

Approximating Independent Sets Using Additive Combinatorics

Shachar Lovett*

University of California, San Diego

slovett@ucsd.edu

October 18, 2026

Abstract

We give a deterministic quasi-polynomial time algorithm for approximating the largest independent set in a graph. We show that if an N -vertex graph contains an independent set of size $2^{-t}N$, then we can find an independent set of size 2^t in quasi-polynomial time, for all t up to a constant factor of $(\log N)^{1/3}$. This gives an approximation factor of $N/\exp((\log N)^{1/3})$, which has the same shape as known hardness results for quasi-polynomial time algorithms, up to a different exponent of the logarithm.

The algorithm builds on a structural principle: a graph containing a large independent set must also contain a large sparse structured set, and such a set can be found efficiently. A greedy algorithm then extracts an independent set from it. Our proof draws on tools in additive combinatorics, in particular spreadness and sifting, to obtain the combinatorial structural results. We then convert these to efficient algorithms using pseudo-marginals obtained by semidefinite programming.

1 Introduction

Finding a maximum independent set is a classical NP-hard problem [Kar72]. Throughout, $G = (V, E)$ is a simple undirected graph on N vertices, and $\alpha(G)$ denotes its maximum independent-set size. An r -approximation algorithm returns an independent set of size at least $\alpha(G)/r$.

Polynomial time approximation. Feige [Fei04] gives the best known polynomial time approximation ratio,

$$O\left(\frac{N(\log \log N)^2}{(\log N)^3}\right).$$

On the hardness side, Håstad [Hås99] proved that, for every constant $\varepsilon > 0$, an $N^{1-\varepsilon}$ -approximation is impossible in polynomial time unless $\text{NP} = \text{ZPP}$. Zuckerman [Zuc07] derandomized this result, establishing NP-hardness for the same factor. So, the gap between known algorithms and hardness results is large: polylogarithmic versus small polynomial savings over the trivial factor of N .

Quasi-polynomial time approximation. We now switch to consider algorithms that are allowed to run in quasi-polynomial time, namely $N^{O(\log N)^a}$ time for any constant $a > 0$. Within this time limit, exhaustive search gives a an approximation ratio of $O(N/(\log N)^a)$, by enumerating all vertex sets of size at most $(\log N)^a$. The enumeration either finds an optimum independent set or finds one of size $(\log N)^a$. On the hardness side, building on Khot [Kho01], Zuckerman [Zuc07] ruled out quasi-polynomial time approximation within $N/2^{(\log N)^{1-\gamma}}$ for some constant $\gamma > 0$, unless NP has deterministic quasi-polynomial time algorithms. Under the stronger assumption that NP has no bounded-error randomized quasi-polynomial time algorithms,

*Supported by Simons Investigator Award #929894 and NSF award CCF-2425349.

Khot and Ponnuswami [KP06] ruled out polynomial time approximation within $N/2^{(\log N)^{3/4+\varepsilon}}$ for every constant $0 < \varepsilon < 1/4$.

In this paper, we design a new quasi-polynomial time algorithm that achieves the approximation ratio

$$N/2^{\Omega(\log N)^{1/3}}.$$

This has the same form as the hardness bound of [KP06], though the exponents differ: $1/3$ in our algorithm and $3/4 + \varepsilon$ in the hardness bound. Formally, we prove the following theorem.

Theorem 1.1 (Finding a large independent set). *There is an absolute constant $c > 0$ such that the following holds. Let t be an integer with $1 \leq t \leq c(\log N)^{1/3}$. If $\alpha(G) \geq 2^{-t}N$, then a deterministic algorithm finds an independent set in G of size at least 2^t in time $N^{O(t)}$.*

To obtain the approximation guarantee, take $t = c(\log N)^{1/3}$ and run the algorithm, returning a singleton if it finds no independent set. If $\alpha(G) \geq N/2^t$, the output has size at least 2^t , giving an approximation factor of at most $N/2^t$. If $\alpha(G) < N/2^t$, a singleton already gives this factor. The running time is $N^{O((\log N)^{1/3})}$ which is quasi-polynomial in N .

