{S,\ell} = \bigcup_{a \in [\ell]} W_a$.

Let $\widehat{w}_a = 6 \ln n / p^a \gamma^2$. Note that if $w_a \leqslant \widehat{w}_a$ for all $a \in [\ell]$, then $|W_a| \leqslant a \cdot \widehat{w}_a$ and

$$|W| \leqslant \frac{6 \log n}{\gamma^2} \sum_{a \in [\ell]} \frac{a}{p^a} \leqslant \frac{6 \log n}{\gamma^2} \left(\frac{2\ell}{p^\ell} \right) = w \ , \tag{8.5}$$

where we use that $p \leqslant 1/2$. We can therefore bound the probability that $|W_{S,\ell}| > w$ by bounding the probability that there exists an $a \in [\ell]$ such that $w_a \geqslant \widehat{w}_a$.

Fix an $a \in [\ell]$. The probability that a given tuple of t_j of size a has too many or too few common neighbors in S can be bounded by the multiplicative Chernoff bound to obtain that

$$\Pr\Big[\big||N^{\cap}(t_j, S)| - p^a |S|\big| > \gamma \cdot p^a |S|\Big] \leqslant 2 \cdot \exp(-\gamma^2 \cdot p^a \cdot |S|/3) \ . \tag{8.6}$$

Note that the presence of edges between disjoint tuples t_j and S are independent events. Thus we can bound the probability that there there are at least $\widehat{w}_a$ many disjoint tuples that have too many or too few common neighbors in S by

$$2 \cdot \exp\big(-\gamma^2 \cdot p^a \cdot s \cdot \widehat{w}_a / 3\big) = 2 \cdot \exp(-2s \ln n) \ . \tag{8.7}$$

By taking a union bound over $\ell \in [k]$, $s = 2w, \ldots, n$ and $a \in [\ell]$, and then for each s and a taking a union bound over $i \in [k]$, over the choices of $S \subseteq V_i$ of size s, and over the choices for the $\widehat{w}_a$ tuples (of which there are at most $\binom{kn}{a}^{\widehat{w}_a} \leqslant (kn)^{a\widehat{w}_a}$), we conclude that the probability that there exists an ℓ and a set S of size at least $2w$ such that $W_{S,\ell} > w$ is at most

$$\Pr\big[\exists S, \ell : |W_{S,\ell}| > w\big] \leqslant \sum_{\ell \in [k]} \sum_{s=2w}^{n} \sum_{a \in [\ell]} 2 \cdot \exp(-2s \ln n + a\widehat{w}_a \ln(kn) + \ln k + s \ln n) \tag{8.8}$$

$$\leqslant \sum_{\ell \in [k]} \sum_{s=2w}^{n} 2 \cdot \exp(-(s-1)\ln n + \ell \widehat{w}_\ell (2 \ln n) + \ln \ell) \tag{8.9}$$

$$\leqslant \sum_{\ell \in [k]} \sum_{s=2w}^{n} 2 \cdot \exp(-(s-2-w)\ln n) \leqslant n^{-\Omega(1)} \ , \tag{8.10}$$

as we wished to prove. ◻

Given a large rectangle Q we would like to argue that after removing the union of all error sets $W = \bigcup_i W_{Q_i}$ we are left with a rectangle in which all small tuples have common neighborhoods of expected size. To this end we rely on the following claim.

Claim 8.3. Let $b > 0$ and $0 < \gamma < 1$. Suppose we have a universe U and a set $S \subseteq U$ satisfying $\frac{|S|}{|U|} \in (1 \pm \gamma)b$. For any set $T \subseteq U$ satisfying $\frac{|T|}{|U|} \leqslant \min\{\gamma/2, b\gamma\}$ it holds that

$$\frac{|S \setminus T|}{|U \setminus T|} \in (1 \pm 3\gamma)b \ .$$

Proof. Let us denote the size of S by s, the size of T by t and the size of U by u. It holds that

$$\frac{|S \setminus T|}{|U \setminus T|} \leqslant \frac{s}{u(1-\gamma/2)} \leqslant (1+\gamma)^2 b \leqslant (1+3\gamma)b \ . \tag{8.11}$$

For the lower bound observe that it holds

$$\frac{|S \setminus T|}{|U \setminus T|} \geqslant \frac{s}{u-t} - \frac{t}{u-t} \geqslant \frac{s}{u} - \frac{t}{u(1-\gamma/2)} \geqslant (1-\gamma)b - \frac{t}{u}(1+\gamma) \geqslant (1-3\gamma)b \ , \tag{8.12}$$

where, for the final inequality, we used that $t/u \leqslant b\gamma$. □

We are now ready to prove that, asymptotically almost surely, $G \sim \mathcal{G}(n,k,p)$ for $p = n^{-2/D}$, satisfies Property 2 of Definition 5.4 which states that for all $\ell \leqslant D/4$, and $s \geqslant Ck^4\ell \ln n/p^{2\ell}$, for a large enough constant C, the graph G has $(2s, s, 1/k, p, \ell)$-bounded error sets, i.e., that for all rectangles $Q = \bigtimes_{i\in[k]} Q_i$ satisfying $|Q_i| \geqslant 2s$ or $|Q_i| = 0$ it holds that there exists a small set W, $|W| \leqslant s$, such that for all $S \subseteq [k]$ of size at most ℓ it holds that all tuples $t \in \bigtimes_{i\in S}(Q_i \setminus W)$ satisfy $|N^\cap(t, Q_j \setminus W)| = (1 \pm 1/k)p^{|t|}|Q_j \setminus W|$ for all $j \in [k] \setminus S$.

