Sherali-Adams Integrality Gaps Matching the Log-Density Threshold

Eden Chlamtáč[1]

Ben-Gurion University, Beer Sheva, Israel
chlamtac@cs.bgu.ac.il

Pasin Manurangsi[2]

University of California, Berkeley, USA
pasin@berkeley.edu

─── **Abstract** ───

The log-density method is a powerful algorithmic framework which in recent years has given rise to the best-known approximations for a variety of problems, including Densest-k-Subgraph and Small Set Bipartite Vertex Expansion. These approximations have been conjectured to be optimal based on various instantiations of a general conjecture: that it is hard to distinguish a fully random combinatorial structure from one which contains a similar planted sub-structure with the same "log-density".

We bolster this conjecture by showing that in a random hypergraph with edge probability $n^{-\alpha}$, $\tilde{\Omega}(\log n)$ rounds of Sherali-Adams cannot rule out the existence of a k-subhypergraph with edge density $k^{-\alpha-o(1)}$, for any k and α. This holds even when the bound on the objective function is lifted. This gives strong integrality gaps which exactly match the gap in the above distinguishing problems, as well as the best-known approximations, for Densest k-Subgraph, Smallest p-Edge Subgraph, their hypergraph extensions, and Small Set Bipartite Vertex Expansion (or equivalently, Minimum p-Union). Previously, such integrality gaps were known only for Densest k-Subgraph for one specific parameter setting.

1 Introduction

The log-density framework [7] is an emerging technique in approximation algorithms that proposes to understand the problems of interest via an average case study. More specifically, the first step in this framework is to consider a distinguishing problem between a "random instance" and a "random instance with planted solution". Instead of considering any algorithm, the focus here is on a simple "witness counting" algorithm, in which one only counts the number of occurrences of a certain substructure (i.e. witness) of the two instances. One then uses the insights from this simple witness to devise an algorithm for worst case instances. Although at first glance this last step may seem like a large leap (i.e. from simple algorithms

---

for average case instances to more complicated algorithms for worst case instances) and even somewhat implausible, the framework has turned out to be quite effective in tackling a number of problems, including Densest k-Subgraph (DkS) [7], Lowest Degree 2-Spanner, Smallest p-Edge Subgraph (SpES) [13], Small Set Bipartite Vertex Expansion (SSBVE) [14], Label Cover and 2-CSPs [15].

To be more concrete, let us consider the Densest k-Subgraph (DkS) problem. In DkS, we are given a graph $G = (V, E)$ and an integer k; the goal is to find subgraph of G of at most k vertices that induces a maximum number of edges. [7] considers the following distinguishing problem:

▶ **Definition 1** (Random vs Planted Subgraph Problem). For certain parameters $\alpha, \beta \in (0, 1), 1 < k < n$, given a graph G, decide whether it is sampled from the Erdős-Rényi distribution $\mathcal{G}(n, n^{-\alpha})$ or from the same distribution, but with a planted subgraph on k randomly chosen vertices, sampled from $\mathcal{G}(k, k^{-\beta})$.

They noted that simple witness counting algorithms can solve the distinguishing problem w.h.p. in time $n^{O(1/\varepsilon)}$ whenever $\beta \leqslant \alpha - \varepsilon$, and fail to distinguish when $\beta \geqslant \alpha$. In particular, when $k = n^{\alpha}$, the witness counting algorithm fails to distinguish between two instances, one in which the densest k-subgraph has at most $\tilde{O}(k)$ edges, and one in which it has $\Omega(k^{1+1-\alpha}) = \Omega(k \cdot n^{\alpha(1-\alpha)})$ edges. They then describe an algorithm for general instances which achieves this exact tradeoff: when $k = n^{\alpha}$, their algorithm gives an $O(n^{\alpha(1-\alpha)+\varepsilon})$-approximation in time $n^{O(1/\varepsilon)}$.