Large sparse structured sets. The algorithm in Theorem 1.1 is based on two steps. First, we find a large sparse *structured* set in the graph, and then extract an independent set from it by a simple greedy algorithm. The first step is the main innovation of this paper. It is a manifestation of the following guiding principle:

*A large exact solution guarantees a large structured approximate solution,
with the structure captured by a small, efficiently searchable family.*

In our case, an exact solution is a large independent set, and an approximate solution is a set with few edges. The specific structure we utilize will become apparent soon.

We now move to the formal statement of our main technical theorem. For a nonempty set of vertices $U \subseteq V$, define its edge density by

$$d(U) := \Pr_{u,v \in U}[uv \in E] = \frac{2|E(G[U])|}{|U|^2}.$$

Here u, v are chosen independently and uniformly from U .

Theorem 1.2 (Finding a large sparse set). *Let $t, h \geq 1$ be integers. If $\alpha(G) \geq 2^{-t}N$, then a deterministic algorithm running in time $N^{O(t)}$ finds a nonempty set $U \subseteq V$ satisfying*

$$|U| \geq 2^{-O(t^2h)}N, \quad d(U) \leq 2^{-h}.$$

Before describing the proof approach for Theorem 1.2, we first show that Theorem 1.1 follows from it by simple greedy pruning.

Proof of Theorem 1.1 given Theorem 1.2. Apply Theorem 1.2 with $h = t + 1$. For a sufficiently small choice of c , the resulting set U satisfies

$$|U| \geq N2^{-O(t^3)} \geq 2^{t+1}, \quad d(U) \leq 2^{-(t+1)}.$$

We work in the induced subgraph $G[U]$. Consider the greedy algorithm that iteratively selects a vertex of minimum degree, adds it to the independent set, and deletes it together with its neighbors. This algorithm is well-known to attain Turán's bound (see, e.g., [HR97]), finding an independent set in $G[U]$ of size at least

$$\frac{|U|}{1 + d(U)|U|} = \frac{1}{|U|^{-1} + d(U)} \geq 2^t.$$

□

1.1 Proof overview

We first illustrate our guiding principle for Cayley graphs over $\mathbb{F}_2^n$, where existing additive-combinatorial results supply the structured sparse sets. Our proof for general graphs uses some of the same ideas directly, in particular spreadness and sifting, developed by Kelley and Meka [KM23] in the arithmetic setting and by Kelley, Lovett, and Meka [KLM24] in the combinatorial setting.

Cayley graphs. Consider a Cayley graph G on $\mathbb{F}_2^n$, with $N = 2^n$ vertices and generating set $S \subseteq \mathbb{F}_2^n \setminus \{0\}$: vertices x, y are adjacent when $x + y \in S$. If A is an independent set in G , then

$$(A + A) \cap S = \emptyset.$$

We will use the additive structure in $A + A$ to produce a structured sparse set in G .

Assume $|A| \geq 2^{-t}N$ for an integer $t \geq 1$. The quasi-polynomial Bogolyubov lemma, following the Kelley-Meka framework [KM23] in the form proved by Hunter and Pohoata [HP25], gives, for every integer $h \geq 1$, a linear subspace $U \leq \mathbb{F}_2^n$ such that

$$\text{codim}(U) = O(t^5 h^2), \quad |U \setminus (A + A)| \leq 2^{-h}|U|.$$

Since S avoids $A + A$, at most a 2^{-h} fraction of U lies in S . For independent uniform $x, y \in U$, the sum $x + y$ is uniform in U . Consequently,

$$|U| \geq N 2^{-O(t^5 h^2)}, \quad d(U) = \frac{|S \cap U|}{|U|} \leq 2^{-h}.$$

This is the structural principle in a concrete form: U is large and sparse, and its structure is that of a linear subspace. A subspace of codimension k is the common zero set of k linear forms, for which there are at most N^k choices. Enumerating subspaces of codimension at most $O(t^5 h^2)$ and checking their edge densities therefore finds a suitable U in deterministic time $N^{O(t^5 h^2)}$.

General graphs. We begin by assuming access to a distribution on independent sets whose expected size is large. This lets us first develop the combinatorial argument. We then show that access to the distribution can be replaced by access to its marginals, namely the probabilities that small sets of vertices are contained in the random independent set. Finally, we show that pseudo-marginals computed by an SDP suffice to implement the argument efficiently.