Corollary 8.4. There exists a constant $C \in \mathbb{R}^+$ such that the following holds for all $D \in \mathbb{R}^+$ and integers $k, n, s \in \mathbb{N}^+$ satisfying $k \leqslant n$ and $p = n^{-2/D} \leqslant 1/2$. A graph G sampled from $\mathcal{G}(n,k,p)$ has asymptotically almost surely $(2s, s, 1/k, p, \ell)$-bounded error sets for all $1 \leqslant \ell \leqslant D/4$ and $s \geqslant Ck^4\ell \ln n/p^{2\ell}$.

Proof. Let $T \subseteq [k]$ and $Q = \bigtimes_{i\in[k]} Q_i$ be such that $|Q_i| \geqslant 2s$ if $i \in T$ and $|Q_i| = 0$ otherwise. Our goal is to show that there exists a small set of vertices $W \subseteq V(G)$, $|W| \leqslant s$, such that for all $T' \subseteq T$ of size at most ℓ it holds that all tuples $t \in \bigtimes_{i\in T'}(Q_i \setminus W)$ satisfy

$$\left|N^\cap(t, Q_j \setminus W)\right| \in (1 \pm 1/k)p^{|t|}\left|Q_j \setminus W\right|$$

for all $j \in [k] \setminus T'$.

Fix $\ell \leqslant D/4$, and the parameters $\gamma = 1/3k$ and $w = 12\ell \ln n/p^\ell\gamma^2$. Note that by choosing $C \geqslant 3^3 \cdot 12$ it holds that $w \leqslant s\gamma p^\ell/k$. For each $i \in T$, let W_i be the set, guaranteed to exist (asymptotically almost surely) by Lemma 8.2, of size at most w such that for all $T' \subseteq T \setminus i$ of size at most ℓ it holds that all tuples $t \in \bigtimes_{j\in T'}(Q_j \setminus W_i)$ satisfy

$$|N^\cap(t, Q_i)| = (1 \pm \gamma)p^{|t|}|Q_i| \ . \tag{8.13}$$

Note that to apply Lemma 8.2 we use the fact that $|Q_i| \geqslant 2w$ for all $i \in T$, which follows since $|Q_i| \geqslant 2s$ and $w \leqslant s\gamma p^\ell/k \leqslant s$.

Let $W = \cup_{i \in T} W_i$. Since $w \leqslant s\gamma p^\ell/k$, we have that $|W| \leqslant kw \leqslant s\gamma p^\ell$. Moreover, for all $T' \subseteq T$ of size at most ℓ and all $j \in [k] \setminus T'$, it holds that all tuples $t \in \bigtimes_{i \in T'}(Q_i \setminus W)$ satisfy

$$|N^\cap(t, Q_j \setminus W)| = (1 \pm 3\gamma)p^{|t|}|Q_j \setminus W| \ , \tag{8.14}$$

by Claim 8.3 where we use the bound $\frac{|W|}{|Q_j|} \leqslant \frac{|W|}{2s} \leqslant \frac{\gamma p^\ell}{2} \leqslant \min\{\gamma/2, \gamma p^\ell\}$. Since $1/k = 3\gamma$ and $|W| \leqslant s\gamma p^\ell \leqslant s$, the statement follows. □

8.2 Random Graphs Have Bounded Character Sums

In order to show that random graphs are asymptotically almost surely well-behaved, we rely on the following probabilistic bound on the weighted sums of Fourier characters, inspired by [AMP21].

Lemma 8.5. Let F be a non-empty graph over the vertex set $[k]$ with $A = V(E(F))$, and let $M \subseteq E(F)$ be any matching in F. Let $Q = \bigtimes_{u \in A} Q_u$ be a rectangle such that for every edge $\{u, v\} \in M$ it holds that $|Q_u \times Q_v| \geqslant \kappa$. For any even $m \leqslant \kappa$, any $r \in \mathbb{R}^+$ and any function $\xi : Q \to [-r, r]$ it holds that

$$\Pr_{G[V_A]}\left[\Big|\sum_{t \in Q} \chi_{F(t)}(G)\xi(t)\Big| > s\right] \leqslant \left(\frac{r \cdot p^{-|E(F)|} \cdot \left(\frac{m}{\kappa}\right)^{|M|/2} \cdot |Q|}{s}\right)^m ,$$

where each edge in $\binom{V_A}{2}$ is sampled independently at random with probability p.

It may be illustrative to consider the case when $F = \{i, j\}$ is a single edge, ξ is the constant 1 function, $|Q_i \times Q_j| = \kappa = |Q| = n$ and $p = 1/2$. In this setting we obtain a bound on the absolute value of n random variables taking values in ± 1 uniformly at random. For some setting of parameters we essentially recover the Chernoff bound: we may, for example, set $s = \sqrt{n \log n}$ and $m = \frac{\log n}{16}$ to obtain the bound

$$\Pr\Big[\Big|\sum_{e \in [n]} \chi_e\Big| > \sqrt{n \log n}\,\Big] \leqslant 2^{-\log(n)/16} , \tag{8.15}$$

which is essentially the bound that Chernoff guarantees.

The proof of Lemma 8.5 goes through a high moment version of the Markov inequality. After some standard manipulations of the resulting expressions we are left to bound the number of ways to map m copies of the graph F into Q such that every 2-tuple in Q is either mapped to at at least twice or not at all. We bound this quantity via an encoding argument.