It has been conjectured (see for example [13, 14]) that not only can the above approximation guarantee not be improved for worst-case instances, but even the Random vs Planted distinguishing problem cannot be solved below the same threshold:

▶ **Conjecture 2** (Planted Dense Subgraph Conjecture). *For all $0 < \alpha < 1$ and sufficiently small $\varepsilon > 0$, and for all $k \leqslant \sqrt{n}$,[3] no polynomial time algorithm can solve the Random vs Planted Subgraph problem for parameters $\beta \geqslant \alpha + \varepsilon$ with non-negligible probability.*

The above conjecture demonstrates another intriguing aspect of the log-density framework: not only that the framework leads to improved algorithms, it also leads to conjectured tight hardness results. For Densest k-Subgraph, this should be contrasted with the fact that no NP-hardness of approximation is known even for a factor of say 1.01! Moreover, while inapproximability results for DkS are known under stronger complexity assumptions [19, 26, 31, 1, 9, 29], none of them achieves a ratio of the form n^{δ} for some $\delta > 0$ and hence they are still very far from giving a tight lower bound as predicted by Conjecture 2.

The importance of such tight lower bounds is also amplified by the known connections between Densest k-Subgraph and numerous other problems. In particular, there are reductions from DkS to many problems, for which either hardness of approximation is not known at all otherwise or the known hardness is very weak compared to known approximations (see e.g. [22, 23, 10, 3, 4, 17, 11, 28, 12, 20, 25]). A hardness for DkS with a concrete inapproximability ratio, such one as established by Conjecture 2, would imply strong inapproximability results for these problems as well. In addition to applications in hardness of approximation, a variant of the conjecture has also been used in public-key cryptography [2].

Despite the aforementioned applications and importance of the conjecture, little progress has been made towards actually justifying it. This is somewhat unsurprising, since, barring few

---

[3] Note that, while for $k \geqslant \sqrt{n}$, a spectral algorithm can distinguish the two cases beyond the log-density threshold, the best *worst-case* approximation guarantees are still at the log-density threshold even for this regime.

exceptions, it is typically hard to argue for validity of average case assumptions. A common approach to support these hypotheses is by proving lower bounds against restricted classes of algorithms, such as certain LP and SDP relaxations; this has been done, for examples, for the planted clique hypothesis [18, 30, 24, 6] and the random 3-SAT hypothesis [21, 32].

Unfortunately, even on this front, not much is known for Densest k-Subgraph. In particular, the only (non-trivial) matching lower bound shown so far is that of Bhaskara et al. [8], who showed that $\Omega(\log n / \log \log n)$ rounds of the Sherali-Adams hierarchy [33] have a matching integrality gap only for the case of $\alpha = 1/2$ and $k = \sqrt{n}$, which happens to be the parameter setting maximizing the log-density gap for DkS. No tight integrality gaps were known for other parameter settings. The situation is not better for other problems that have been studied in the log-density framework; either no lower bound of this type is known for them at all, or a lower bound is known only for one specific instantiation of parameters. Note that evidence for more specific hardness of this form is crucial in hardness-of-approximation results based on DkS and related problems, as an optimal *reduction* may use a different parameter setting than the worst-case setting for DkS (e.g. [11]), or a different problem such as SpES (e.g. [28]) or SSBVE (e.g. [17]).

1.1 Our Results

We show that for every possible parameter setting, Sherali-Adams requires a super-constant number of rounds to distinguish between a random hypergraph, and one which contains a subhypergraph with the same log-density:

▶ **Theorem 3.** *[informal; see Theorem 9] For every $c = c(n) = O(\log \log n)$, every $k \in [n]$, every constant $\varepsilon > 0$ and every $\varepsilon < \alpha < c - 1$, for the random Erdős-Rényi c-uniform hypergraph $\mathcal{G}_c(n, n^{-\alpha})$, there is a $\tilde{\Omega}(\log n)$-round Sherali-Adams solution for the existence of a subhypergraph of size at most k, and at least $k^{c-\alpha-o(1)}$ edges, where both the bound on the size and the bound on the number of edges are lifted.*

As immediate corollaries of our main theorem, we obtain integrality gaps matching the log-density threshold for a variety of problems, which we outline here. Firstly, for DkS, we have an integrality gap matching the approximation guarantee of [7] for every log-density:

▶ **Corollary 4.** *For any constant $\alpha \in (0, 1)$, for $G \sim \mathcal{G}(n, n^{-\alpha})$ and $k = n^\alpha$, $\tilde{\Omega}(\log n)$ rounds of Sherali-Adams applied to a Densest k-Subgraph relaxation with a lifted objective function have an integrality gap of $n^{\alpha(1-\alpha)-o(1)}$ with high probability.*