The first step is to increase inclusion probabilities by conditioning. If a set of vertices contains many edges and each vertex has a reasonably large inclusion probability, Lemma 2.1 shows that conditioning on one vertex belonging to the independent set increases the inclusion probabilities of many other vertices by a constant factor.

We then iterate this step, retaining the vertices whose inclusion probabilities increase. Since probabilities cannot exceed one, the process must stop with a sparse set. In Lemma 2.2, we track the number of retained vertices to show that this sparse set is large.

To make the argument algorithmic, we show that it uses only small marginals and remains valid for pseudo-marginals obtained from a semidefinite relaxation. Each candidate sparse set is determined by a small conditioning set. Enumerating these conditioning sets and checking the resulting candidates gives an efficient algorithm.

1.2 Related work

Our work builds on a long line of research in additive combinatorics and approximation algorithms. Its main contribution is to use additive-combinatorial techniques to obtain strong quantitative guarantees from semidefinite relaxations of the independent set problem.

Finding almost-cliques and sparse sets. Algorithms for finding almost-cliques [FS97, Bar18] already yield sparse sets when large independent sets exist. These algorithms preserve the set size; we allow it to shrink while achieving much smaller edge density within quasi-polynomial time. Related hardness results appear in [BKRW17].

From distributions to SDP rounding. Barak, Kelner, and Steurer [BKS14] develop a framework for converting arguments about distributions into SDP rounding algorithms. We follow this approach, using only small marginals. High entrywise powers of the correlation matrix require no additional SDP levels.

Correlation rounding and conditioning. Conditioning is central to correlation rounding [BRS11, RT12], where progress is measured through variance or entropy. Our iteration increases all retained inclusion marginals by a constant factor, bounding the number of conditioning steps.

Odd correlation moments and spectral methods. Jin, Milojević, Tomon, and Zhang [JMTZ25] use entrywise cubes and the Schur product theorem to obtain dense subgraphs under spectral assumptions. We apply a related odd-moment argument to a normalized covariance matrix, obtaining positive correlations that yield a marginal increment.

Hierarchies for independent set and densest subgraph. Local distributions and small conditioning sets also appear in independent-set algorithms [GS13, BGG18] and densest-subgraph algorithms [BCC⁺10]. Our argument uses these tools to find sparse sets, with guarantees depending only on the size of the promised independent set.

Containers and additive-combinatorial sifting. Container theorems [BMS15, ST15] provide another precedent for sparse sets encoded by small vertex subsets. Our density-increment argument draws on Kelley–Meka and Kelley–Lovett–Meka [KM23, KLM24]; Bloom–Sisask [BS23] explain the connection between sifting and dependent random choice.

1.3 Open problems

The most immediate open problem is to narrow the gap between algorithms and hardness results for independent set approximation. A broader and, in our view, more interesting question is where else techniques from additive combinatorics can yield stronger quantitative guarantees for approximation algorithms. In particular, can we find further realizations of our guiding principle, where a large exact solution guarantees a large structured approximate solution belonging to a small, efficiently searchable family?

Paper organization. We prove Theorem 1.2 in Section 2 and describe our AI methodology in Section 3.

2 Finding large sparse structured sets, efficiently

We prove Theorem 1.2 in this section. We first work with a distribution μ over independent sets whose expected size is large. We then show that the argument uses only small marginals and remains valid for pseudo-marginals, which can be computed efficiently by semidefinite programming.

Notation and definitions. Let $G = (V, E)$ be a graph on N vertices. For a distribution μ on the independent sets of G and a set $T \subseteq V$, write

$$\mu(T) := \Pr_{I \sim \mu}[T \subseteq I].$$

Thus $\mu(T)$ denotes the probability that the random independent set *contains* T . We abbreviate $\mu(\{v\})$ as $\mu(v)$ and $\mu(\{u, v\})$ as $\mu(u, v)$. In particular,

$$0 \leq \mu(v) \leq 1, \quad 0 \leq \mu(u, v) \leq \min\{\mu(u), \mu(v)\}, \quad \mu(u, v) = 0 \quad \text{if } uv \in E. \quad (1)$$

For $S \subseteq V$ with $\mu(S) > 0$, let μ_S be the distribution of $I \sim \mu$ conditioned on $S \subseteq I$, where

$$\mu_S(T) = \frac{\mu(S \cup T)}{\mu(S)}.$$

We write μ_v for $\mu_{\{v\}}$. Successive conditionings satisfy $(\mu_S)_T = \mu_{S \cup T}$ whenever $\mu(S \cup T) > 0$.