The main idea of the encoding argument is to only consider the edges in the matching M. Each such edge can be mapped in at least κ ways. As $m \leqslant \kappa$, we argue that it is beneficial to point to the other copy of F that maps this edge to the same place. This drives the upper bound on the failure probability. The details of the argument are worked out in Section 8.3.

With this lemma in hand we are ready to prove the main result of this section.

Proof of Theorem 5.5. Properties 1 and 2 hold asymptotically almost surely by Lemma 8.1 and Corollary 8.4, respectively.

For Properties 3 and 4 we will apply Lemma 8.5 to bound the probability that for a fixed core graph F and rectangle Q that satisfy the conditions stated G does not have bounded character sums over Q for F. We will then apply a union bound over all such F's and Q's.

We start by defining a function $\xi_{F,Q,G,B}$ for every core F, rectangle Q, graph G and subset $B \subseteq [k]$ that maps tuples $t_A \in Q_A$, for $A = V(E(F)) \cap B$, to real numbers as

$$\xi_{F,Q,G,B}(t_A) = \sum_{t \in Q_B : t \supseteq t_A} \sum_{E \subseteq E_F^\star[B]} \chi_{E(t)}(G) = p^{-|E_F^\star[B]|} \cdot \left|\left\{t \in Q_B : t \supseteq t_A \wedge E_F^\star[B](t) \subseteq E(G)\right\}\right| . \tag{8.16}$$

Observe that

$$\sum_{t \in Q_B} \sum_{H \in \mathcal{H}(F, E_F^\star[B])} \chi_{H[B](t)}(G) = \sum_{t \in Q_B} \chi_{F[B](t)}(G) \sum_{E \subseteq E_F^\star[B]} \chi_{E(t)}(G) \tag{8.17}$$

$$= \sum_{t_A \in Q_A} \chi_{F[B](t_A)}(G)\, \xi_{F,Q,G,B}(t_A) . \tag{8.18}$$

Note that $\xi_{F,Q,G,B}$ only depends on F, Q and $G[B] \setminus G[V_A]$, that is, it does not depend on the edges between vertices in V_A. This will be crucial for our analysis as it implies that once we fix B, F, Q and $G_{\bar{A}} = G \setminus G[V_A]$ we can determine whether $\xi_{F,Q,G,B}$ is r-bounded.

For a fixed r, let $\mathbf{X}_r(F, Q, G, B)$ be the event that $\xi_{F,Q,G,B}$ is r-bounded. By Lemma 8.5, we claim that for any s if Q_B is (κ, s)-large, then for any $m \leqslant \kappa$ it holds that

$$\Pr_G\left[\left|\sum_{t \in Q_B} \sum_{H \in \mathbf{H}(F, E_F^\star[B])} \chi_{H[B](t)}(G)\right| > s \;\middle|\; \mathbf{X}_r(F, Q, G, B)\right] \leqslant \left(\frac{r\, p^{-|E(F[B])|} \left(\frac{m}{\kappa}\right)^{\mathrm{vc}(F[B])/4} |Q_A|}{s}\right)^m . \tag{8.19}$$

Indeed, this follows by rewriting the left hand-side according to Equation (8.17) to obtain

$$\begin{aligned}
&\Pr_G\left[\left|\sum_{t \in Q_A} \chi_{F[B](t)}(G)\, \xi_{F,Q,G,B}(t)\right| > s \;\middle|\; \mathbf{X}_r(F, Q, G, B)\right] \\
&\quad = \sum_{\substack{G_{\bar{A}}: \\ \mathbf{X}_r(F,Q,G,B)}} \Pr_{G \setminus G[V_A]}\left[G_{\bar{A}} \mid \mathbf{X}_r(F, Q, G, B)\right] \cdot \Pr_{G[V_A]}\left[\left|\sum_{t \in Q_A} \chi_{F[B](t)}(G)\, \xi_{F,Q,G,B}(t)\right| > s\right] ,
\end{aligned} \tag{8.20}$$

and applying Lemma 8.5 for each fixed $G_{\bar{A}}$.

We are now ready to prove that asymptotically almost surely G satisfies Properties 3 and 4. Fix a core F and a rectangle Q.

For Property 3 we may assume that Q is $(n/2, F)$-large. Let $B = [n]$ and $C = [k] \setminus V(E(F))$. By Property 1 G has $(1/k, p, D/4)$-bounded common neighborhoods in every block hence $\xi_{F,Q,G,B}$ is r-bounded for $r = (1 + 1/k)^{|C|} n^{|C|} \leqslant 3 \cdot n^{|C|}$. Applying Equation (8.19) for this value of r and for $s = 6 \cdot p^{-|E(F)|} \cdot n^{-\lambda\, \mathrm{vc}(F)/4} \cdot n^k$, $\kappa = n^2/4$ and $m = n^{2-\lambda}/4$, we derive that the probability that G does not have s-bounded character sums over Q for F is at most $2^{-m} = 2^{-n^{2-\lambda}/4}$. We conclude Property 3 by taking a union bound over all cores of vertex cover at most $D/4$, of which, according to Lemma 2.4, there are at most $\sum_{i=1}^{D/4} 2^{3i(i+\log k)} \leqslant 2 \cdot 2^{D(D/4+\log k)} \ll 2^{n^{2-\lambda}/4}$, and all $2^{nk} \ll 2^{n^{2-\lambda}/4}$ rectangles, using the fact that $\lambda < 1 - (\log k)/(\log n)$ and $k \leqslant n^{1/3}$.