In the Smallest p-Edge Subgraph problem (SpES), we are given a graph G and a parameter p, and are asked to find a subgraph of G with p edges and a minimum number of vertices. For this problem we get integrality gaps matching the approximation guarantee of [13]:

▶ **Corollary 5.** *For any constant $\alpha \in (0, 1)$, for $G \sim \mathcal{G}(n, n^{-\alpha})$ and $p = n^\alpha$, $\tilde{\Omega}(\log n)$ rounds of Sherali-Adams applied to a SpES relaxation with a lifted objective function have an integrality gap of $n^{\frac{\alpha(1-\alpha)}{2-\alpha}-o(1)}$ w.h.p. In particular, for $\alpha = 2 - \sqrt{2}$, we get an integrality gap of $n^{3-2\sqrt{2}-o(1)}$.*

We remark that the algorithms in [7, 13] can be stated in terms of the Sherali-Adams relaxations used in our work. Hence, we give an essentially tight (up to a sub-polynomial factor) integrality gaps for these relaxations of DkS and SpES for every setting of parameters.

Since our main theorem works not only for random graphs but also for random hypergraphs, our integrality gaps extend to the hypergraph variants of DkS and SpES. The hypergraph

extension of DkS, called Densest k-Subhypergraph (DkSH), is to find, given a c-uniform hypergraph G and an integer k, a subhypergraph of k vertices that contains as many hyperedges as possible. Our integrality gap for Densest k-Subhypergraph is stated below.

▶ **Corollary 6.** *For any constants $c \geqslant 2$ and $\alpha \in (0, c-1)$, for $G \sim \mathcal{G}_c(n, n^{-\alpha})$ and $k = n^{\alpha/(c-1)}$, $\tilde{\Omega}(\log n)$ rounds of Sherali-Adams applied to a DkSH relaxation with a lifted objective function have an integrality gap of $n^{\alpha(1-\alpha/(c-1))-o(1)}$ w.h.p. In particular, for $\alpha = (c-1)/2$, the integrality gap is $n^{(c-1)/4-o(1)}$.*

In the Minimum p-Union (MpU) problem, the goal is to find p hyperedges whose union is as small as possible in a given c-uniform hypergraph. Our integrality gap for MpU is stated below.

▶ **Corollary 7.** *For any constants $c \geqslant 2$ and $\alpha \in (0, c-1)$, for $G \sim \mathcal{G}_c(n, n^{-\alpha})$ and $p = n^{\alpha/(c-1)}$, $\tilde{\Omega}(\log n)$ rounds of Sherali-Adams applied to a MpU relaxation with a lifted objective function have an integrality gap of $n^{\frac{\alpha(c-1-\alpha)}{(c-1)(c-\alpha)}-o(1)}$ w.h.p. In particular, for $\alpha = c - \sqrt{c}$, we get an integrality gap of $n^{1-2/(1+\sqrt{c})-o(1)}$.*

We remark that our integrality gaps for both DkSH and MpU match the log-density threshold. However, unlike their graph counterparts DkS and SpES, no approximation algorithm whose ratios match the ones predicted by the log-density framework is known for DkSH and MpU for c-uniform hypergraphs where $c \geqslant 4$. For $c = 3$, such algorithms are known only for the semi-random instances [16] but not for the general (worst case) instances.

The extension to super-constant edge size also has implications for the recently studied Small Set Bipartite Vertex Expansion (SSBVE) problem. In this problem, we are given a bipartite graph $G = (L, R, E)$ and a parameter p, and are asked to find $U \subseteq L$ of size p with a minimum number of total neighbors in R. We also match the approximation guarantee of [14] for SSBVE:

▶ **Corollary 8.** *For any constant $\alpha \in (0, 1)$, there is an infinite family of bipartite graphs, such that for any bipartite graph $G = (L, R, E)$ in this family, and $p = |L|^\alpha$, $\omega(1)$ rounds of Sherali-Adams applied to an SSBVE relaxation with a lifted objective function have an integrality gap of $|L|^{\alpha(1-\alpha)-o(1)}$.*

Note that [14] also gave a Sherali-Adams integrality gap for the above problem, though only for $\alpha = 1/2$ (as in the integrality gap of [8]), and without lifting the objective function.