2.1 Increasing marginals by conditioning

The first step is inspired by the sifting and spectral-positivity arguments of Kelley and Meka [KM23] and their extension to general bipartite graphs through grid norms by Kelley, Lovett, and Meka [KLM24]. The main idea is to use high moments of correlations in a distribution on independent sets to obtain a density increment by conditioning.

Lemma 2.1 (Increased marginals in dense subgraphs). *Let μ be a distribution on independent sets of G , and let $U \subseteq V$. Let $\delta, q \in (0, 1)$ and suppose $d(U) \geq \delta$ and $\mu(u) \geq q$ for every vertex $u \in U$. Let ℓ be a positive odd integer satisfying $2^{-\ell} \leq \delta/2$. Then some $v \in U$ satisfies*

$$|\{u \in U : \mu_v(u) \geq \frac{3}{2}\mu(u)\}| \geq \frac{\delta}{2} q^\ell |U|.$$

Proof. Define a matrix M indexed by $u, v \in U$ as

$$M(u, v) = \frac{\mu(u, v)}{\mu(u)\mu(v)} - 1.$$

We claim that M is positive semidefinite. To see this, consider the matrix R indexed by $u \in U$ and an independent set I , given by $R(u, I) = \frac{\mathbf{1}_{[u \in I]}}{\mu(u)} - 1$, and note that $M(u, v) = \mathbb{E}_{I \sim \mu} R(u, I) R(v, I)$. The marginal bounds on μ in Equation (1) and our assumptions give

$$M(u, v) = -1 \quad \text{if } uv \in E, \quad M(u, v) \leq q^{-1} \quad \text{for all } u, v \in U.$$

Write $M^{\circ \ell} = (M(u, v)^\ell)_{u, v \in U}$ for the entrywise ℓ -th power of M . Since M is positive semidefinite, by the Schur product theorem [Sch11], $M^{\circ \ell}$ is also positive semidefinite, and so

$$\mathbb{E}_{u, v \in U} M(u, v)^\ell = \frac{\mathbf{1}^\top M^{\circ \ell} \mathbf{1}}{|U|^2} \geq 0,$$

where $\mathbf{1}$ is the all-ones vector. This nonnegativity can also be seen directly:

$$\mathbb{E}_{u, v \in U} [M(u, v)^\ell] = \mathbb{E}_{I_1, \dots, I_\ell \sim \mu} \left(\mathbb{E}_{u \in U} \prod_{j=1}^{\ell} R(u, I_j) \right)^2 \geq 0.$$

We next consider the contribution of each term $M(u, v)$ to the ℓ -th moment. Since ℓ is odd, each edge contributes -1 and hence edges contribute $-d(U)$ in total. Other negative terms contribute a negative amount. Terms with $0 \leq M(u, v) < 1/2$ contribute at most $2^{-\ell}$ in total, while those with $M(u, v) \geq 1/2$ contribute at most $\rho q^{-\ell}$, where $\rho = \Pr_{u, v \in U} [M(u, v) \geq 1/2]$. Thus

$$0 \leq \mathbb{E}_{u, v \in U} [M(u, v)^\ell] \leq -d(U) + 2^{-\ell} + \rho q^{-\ell} \leq -\frac{\delta}{2} + \rho q^{-\ell}.$$

This implies that $\rho \geq (\delta/2)q^\ell$. By averaging, some $v \in U$ has at least $\rho|U|$ vertices $u \in U$ with $M(u, v) \geq 1/2$. For each such vertex u ,

$$\mu_v(u) = \frac{\mu(u, v)}{\mu(v)} \geq \frac{3}{2}\mu(u),$$

which proves the claim. $\square$

2.2 Finding a large sparse set by repeated conditioning

We now iterate Lemma 2.1. Each conditioning multiplies the marginal threshold by $3/2$. This process must stop because inclusion marginals are bounded by 1.