Property 4 follows by a similar argument. Further fix an integer $\Lambda \in [n]$, $\Lambda \geqslant 20\,k\log n$ and a subset $B \subseteq [k]$, let $A = V\big(E(F)\big) \cap B$ and $C = B \setminus A$. We assume Q is (4Λ)-small and (Λ, F)-large. Note that if G is such that for every $i \in C$, G has $(3/k, p)$-bounded common neighborhoods from Q_A to Q_i, then $\xi_{F,Q,G,B}$ is r-bounded for $r = (1+3/k)^{|C|}\,|Q_C| < 30\,|Q_C|$.

We can therefore apply Equation (8.19) with parameters $r = 30\,|Q_C|$, $\kappa = \Lambda^2$, $m = 10\Lambda k\log n \leqslant \kappa$, and $s = 60\,p^{-|E(F[B])|}(m/\kappa)^{vc(F[B])/4}|Q|$, to obtain that the probability that G does not have s-bounded character sums over Q_B for F is at most $2^{-m} = 2^{-10\Lambda k\log n}$. Applying a union bound over all $\Lambda \in [n]$ such that $\Lambda \geqslant 20\,k\log n$, all cores of vertex cover at most $D/4$, all (4Λ)-small rectangles and all subsets $B \subseteq [k]$ we conclude that the probability that G does not satisfy Property 4 is at most $2^{-\Lambda k\log n}$ since, by Lemma 2.4, there are at most $\sum_{i=1}^{D/4} 2^{3i(i+\log k)} \leqslant 2\cdot 2^{D(D/4+\log k)} \ll 2^{\Lambda k\log n}$ many core graphs, $\binom{n}{\leqslant 4\Lambda}^k \leqslant 2\cdot 2^{4\Lambda k\log n}$ rectangles that are (4Λ)-small and only 2^k subsets B. ◻

8.3 Probabilistic Bound on Sums of Fourier Characters

The rest of this section is dedicated to the proof of Lemma 8.5, restated here for convenience.

Lemma 8.5 (Restated). Let F be a non-empty graph over the vertex set $[k]$ with $A = V\big(E(F)\big)$, and let $M \subseteq E(F)$ be any matching in F. Let $Q = \bigtimes_{u\in A} Q_u$ be a rectangle such that for every edge $\{u,v\} \in M$ it holds that $|Q_u \times Q_v| \geqslant \kappa$. For any even $m \leqslant \kappa$, any $r \in \mathbb{R}^+$ and any function $\xi : Q \to [-r, r]$ it holds that

$$\Pr_{G[V_A]}\left[\Big|\sum_{t\in Q} \chi_{F(t)}(G)\xi(t)\Big| > s\right] \leqslant \left(\frac{r\cdot p^{-|E(F)|}\cdot\left(\frac{m}{\kappa}\right)^{|M|/2}\cdot|Q|}{s}\right)^m ,$$

where each edge in $\binom{V_A}{2}$ is sampled independently at random with probability p.

Proof. In order to bound the probability $\Pr_G\big[|\sum_{t\in Q}\chi_{F(t)}(G)\xi(t)| > s\big]$ we resort to a high moment version of the Markov inequality and conclude that

$$\Pr_{G[V_A]}\left[\Big|\sum_{t\in Q} \chi_{F(t)}(G)\xi(t)\Big| > s\right] \leqslant \frac{\mathbb{E}_{G[V_A]}\Big[\big(\sum_{t\in Q}\chi_{F(t)}(G)\xi(t)\big)^m\Big]}{s^m} , \tag{8.21}$$

for even m. For conciseness, let us write G instead of $G[V_A]$ and rewrite above expectation to obtain that

$$\begin{aligned}
\mathbb{E}_G\Big[\Big(\sum_{t\in Q}\chi_{F(t)}(G)\xi(t)\Big)^m\Big] &= \sum_{t_1,\dots,t_m\in Q} \mathbb{E}_G\Big[\prod_{i\in[m]}\chi_{F(t_i)}(G)\Big]\cdot\prod_{i\in[m]}\xi(t_i) &\quad(8.22)\\
&\leqslant \sum_{t_1,\dots,t_m\in Q} \Big|\mathbb{E}_G\Big[\prod_{i\in[m]}\chi_{F(t_i)}(G)\Big]\Big|\cdot\Big|\prod_{i\in[m]}\xi(t_i)\Big| &\quad(8.23)\\
&\leqslant \max_{t\in Q}|\xi(t)|^m\cdot\sum_{t_1,\dots,t_m\in Q}\Big|\mathbb{E}_G\Big[\prod_{i\in[m]}\chi_{F(t_i)}(G)\Big]\Big| . &\quad(8.24)
\end{aligned}$$

Observe that the magnitude of each expectation in Equation (8.22) is 0 unless each edge appears at least twice in the product, in which case it is at most $(1/p)^{|E(F)|\cdot m}$. Indeed, for $\ell \in \mathbb{N}$ it holds that

$$\left|\mathop{\mathbb{E}}_{G}\left[\chi_{e(t)}(G)^{\ell}\right]\right| \leqslant (1-p)\cdot 1 + p\cdot\frac{(1-p)^{\ell}}{p^{\ell}} \leqslant (1-p)\cdot\frac{1}{p^{\ell}} + p\cdot\frac{1}{p^{\ell}} = \frac{1}{p^{\ell}} \ . \tag{8.25}$$

Therefore, given $t_1,\ldots,t_m \in Q$ such that each edge appears at least twice, we have that

$$\left|\mathop{\mathbb{E}}_{G}\Big[\prod_{i\in[m]}\chi_{F(t_i)}(G)\Big]\right| \leqslant p^{-|E(F)|\cdot m} \ . \tag{8.26}$$

It remains to bound the number of tuples $t_1,\ldots,t_m$ such that each edge appears at least twice. In more detail, we want to count the number of ordered sequences $(t_1,\ldots,t_m)$ such that if F is mapped to each tuple t_i, then in the resulting multi-graph $\mathcal{F} = \cup_{i\in[m]}F(t_i)$ each edge appears at least twice.