1.2   Related Work

As mentioned earlier, Sherali-Adams gaps for Densest k-Subgraph were studied in [8]. In their paper, they focused exclusively on the case of $\mathcal{G}(n, n^{-\alpha})$ and $k = n^\alpha$ for $\alpha = 1/2$. They defined an LP solution which, for every small vertex set $S \subseteq V$, assigns to the variable y_S some value which depends only on the size of the smallest Steiner tree for S. As they explain, such a solution is intuitive for the specific statistical properties of $\mathcal{G}(n, n^{-1/2})$. However this intuition, and their analysis, break down for any other value of α.

We also remark that, while lower bounds for the stronger Sum-of-Square (SoS) relaxations for DkS have been studied [8, 15], the integrality gaps given do not match the gaps predicted by log-density framework; in fact, the graphs used in these results are not random graphs but rather graphs resulting from a reduction from random instances of a constraint satisfaction problem (CSP). Hence, these SoS lower bounds are unrelated to the Planted Dense Subgraph Conjecture, and it is unlikely that the approach can yield similar gaps as predicted by the conjecture.

We turn instead to more recent work on integrality gaps for lift-and-project relaxations, namely the pseudo-calibration technique. This technique was introduced by Barak et al. [6], to give tight Sum-of-Squares integrality gaps for the closely related Planted Clique problem. The pseudo-calibration heuristic involves choosing a pseudo-distribution whose low-degree Fourier coefficients match those of a random graph with a planted solution. This technique is a very powerful heuristic, however, it does not automatically guarantee that the pseudo-distribution suggested by the technique will satisfy the required constraints. It is not clear at this point whether such a solution for Densest k-Subgraph and related problems is feasible as a Sherali-Adams solution, much less Sum-of-Squares. Our solution, however, is inspired by pseudo-calibration. We are able to modify the solution suggested by pseudo-calibration in such a way that in fact it does satisfy at least the Sherali-Adams constraints. This connection is explained in Appendix A.

1.3 Our Technique

In our analysis, we must show that three kinds of lifted constraints hold: monotonicity ($y_T \leqslant y_S$ for any $S \subseteq T \subseteq V$), size constraints (bounding the number of vertices selected by the solution), and density constraints (bounding the number of edges among selected vertices).

As explained below, for every small subset $S \subseteq V$, our solution assigns to y_S the maximum value attained for some function of $G|_{S'}$ over all possible vertex sets $S' \supseteq S$ up to a certain size bound. While at a first glance, both the monotonicity constraint and the density constraint seem to follow almost immediately from the definition, it turns out that this "immediate" proof of the density constraint relies on two different size bounds for S'. To reconcile the two conditions, we show that in fact in $\mathcal{G}(n, n^{-\alpha})$, w.h.p. any optimal set S' must already be bounded in size, thus making the difference in the requirements irrelevant.

The main technical aspect of our proof involves the size constraints. As we will show, a careful analysis of the properties of optimal sets does in fact guarantee the size constraints in expectation (over the random choices in $\mathcal{G}(n, n^{-\alpha})$). However, this is not enough, since we also need to show strong concentration (especially since this must hold for all possible small subsets $S \subseteq V$). Unfortunately, the general kind of concentration result we need is not true for random graphs.[4] We again strongly rely on the properties of optimal sets, showing that they *precisely* give the condition which guarantees concentration. This is shown by bounding the left-hand-side of the LP constraint by a polynomial in the incidence vector of $\mathcal{G}(n, n^{-\alpha})$, and applying a concentration bound of Kim and Vu [27] for low-degree polynomials.

Organization.   In Section 2, we will describe our solution for the Sherali-Adams relaxation and show that it satisfies the three aforementioned properties; this is the main contribution of our paper. Due to space constraint, we defer the descriptions of the relaxations for DkSH, MpU and SSBVE, and the parameters settings that give the claimed integrality gaps to the full version. Finally, we discuss several open questions in Section 3.