Lemma 2.2 (A large sparse set after conditioning). *Let $t, h \geq 1$ be integers, and assume $N \geq 2^{3(t+2)^2(h+2)}$. Let μ be a distribution on independent sets of G such that $\mathbb{E}_{I \sim \mu} |I| \geq 2^{-t}N$. For each $S \subseteq V$ with $\mu(S) > 0$, define*

$$U_S = \left\{ u \in V \setminus S : \mu_S(u) \geq (3/2)^{|S|} 2^{-(t+1)} \right\}.$$

There is such an S with $|S| \leq 2(t+1)$ for which

$$|U_S| \geq 2^{-3(t+2)^2(h+2)}N, \quad d(U_S) \leq 2^{-h}.$$

Proof. Start with $S = \emptyset$. Since

$$2^{-t}N \leq \mathbb{E}_{I \sim \mu} |I| = \sum_{u \in V} \mu(u) \leq |U_\emptyset| + 2^{-(t+1)}N,$$

we have $|U_\emptyset| \geq 2^{-(t+1)}N$. Let $K = 2^{(t+2)(h+2)}$. We will maintain the invariant $|U_S| \geq 2^{-(t+1)}K^{-|S|}N$ throughout the process.

Suppose $\mu(S) > 0$ and $d(U_S) > 2^{-h}$. Apply Lemma 2.1 to μ_S and U_S , using the marginal lower bound $\mu_S(u) \geq (3/2)^{|S|} 2^{-(t+1)} \geq 2^{-(t+1)}$ for all $u \in U_S$, and choosing $\ell \in \{h+1, h+2\}$ to be odd. We obtain a vertex $v \in U_S$ for which at least

$$2^{-(h+1)-(t+1)\ell} |U_S| \geq \frac{2|U_S|}{K}$$

vertices $u \in U_S$ satisfy $\mu_{S \cup \{v\}}(u) \geq \frac{3}{2} \mu_S(u)$. After removing v , these vertices lie in $U_{S \cup \{v\}}$. We note that the lower bound on $|U_S|$ and our assumption that N is large enough imply $|U_S| \geq K$. Hence

$$|U_{S \cup \{v\}}| \geq \frac{2|U_S|}{K} - 1 \geq \frac{|U_S|}{K} \geq 2^{-(t+1)}K^{-(|S|+1)}N,$$

which maintains the invariant. Since $S \cap U_S = \emptyset$, this adds one vertex to S ; moreover, $\mu(S \cup \{v\}) = \mu(S)\mu_S(v) > 0$.

Repeat this process as long as $d(U_S) > 2^{-h}$. The process must stop before $|S| = 2(t+1)$: at that point the marginal threshold would be $(3/2)^{2(t+1)} 2^{-(t+1)} = (9/8)^{t+1} > 1$, forcing $U_S = \emptyset$ and contradicting the size bound $|U_S| \geq K$ that we showed. When the process terminates at some set S , we have $d(U_S) \leq 2^{-h}$ and

$$|U_S| \geq 2^{-(t+1)}K^{-|S|}N \geq 2^{-(t+1)}K^{-2(t+1)}N \geq 2^{-3(t+2)^2(h+2)}N.$$

□

2.3 Finding large sparse sets using pseudo-marginals

We now turn our combinatorial observations into efficient algorithms. The first step is to note that in order to apply Lemma 2.2 we do not need full access to μ ; it suffices to get access to small marginals of μ . In fact, these do not have to be marginals of a real distribution; a pseudo-distribution is sufficient. The marginals of such a pseudo-distribution can be found by a semidefinite program.

Marginals suffice. Suppose we have access to the marginals $\mu(T)$ for all $|T| \leq 2t+3$, where μ satisfies the assumptions of Lemma 2.2. Enumerate all $S \subseteq V$ with $|S| \leq 2(t+1)$ and $\mu(S) > 0$. For each S , compute

$$\mu_S(u) = \frac{\mu(S \cup \{u\})}{\mu(S)}, \quad U_S = \left\{ u \in V \setminus S : \mu_S(u) \geq (3/2)^{|S|} 2^{-(t+1)} \right\}.$$

Check the size and edge density of U_S directly in G and return a set satisfying the conclusion of the lemma. The lemma guarantees that one exists. There are $N^{O(t)}$ candidates, and each can be checked in polynomial time. Thus small marginals suffice; we do not need to sample full independent sets from μ .