To this end, let $\mathcal{T}$ be the family of ordered sequences $(t_i)_{i\in[m]}$ that give rise to such multi-graphs. In what follows we bound the cardinality of $\mathcal{T}$ by an encoding argument: we encode each element of $\mathcal{T}$ with few bits and thus establish that this set is small.

Let us fix one such sequence $(t_i)_{i\in[m]} \in \mathcal{T}$. Encoding each tuple separately is too costly. To minimize the bits needed we use the property that in the resulting multi-graph $\mathcal{F} = \cup_{i\in[m]}F(t_i)$ each edge is covered at least twice—in fact, we only use this property for edges in the matching. We encode the sequence as follows.

- First, for each $i \in [m]$, we go through the edges of the ith copy of M in some fixed order. Suppose we consider an edge $\{u,v\} \in M$ in the ith copy of F. If this edge has not been matched yet, then we write down an integer $0 \leqslant j \leqslant m$ indicating that the tuples t_i and t_j map the vertices u and v to the same place. The edge $\{u,v\}$ of the jth copy of F is then said to be matched. Otherwise, if the edge $\{u,v\}$ of the ith copy of F has already been matched, then we continue with the next edge.
- After completing above step we encode the tuples. We use the knowledge of the previous step: if we indicated that the uth vertex of the tuples t_i and t_j are mapped to the same place, then we encode the target only once. More precisely, we encode the tuples as follows. Iterate over each $i \in [m]$ and $u \in V(E(F))$. If the uth vertex of t_i has already been mapped, then continue with the next vertex. Otherwise, we write down an integer $0 \leqslant v \leqslant |Q_u|$ indicating that the uth vertex of t_i is equal to $v \in Q_u$. If the uth vertex of t_i has matched edges incident, then record the corresponding uth vertices as mapped.

It should be evident that this procedure can be inverted, i.e., from the information written we can uniquely decode the sequence $(t_i)_{i\in[m]}$, provided we know the graph F, the matching M, the rectangle Q and the integer m.

Let us tally the bits used in the above encoding. Suppose that in the first step we matched a edges.

- For each matched edge we wrote down $\log m$ bits for a total of $a\cdot\log m$ bits, and
- at most $m\cdot\sum_{u\in V(E(F))}\log|Q_u| - a\cdot\log\kappa$ many bits in the second step: each matched edge reduces the total number of vertices to be mapped by 2 (using that we only match edges in M). As for every edge $\{u,v\} \in M$ it holds that $|Q_u \times Q_v| \geqslant \kappa$, we see that the number of bits required in the second step of the encoding is reduced by each matched edge by at least $\log\kappa$ bits.

Hence we need at most $m \cdot \sum_{u \in V(E(F))} \log |Q_u| - \min_{\frac{m \cdot |M|}{2} \leqslant a \leqslant m \cdot |M|} a(\log \kappa - \log m)$ many bits to encode such a sequence of tuples. As we assumed that $m \leqslant \kappa$ we conclude that

$$\mathbb{E}_{G}\Big[\Big(\sum_{t \in Q} \chi_{F(t)}(G)\xi(t)\Big)^m\Big] \leqslant \max_{t \in Q} |\xi(t)|^m \cdot p^{-|E(F)| \cdot m} \cdot |\mathcal{T}| \tag{8.27}$$

$$\leqslant \left(a \cdot p^{-|E(F)|} \cdot \left(\frac{m}{\kappa}\right)^{|M|/2} \prod_{u \in V(E(F))} |Q_u|\right)^m . \tag{8.28}$$

Substituting this bound in Equation (8.21) results in the desired statement. □

9 Concluding Remarks

For $k \leqslant n^{1/100}$ we prove an essentially tight average-case $n^{\Omega(D)}$ coefficient size lower bound on Sherali-Adams refutations of the k-clique formula for Erdős-Rényi random graphs with maximum clique of size D. In fact, we obtain a lower bound on the sum of the magnitude of the coefficients appearing in a (general) Sherali-Adams refutation. The obvious problem left open is to prove an $n^{\Omega(D)}$ *monomial* size lower bound on Sherali-Adams refutations of the clique formula.

One possible avenue to prove such a monomial size lower bound is to argue that any Sherali-Adams proof of the clique formula can be converted into a proof of the same monomial size but with small coefficients. As shown in [GHJ+22] this is in general not possible and one would hence have to leverage the structure of the clique formula to argue that such a conversion exists. In fact, a slightly weaker statement would suffice: recall that our lower bound only counts the size of the coefficients of generalized monomials as well as of monomials multiplied by edge axioms. As such we would just need to be able to convert a general Sherali-Adams refutation into a refutation with low coefficients for such monomials.

In contrast to previous lower bounds for clique, our proof strategy is not purely combinatorial. It might be fruitful to obtain an explicit combinatorial description of μ_d—we believe this could potentially be used to prove average-case clique lower bounds for other proof systems, including resolution.

A strength of our lower bound approach is that it is quite oblivious to the encoding: one can introduce all possible extension variables depending on a *single* block and the lower bound argument still goes through. This is because the only property we require of a monomial m is that the set of tuples $Q_m = \{t \mid \rho_t(m) \neq 0\}$ whose associated assignment ρ_t sets m to non-zero is a rectangle. By extending ρ in the natural manner to extension variables it is easy to see that Q_m is still a rectangle.