---

[4] As an example of this issue, consider the number of copies of H in $\mathcal{G}(n, n^{-3/5})$, where H is a clique of size 5 connected to a path of length 5. The expected number of copies of H in G is $n^{|V(H)|-3/5|E(H)|} = n^{10-3/5(15)} = n$. However, w.h.p. *no* copies of H will appear, since H also contains K_5, which only has an expected $n^{5-3/5(10)} = 1/n$ copies.

2   Our Sherali-Adams Solution and Its Properties

The goal of this section is to prove our main theorem, which is stated below.

▶ **Theorem 9.** *For any constants $0 < \beta, \varepsilon < 1$, any $c = c(n) = O(\log\log n)$ and any $\varepsilon < \alpha < c(n) - \varepsilon$, let $G = (V, E)$ be a hypergraph sampled from the Erdős-Rényi random c-uniform hypergraph distribution $\mathcal{G}_c(n, n^{-\alpha})$, then, w.h.p., there exists $\{y_S\}_{|S| \leqslant r}$ for some $r = \tilde{\Omega}(\log n)$ that satisfies the following:*

$$\forall S, T \subseteq V \text{ such that } |S| + |T| \leqslant r :$$

$$\sum_{i \in V} \sum_{J \subseteq T} (-1)^{|J|} y_{S \cup J \cup \{i\}} \leqslant n^{\beta + o(1)} \sum_{J \subseteq T} (-1)^{|J|} y_{S \cup J} \tag{1}$$

$$\sum_{e \in E} \sum_{J \subseteq T} (-1)^{|J|} y_{S \cup J \cup e} \geqslant n^{\beta(c-\alpha) - o(1)} \sum_{J \subseteq T} (-1)^{|J|} y_{S \cup J} \tag{2}$$

$$0 \leqslant \sum_{J \subseteq T} (-1)^{|J|} y_{S \cup J} \leqslant 1 \tag{3}$$

$$y_\emptyset = 1. \tag{4}$$

To illustrate the above constraints, it is perhaps best to think about the Densest k-Subhypergraph problem as a concrete example. In this case, y_i should be thought of as the indicator variable of whether the vertex i is selected as part of the solution (i.e. the subset of k vertices); more generally, y_S should be thought of as an indicator variable whether S is contained in the solution.

Constraint (1), which we refer to as the *size constraint* is the resulting lift of the constraint (i.e., "at most k vertices are selected") where $k = n^{\beta + o(1)}$. Now, recall that our objective is to maximize the number of induced hyperedges, i.e., to maximize $\sum_{e \in E} y_e$. Constraint (2) is the lift of the constraint $\sum_{e \in E} y_e \geqslant k^{(c-\alpha) - o(1)}$. That is of the requirement that the solution must induce at least $k^{(c-\alpha) - o(1)}$ hyperedges; we call this constraint the *density constraint.* In other words, the y_S's given in Theorem 9 behave as if they are a valid solution of the planted case of Conjecture 2 where a random graph from $\mathcal{G}_c(k, k^{-\alpha - o(1)})$ is planted into the instance. We remark also that in the typical LP relaxation, we also need to require that $y_e \leqslant y_i$ for all $i \in e$, but this is already implied by Constraint (3), which results from lifting the constraint $0 \leqslant y_S \leqslant 1$.

We will now proceed to prove Theorem 9, starting by describing our solution.

The Solution.   For all $S \subseteq V$ such that $|S| \leqslant r(n)$, we define y_S as follows:

$$y_S = \frac{1}{L^{|S|}} \max_{\substack{S' \supseteq S \\ |S'| \leqslant \tau(n)}} n^{(1-\beta)(\alpha'|E(S')| - |S'|)} \tag{5}$$

where L is some sub-polynomial dampening factor and $\alpha' = \alpha - o(1)$ is a number slightly smaller than α; we will define them more precisely later.

For brevity, for every $S \subseteq V$, let $\Phi(S)$ be $\alpha'|E(S)| - |S|$, and let $\Phi_{\text{OPT}}(S) = \max_{S' \supseteq S, |S'| \leqslant \tau(n)} \Phi(S')$. Moreover, we say that a subset S' is a *candidate set* for S if $S' \supseteq S$ and $|S'| \leqslant \tau(n)$. Note that our solution can be rewritten simply as $y_S = \frac{1}{L^{|S|}} \cdot n^{(1-\beta)\Phi_{\text{OPT}}(S)}$.