Computing pseudo-marginals by SDP. A level- r pseudo-distribution on independent sets is represented by numbers $\tilde{\mu}(T)$ for $T \subseteq V$, $|T| \leq 2r$, satisfying

$$\tilde{\mu}(\emptyset) = 1, \quad \tilde{\mu}(T) = 0 \text{ if } T \text{ contains an edge}, \quad (\tilde{\mu}(A \cup B))_{|A|, |B| \leq r} \succeq 0. \quad (2)$$

These numbers need not be the marginals of an actual distribution. The last constraint requires the matrix indexed by sets A, B of size at most r , with entry $\tilde{\mu}(A \cup B)$, to be positive semidefinite. This matrix is often called the *moment matrix*. Maximizing $\sum_{u \in V} \tilde{\mu}(u)$ subject to (2) is the standard Lasserre SDP relaxation for independent set [Lau03]. We use the standard algorithmic guarantee: if $\alpha(G) \geq 2^{-t}N$, then in time $N^{O(r)} \text{poly}(t)$ one can compute rational feasible pseudo-marginals with

$$\sum_{u \in V} \tilde{\mu}(u) \geq 2^{-(t+1)}N. \quad (3)$$

The factor-two slack allows approximate optimization of the objective; the marginal constraints remain exact. For the bit-complexity guarantee for this relaxation, see [RW17, GPS23].

We verify that these pseudo-marginals can replace μ in the combinatorial argument.

Claim 2.3. *At level $r \geq 2$, the pseudo-marginals satisfy the bounds in Equation (1), with $\tilde{\mu}$ in place of μ .*

Proof. The principal submatrix on $\emptyset, \{u\}$ gives $0 \leq \tilde{\mu}(u) \leq 1$, the diagonal entry indexed by $\{u, v\}$ gives nonnegativity, and the quadratic form on the difference of the indices $\{u\}$ and $\{u, v\}$ gives $\tilde{\mu}(u) - \tilde{\mu}(u, v) \geq 0$. The edge marginals vanish by Equation (2). $\square$

Conditioning uses the same notation and formula as before:

$$\tilde{\mu}_S(T) = \frac{\tilde{\mu}(S \cup T)}{\tilde{\mu}(S)} \quad \text{when } |S| \leq r \text{ and } \tilde{\mu}(S) > 0.$$

This gives a level- $(r - |S|)$ pseudo-distribution. Its moment matrix has entries

$$\tilde{\mu}_S(A \cup B) = \frac{\tilde{\mu}(A \cup B \cup S)}{\tilde{\mu}(S)}, \quad |A|, |B| \leq r - |S|,$$

so it is obtained by selecting (possibly repeating) rows and columns $A \cup S, B \cup S$ of the original moment matrix and dividing by $\tilde{\mu}(S) > 0$. It is therefore also positive semidefinite. Normalization and the edge constraints are also preserved. Thus each conditioning consumes one level of the pseudo-distribution.

It remains to justify the nonnegative moment used in Lemma 2.1. On any set U with positive marginals, define

$$M(u, v) = \frac{\tilde{\mu}(u, v)}{\tilde{\mu}(u)\tilde{\mu}(v)} - 1.$$

Claim 2.4. *The matrix M is positive semidefinite.*

Proof. Let $(a_u)_{u \in U}$ be arbitrary real coefficients. Define coefficients b_A , indexed by sets A of size at most r , by $b_{\{u\}} = a_u/\tilde{\mu}(u)$ for $u \in U$, $b_\emptyset = -\sum_{u \in U} a_u$, and $b_A = 0$ elsewhere. Positive semidefiniteness of the moment matrix gives

$$\begin{aligned} 0 &\leq \sum_{\substack{A, B \subseteq V \\ |A|, |B| \leq r}} b_A b_B \tilde{\mu}(A \cup B) \\ &= \sum_{u, v \in U} a_u a_v \frac{\tilde{\mu}(u, v)}{\tilde{\mu}(u)\tilde{\mu}(v)} - 2 \left(\sum_{u \in U} a_u \right)^2 + \left(\sum_{u \in U} a_u \right)^2 \\ &= \sum_{u, v \in U} a_u a_v M(u, v). \end{aligned}$$

Since this holds for every choice of the coefficients, $M \succeq 0$. $\square$

By the Schur product theorem [Sch11], every entrywise positive integer power $M^{\circ\ell} = (M(u, v)^\ell)_{u, v \in U}$ is positive semidefinite. Hence

$$\mathbb{E}_{u, v \in U} M(u, v)^\ell = \frac{\mathbf{1}^\top M^{\circ\ell} \mathbf{1}}{|U|^2} \geq 0,$$

where $\mathbf{1}$ is the all-ones vector. The rest of Lemma 2.1 uses only the marginal bounds and the conditioning formula, so its proof applies unchanged. In particular, taking a high entrywise power requires no additional levels. Consequently, Lemma 2.2 holds for pseudo-distributions as well, provided at least two levels remain after all the conditionings.