Our lower bound strategy seems to fail quite spectacularly once the edge probability is increased well beyond $1/2$. More precisely, once $D = \omega(\log n)$, we fail to counter exponential in d^2 factors that arise from encoding the core graphs: as long as $D = O(\log n)$ we can counter these with s^{-d} terms, where s is the minimum block size of a good rectangle. As s is clearly bounded by the block size n, this approach fails once $D = \omega(\log n)$. We leave it as an open problem to extend our result to the dense setting.

We rely on rather unorthodox pseudorandomness properties of the underlying graph. It is natural to wonder whether these properties follow from a previously studied notion of pseudorandomness. Furthermore, it is wide open whether our lower bound can be made explicit.

In particular, we have not investigated whether graphs that satisfy our pseudorandomness property can be constructed deterministically.

Another application of our pseudo-measure μ_d is in communication complexity. Suppose we consider the k-player number-in-hand model, where player i obtains a single node u_i from block V_i. The goal of the k players is to find an edge missing in the induced subgraph by the tuple $(u_1, \dots, u_k)$. Consider the leaves of such a communication protocol. Note that each leaf ℓ is associated with a subrectangle Q_ℓ of an edge axiom. As the family of these associated rectangles Q_ℓ partition the whole space, but $|\mu_d(Q_\ell)| \leqslant n^{-\Omega(D)}$, there must be at least $n^{\Omega(D)}$ leaves.

Finally, we have not investigated whether our technique can be used to obtain lower bounds for other proof systems. For example, is it possible that with similar ideas one could obtain tree-like cutting planes lower bounds with bounded coefficients? Possibly even with unbounded coefficients? The communication complexity view of the problem suggests that this may be a viable approach.

Acknowledgements

The authors are grateful to Albert Atserias, Per Austrin, Johan Håstad, Jakob Nordström, Pavel Pudlák, Dmitry Sokolov, Joseph Swernofsky, Neil Thapen and Marc Vinyals for helpful discussions and feedback. In particular we would like to thank Albert Atserias who observed that cores seem to be related to kernels.

We benefited from feedback of the participants of the Oberwolfach workshop 2413 "Proof Complexity and Beyond" and would like to thank the anonymous FOCS reviewers whose comments helped us improve the exposition in the paper considerably.

This work was supported by the Approximability and Proof Complexity project funded by the Knut and Alice Wallenberg Foundation. Part of this work was carried out while all authors were associated with KTH Royal Institute of Technology. Other parts were carried out while taking part in the semester program *Lower Bounds in Computational Complexity* in the fall of 2018 and the semster programs *Meta-Complexity* and *Satisfiability: Extended Reunion* in the spring of 2023 at the Simons Institute for the Theory of Computing at UC Berkeley.

Susanna F. de Rezende received funding from ELLIIT, from Knut and Alice Wallenberg grants KAW 2018.0371 and KAW 2021.0307, and from the Swedish Research Council grant 2021-05104. Aaron Potechin was supported by NSF grant CCF-2008920. Kilian Risse is supported by Swiss National Science Foundation project 200021-184656 "Randomness in Problem Instances and Randomized Algorithms".

A Explicit Characterization of $E^\star_F$

Instead of implicitly describing $E^\star_F$ as done in Section 4 we can describe this edge set explicitly as follows. Given a graph F and a set of vertices W, for every $U \subseteq V\big(E(F)\big) \setminus W$ such that there is a matching from U to W that covers U, we let $A_U \subseteq W$ be the set of vertices $w \in W$ that could potentially be used to extend one of these matchings. More formally, A_U is the set of vertices $w \in W$ such that there is a matching from U to $W \setminus \{w\}$ that covers U; in this way, if any vertex $v \in [k] \setminus (W \cup U)$ is adjacent to w, then there is a matching from $U \cup \{v\}$ to W that covers $U \cup \{v\}$. With this definition we can define $E^\star_F$ formally as follows.

Definition A.1 ($E_F^\star$)**.** Let $F \in \mathrm{img}(\mathrm{core})$ and let W, U_1, U_2 denote the objects from Algorithm 1 run on F. For a set $U \subseteq V(E(F)) \setminus W$ with a matching between U and W of size $|U|$, we let $A_U \subseteq W$ be the set of vertices

$$A_U = \{w \in W \mid \exists \text{ matching of size } |U| \text{ between } W \setminus \{w\} \text{ and } U\} \ ,$$

and define

$$E_F^\star = \bigcup_{v \in [k] \setminus V(E(F))} \{\{v, w\} \mid w \in W \setminus (A_{U_1 \cap [v-1]} \cup A_{U_2 \cap [v-1]})\} \ .$$

Lemma A.2. For every $F \in \mathrm{img}(\mathrm{core})$ it holds that $\mathrm{core}(H) = F$ if and only if $H = F \dot\cup E$ for some $E \subseteq E_F^\star$.

Proof. We first argue that given $F \in \mathrm{img}(\mathrm{core})$ then for any $E \subseteq E_F^\star$ it holds that $\mathrm{core}(F \dot\cup E) = F$. Let $H = F \dot\cup E$ for some $E \subseteq E_F^\star$ and denote by W, U_1 and U_2 the sets obtained by Algorithm 1 on F. Note that H has the same edges as F on the vertices $W \cup U_1 \cup U_2$. Let us argue that a run of Algorithm 1 on H results in the same sets W, U_1 and U_2.