Proof of Theorem 1.2. If $N < 2^{3(t+3)^2(h+2)}$, a singleton satisfies the required bounds. Otherwise, compute a level- r pseudo-distribution with $r = 2(t+2) + 2$ satisfying Equation (3). Apply the pseudo-distribution version of Lemma 2.2 with $t+1$ in place of t . There are at most $2(t+2)$ conditionings, so two levels remain throughout. The enumeration described above, using $\tilde{\mu}$ in place of μ , therefore finds a set U with

$$|U| \geq 2^{-3(t+3)^2(h+2)} N = 2^{-O(t^2 h)} N, \quad d(U) \leq 2^{-h}.$$

Both the SDP computation and the enumeration take time $N^{O(t)}$. □

3 AI methodology

This project began with an idea I had been considering for some time: additive combinatorics can be used to show that a large independent set in a Cayley graph implies the existence of a large sparse structured set, which can be found in quasi-polynomial time. More recently, I realized that the spreadness techniques we developed in [KLM24] could extend this approach to general graphs, but I did not know how to make the argument algorithmic. I used the web version of ChatGPT Astra to help develop the proof. The interaction proceeded as follows:

1. I first asked it to prove the conjecture for Cayley graphs over $\mathbb{F}_2^n$, where I already knew a proof, to see whether it could find one in the literature or derive one itself. It did neither.
2. I explained my proof for Cayley graphs and asked it to extend the argument to general graphs, without further hints. It did not find an extension.
3. I then suggested using spreadness and sifting from [KLM24] to obtain a density increment, and asked it to work out the details. This and further interactions led to the combinatorial results in this paper.
4. Finally, I asked how to turn these combinatorial observations into an efficient algorithm. It identified the connection to SDP relaxations and pseudo-marginals, which I had not been aware of.

AI was also used to copy-edit the paper, calculate parameters and do a literature search.

References

- [Bar18] Siddharth Barman. Approximating Nash equilibria and dense subgraphs via an approximate version of Carathéodory’s theorem. *SIAM Journal on Computing*, 47(3):960–981, 2018. <https://arxiv.org/abs/1406.2296>.
- [BCC⁺10] Aditya Bhaskara, Moses Charikar, Eden Chlamtáč, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities—an $O(n^{1/4})$ approximation for densest k -subgraph. In *Proceedings of the ACM Symposium on Theory of Computing (STOC)*, pages 201–210, 2010. <https://arxiv.org/abs/1001.2891>.

- [BGG18] Nikhil Bansal, Anupam Gupta, and Guru Guruganesh. On the Lovász theta function for independent sets in sparse graphs. *SIAM Journal on Computing*, 47(3):1039–1055, 2018. <https://arxiv.org/abs/1504.04767>.
- [BKRW17] Mark Braverman, Young Kun Ko, Aviad Rubinfeld, and Omri Weinstein. ETH hardness for densest- k -subgraph with perfect completeness. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1326–1341, 2017. <https://arxiv.org/abs/1504.08352>.
- [BKS14] Boaz Barak, Jonathan A. Kelner, and David Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the ACM Symposium on Theory of Computing (STOC)*, pages 31–40, 2014. <https://arxiv.org/abs/1312.6652>.
- [BMS15] József Balogh, Robert Morris, and Wojciech Samotij. Independent sets in hypergraphs. *Journal of the American Mathematical Society*, 28:669–709, 2015. <https://arxiv.org/abs/1204.6530>.
- [BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS)*, 2011. <https://arxiv.org/abs/1104.4680>.
- [BS23] Thomas F. Bloom and Olof Sisask. The Kelley–Meka bounds for sets free of three-term arithmetic progressions. *Essential Number Theory*, 2(1):15–44, 2023. <https://arxiv.org/abs/2302.07211>.
- [Fei04] Uriel Feige. Approximating maximum clique by removing subgraphs. *SIAM Journal on Discrete Mathematics*, 18(2):219–225, 2004. <https://doi.org/10.1137/S089548010240415X>.
- [FS97] Uriel Feige and Michael Seltser. On the densest k -subgraph problem. Technical Report CS97-16, Weizmann Institute of Science, 1997.
- [GPS23] Sander Gribling, Sven Polak, and Lucas Slot. A note on the computational complexity of the moment-SOS hierarchy for polynomial optimization, 2023. <https://arxiv.org/abs/2305.14944>.
- [GS13] Venkatesan Guruswami and Ali Kemal Sinop. Rounding Lasserre SDPs using column selection and spectrum-based approximation schemes for graph partitioning and quadratic IPs, 2013. <https://arxiv.org/abs/1312.3024>.
- [Hås99] Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182(1):105–142, 1999. <https://doi.org/10.1007/BF02392825>.
- [HP25] Zach Hunter and Cosmin Pohoata. On off-diagonal Ramsey numbers for vector spaces over $\mathbb{F}_2$. *Mathematical Proceedings of the Cambridge Philosophical Society*, 179(2):503–518, 2025. <https://arxiv.org/abs/2309.02424v2>.
- [HR97] Magnús M. Halldórsson and Jaikumar Radhakrishnan. Greed is good: Approximating independent sets in sparse and bounded-degree graphs. *Algorithmica*, 18(1):145–163, 1997. <https://doi.org/10.1007/BF02523693>.
- [JMTZ25] Zhihan Jin, Aleksa Milojević, István Tomon, and Shengtong Zhang. From small eigenvalues to large cuts, and Chowla’s cosine problem, 2025. <https://arxiv.org/abs/2509.03490>.
- [Kar72] Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972. https://doi.org/10.1007/978-1-4684-2001-2_9.
- [Kho01] Subhash Khot. Improved inapproximability results for MaxClique, chromatic number and approximate graph coloring. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 600–609, 2001. <https://doi.org/10.1109/SFCS.2001.959936>.

- [KLM24] Zander Kelley, Shachar Lovett, and Raghu Meka. Explicit separations between randomized and deterministic Number-on-Forehead communication. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1299–1310, 2024. <https://doi.org/10.1145/3618260.3649721>.
- [KM23] Zander Kelley and Raghu Meka. Strong bounds for 3-progressions. In *Proceedings of the 64th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 933–973, 2023. <https://doi.org/10.1109/FOCS57990.2023.00059>.
- [KP06] Subhash Khot and Ashok Kumar Ponnuswami. Better inapproximability results for MaxClique, chromatic number and Min-3Lin-Deletion. In *Automata, Languages and Programming (ICALP)*, volume 4051 of *Lecture Notes in Computer Science*, pages 226–237. Springer, 2006. https://doi.org/10.1007/11786986_21.
- [Lau03] Monique Laurent. A comparison of the Sherali–Adams, Lovász–Schrijver, and Lasserre relaxations for 0–1 programming. *Mathematics of Operations Research*, 28(3):470–496, 2003. <https://doi.org/10.1287/moor.28.3.470.16391>.
- [RT12] Prasad Raghavendra and Ning Tan. Approximating CSPs with global cardinality constraints using SDP hierarchies. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2012. <https://arxiv.org/abs/1110.1064>.
- [RW17] Prasad Raghavendra and Benjamin Weitz. On the bit complexity of sum-of-squares proofs. In *44th International Colloquium on Automata, Languages, and Programming (ICALP 2017)*, volume 80 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 80:1–80:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017. <https://doi.org/10.4230/LIPIcs.ICALP.2017.80>.
- [Sch11] Issai Schur. Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen. *Journal für die reine und angewandte Mathematik*, 140:1–28, 1911. <https://doi.org/10.1515/crll.1911.140.1>.
- [ST15] David Saxton and Andrew Thomason. Hypergraph containers. *Inventiones Mathematicae*, 201:925–992, 2015. <https://arxiv.org/abs/1204.6595>.
- [Zuc07] David Zuckerman. Linear degree extractors and the inapproximability of Max Clique and Chromatic Number. *Theory of Computing*, 3:103–128, 2007. <https://doi.org/10.4086/toc.2007.v003a006>.