1. All edges in $E_F^\star$ are incident to a vertex in W and hence W is a vertex cover of H. Since W is the lexicographically first minimum vertex cover of F it must be the lexicographically first minimum vertex cover of H.
2. We now argue that the set U_1 is the lexicographically first maximal subset of $[k] \setminus W$ with a matching in H of size $|U_1|$ from U_1 to W. To see this, let $U_1' \subseteq [k] \setminus W$ be the lexicographically first maximal set such that there is a matching M in H between U_1' and W that covers U_1', and suppose for the sake of contradiction that $U_1' \neq U_1$. This implies that either $U_1' \supsetneq U_1$ or U_1' is lexicographically smaller than U_1. Let v be the lexicographically first vertex in the symmetric difference of U_1' and U_1. Note that it must be the case that $v \in U_1'$ (since either $U_1' \supsetneq U_1$ or U_1' is maximal and lexicographically smaller than U_1). Let w be the vertex in W that v is matched to in M. Note that $\{v, w\}$ cannot be in F since the set $(U_1' \cap U_1) \cup \{v\}$ would contradict the choice of U_1. Moreover, it holds that $w \in A_{U_1' \cap [v-1]} = A_{U_1 \cap [v-1]}$ hence $\{v, w\}$ is not in $E_F^\star$. But $\{v, w\} \notin E_F^\star \cup F$ contradicts $\{v, w\} \in H \subseteq E_F^\star \cup F$.
3. Similarly, we argue that the set U_2 is the lexicographically first maximal subset disjoint of $W \cup U_1$ with a matching in H of size $|U_2|$ from U_2 to W. To see this, let $U_2' \subseteq [k] \setminus (W \cup U_1)$ be the lexicographically first maximal set such that there is a matching M in H between U_2' and W that covers U_2', and suppose for the sake of contradiction that $U_2' \neq U_2$. This implies that either $U_2' \supsetneq U_2$ or U_2' is lexicographically smaller than U_2. Let v be the lexicographically first vertex in the symmetric difference of U_2' and U_2. Note that it must be the case that $v \in U_2'$ (since either $U_2' \supsetneq U_2$ or U_2' is maximal and lexicographically smaller than U_2). Let w be the vertex in W that v is matched to in M. Note that $\{v, w\}$ cannot be in F since the set $(U_2' \cap U_2) \cup \{v\}$ would contradict the choice of U_2. Moreover, it holds that $w \in A_{U_2' \cap [v-1]} = A_{U_2 \cap [v-1]}$, hence $\{v, w\}$ is not in $E_F^\star$. But $\{v, w\} \notin E_F^\star \cup F$ contradicts $\{v, w\} \in H \subseteq E_F^\star \cup F$.

For the other direction, let W, U_1 and U_2 be the sets obtained by Algorithm 1 on H, and let $F = H[W \cup U_1 \cup U_2]$. Assume, for sake of contradiction, that H contains an edge $e \notin F \cup E_F^\star$. We analyse four cases depending on where the endpoints of e are located.

1. If e is not incident to a vertex in W, then this contradicts the fact that W is a vertex cover of H.
2. If both endpoints of e are in $V\big(E(F)\big) = W \cup U_1 \cup U_2$ then H has different edges than F on the vertices $V\big(E(F)\big)$, contradicting that $F = H[W \cup U_1 \cup U_2]$.

We are left to analyse the case when $e = \{v, w\}$ for some $w \in W$ and $v \in [k] \setminus V\big(E(F)\big)$. Note that since $e \notin E^\star_F$ it must hold that $w \in W \cap A_{U_1 \cap [v-1]}$ or $w \in W \cap A_{U_2 \cap [v-1]}$.

3. If $e = \{v, w\}$ for some $w \in W \cap A_{U_1 \cap [v-1]}$ and $v \in [k] \setminus V\big(E(F)\big)$ (hence $v \notin U_1$), then there is a matching between $U = (U_1 \cap [v-1]) \cup \{v\}$ and W that covers U. Since U precedes U_1 lexicographically, this contradicts our choice of U_1.
4. If $e = \{v, w\}$ for some $w \in W \cap A_{U_2 \cap [v-1]}$ and $v \in [k] \setminus V\big(E(F)\big)$ (hence $v \notin U_2$), then there is a matching between $U = (U_2 \cap [v-1]) \cup \{v\}$ and W that covers U. Since U precedes U_2 lexicographically, this contradicts our choice of U_2.

In each case, we obtained a contradiction and therefore we conclude that if $\mathrm{core}(H) = F$ then H cannot contain an edge $e \notin F \cup E^\star_F$. This completes the proof of the lemma. □

References

[ABdR+21] Albert Atserias, Ilario Bonacina, Susanna F. de Rezende, Massimo Lauria, Jakob Nordström, and Alexander A. Razborov. Clique is hard on average for regular resolution. *J. ACM*, 68(4):23:1–23:26, 2021. 1, 2, 6, 19, 40

[AH19] Albert Atserias and Tuomas Hakoniemi. Size-Degree Trade-Offs for Sums-of-Squares and Positivstellensatz Proofs. In Amir Shpilka, editor, *34th Computational Complexity Conference (CCC 2019)*, volume 137 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 24:1–24:20, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. 2

[AMP21] Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. Graph matrices: Norm bounds and applications, 2021. 40, 43

[BGLR12] Olaf Beyersdorff, Nicola Galesi, Massimo Lauria, and Alexander A. Razborov. Parameterized bounded-depth Frege is not optimal. *ACM Transactions on Computation Theory*, 4(3):7:1–7:16, September 2012. Preliminary version in *ICALP '11*. 2, 19, 40

[BHK+19] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K. Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019. 2, 9, 10

[BIS07] Paul Beame, Russell Impagliazzo, and Ashish Sabharwal. The resolution complexity of independent sets and vertex covers in random graphs. *Comput. Complex.*, 16(3):245–297, 2007. 2, 6, 7, 19, 40

[BT06] Andrej Bogdanov and Luca Trevisan. On worst-case to average-case reductions for NP problems. *SIAM J. Comput.*, 36(4):1119–1159, 2006. 1

[CHKX04] Jianer Chen, Xiuzhen Huang, Iyad A. Kanj, and Ge Xia. Linear FPT reductions and computational lower bounds. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC '04)*, pages 212–221, June 2004. 1

[Coo71] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing (STOC '71)*, pages 151–158, May 1971. 1

[DF95] Rodney Downey and Michael R. Fellows. Fixed-parameter tractability and completeness II: Completeness for W[1]. *Theoretical Computer Science A*, 141(1–2):109–131, April 1995. 1

[DGGM20] Stefan S. Dantchev, Nicola Galesi, Abdul Ghani, and Barnaby Martin. Proof complexity and the binary encoding of combinatorial principles. *CoRR*, abs/2008.02138, 2020. 2

[DM13] Stefan S. Dantchev and Barnaby Martin. Rank complexity gap for lovász-schrijver and sherali-adams proof systems. *Comput. Complex.*, 22(1):191–213, 2013. 2, 3

[DRPR23] Susanna F. De Rezende, Aaron Potechin, and Kilian Risse. Clique is hard on average for unary sherali-adams. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 12–25, 2023. 7

[FHR+24] Yuval Filmus, Edward A. Hirsch, Artur Riazanov, Alexander Smal, and Marc Vinyals. Proving Unsatisfiability with Hitting Formulas. In Venkatesan Guruswami, editor, *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, volume 287 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 48:1–48:20, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 3, 4

[FJK+18] Michael R. Fellows, Lars Jaffke, Aliz Izabella Király, Frances A. Rosamond, and Mathias Weller. *What Is Known About Vertex Cover Kernelization?*, pages 330–356. Springer International Publishing, Cham, 2018. 15

[FMSV23] Yuval Filmus, Meena Mahajan, Gaurav Sood, and Marc Vinyals. MaxSAT resolution and subcube sums. *ACM Transactions on Computational Logic*, 24(1):1–27, January 2023. 2

[GHJ+22] Mika Göös, Alexandros Hollender, Siddhartha Jain, Gilbert Maystre, William Pires, Robert Robere, and Ran Tao. Separations in proof complexity and TFNP. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 1150–1161. IEEE, 2022. 2, 47

[Hås99] Johan Håstad. Clique is hard to approximate within $n^{1-\varepsilon}$. *Acta Mathematica*, 182:105–142, 1999. Preliminary version in *FOCS '96*. 1

[HKT24] Pavel Hubáček, Erfan Khaniki, and Neil Thapen. TFNP Intersections Through the Lens of Feasible Disjunction. In Venkatesan Guruswami, editor, *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, volume 287 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 63:1–63:24, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 8

[Kar72] Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, The IBM Research Symposia Series, pages 85–103. Springer, 1972. 1

[Lau18] Massimo Lauria. Cliques enumeration and tree-like resolution proofs. *Inf. Process. Lett.*, 135:62–67, 2018. 2

[LPRT17] Massimo Lauria, Pavel Pudlák, Vojtěch Rödl, and Neil Thapen. The complexity of proving that a graph is Ramsey. *Combinatorica*, 37(2):253–268, April 2017. Preliminary version in *ICALP '13*. 2

[Mar08] Susan Margulies. *Computer Algebra, Combinatorics, and Complexity: Hilbert's Nullstellensatz and NP-complete Problems*. PhD thesis, University of California, Davis, 2008. 2

[MPW15] Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC '15)*, pages 87–96, June 2015. 2

[NP85] Jaroslav Nešetřil and Svatopluk Poljak. On the complexity of the subgraph problem. *Commentationes Mathematicae Universitatis Carolinae*, 026(2):415–419, 1985. 1

[Pan21a] Shuo Pang. Large clique is hard on average for resolution. In Rahul Santhanam and Daniil Musatov, editors, *Computer Science - Theory and Applications - 16th International Computer Science Symposium in Russia, CSR 2021, Sochi, Russia, June 28 - July 2, 2021, Proceedings*, volume 12730 of *Lecture Notes in Computer Science*, pages 361–380. Springer, 2021. 2

[Pan21b] Shuo Pang. SOS lower bound for exact planted clique. In Valentine Kabanets, editor, *36th Computational Complexity Conference, CCC 2021, July 20-23, 2021, Toronto, Ontario, Canada (Virtual Conference)*, volume 200 of *LIPIcs*, pages 26:1–26:63. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. 2

[PZ22] Aaron Potechin and Aaron Zhang. Bounds on the total coefficient size of nullstellensatz proofs of the pigeonhole principle and the ordering principle, 2022. 8

[Ros08] Benjamin Rossman. On the constant-depth complexity of k-clique. In Cynthia Dwork, editor, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 721–730. ACM, 2008. 1

[Ros10] Benjamin Rossman. The monotone complexity of k-clique on random graphs. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 193–201. IEEE Computer Society, 2010. 1

[SA94] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations and convex hull characterizations for mixed-integer zero—one programming problems. *Discrete Applied Mathematics*, 52(1):83–106, 1994. 2, 3

[Zuc07] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3(6):103–128, August 2007. Preliminary version in *STOC '06*. 1