For every $S \subseteq V$ of cardinality at most $r(n)$, we use $\mathcal{S}'(S)$ to denote the collection of all candidate sets S' of S that attain the optimum in our solution y_S defined in (5). In other words, a set $S' \supseteq S$ of size at most $\tau(n)$ belongs to $\mathcal{S}'(S)$ if and only if $\Phi(S') = \Phi_{\text{OPT}}(S)$.

Parameter Selection. The exact parameters that we will use are as follows:

- $\tau = \tau(n) = \frac{\log n}{c(\log \log n)^2} = \tilde{\Theta}(\log n)$.
- $r = r(n) = \frac{\tau(n)}{(\log \log n)^2} = \tilde{\Theta}(\log n)$.
- $L = 2\tau(n) = \tilde{\Theta}(\log n)$.
- $\alpha' = \alpha \left(\frac{1 - 10r(n)/\tau(n)}{1 + 3c\tau(n)/\log n} \right) = \left(1 - O(1/(\log \log n)^2) \right) \alpha = \alpha - o(1)$.

2.1 Dampening Factor and Simplified Constraints

Our use of the dampening factor L is exactly the same as that in [8]; namely, it will allow us to simplify all constraints to the case where $T = \emptyset$. To see this, first observe that our solution has the following strong monotonicity property:

▶ **Lemma 10.** *For every $S \subseteq V$ and $i \in V$ such that $|S \cup \{i\}| \leqslant r(n)$, we have $y_S \geqslant L \cdot y_{S \cup \{i\}}$*

Proof. Since $S' \in \mathcal{S}'(S \cup \{i\})$ is a candidate set for S, $y_S \geqslant L^{-|S|} n^{(1-\beta)\Phi(S')} = L \cdot y_{S \cup \{i\}}$. ◀

We can then arrive at the following lemma, which states that $\sum_{J \subseteq T} (-1)^{|J|} y_{S \cup J}$ is within a factor of two of y_S, as stated below. Since the proof is exactly identical to the proof[5] of Lemma 3.2 of [8], we do not repeat it here.

▶ **Lemma 11** ([8]). *For any $S, T \subseteq V$ such that $|S \cup T| \leqslant r$ and $S \cap T = \emptyset$, we have*

$$y_S \geqslant \sum_{J \subseteq T} (-1)^{|J|} y_{S \cup J} \geqslant y_S/2.$$

Hence, the above lemma readily implies (3). Moreover, as observed in [8], we can "replace" each side in the constraints (1) and (2) by the case where $T = \emptyset$ and lose a factor of at most two each time. (Note that when $S \cap T \neq \emptyset$, both sides are already zero.) In other words, since these constant factors can be absorbed in the $o(1)$ term, it suffices to show that the two constraints hold when $T = \emptyset$. That is, we only need to show the following for all $S \subseteq V$ such that $|S| \leqslant r$:

$$\sum_{i \in V} y_{S \cup \{i\}} \leqslant n^{\beta + o(1)} y_S \tag{6}$$

$$\sum_{e \in E} y_{S \cup e} \geqslant n^{\beta(c - \alpha) - o(1)} y_S \tag{7}$$

$$y_\emptyset = 1 \tag{8}$$

2.2 Subhypergraph Exceeding the Log-Density Threshold and the Size Cutoff

Our analysis will rely heavily on the fact that w.h.p. random hypergraphs do not contain subhypergraphs with density strictly above the log-density threshold:

▶ **Lemma 12.** *With high probability, $|E(S)| \leqslant |S| \cdot \frac{(1 + 3c\tau/\log n)}{\alpha}$ for all $S \subseteq V$ of size at most τ.*

---

[5] Bhaskara et al.'s proof [8] requires $L \geqslant r(n)$; this also holds for our choice of parameter.

The lemma follows easily by bounding the number of labeled k-vertex hypergraphs with a sufficiently large number of edges, and the expected number of copies of each in a random hypergraph. The full calculation is deferred to Appendix B.

Note that since we select $\alpha' < \frac{\alpha}{1+3c\tau(n)/\log n}$, we can immediately conclude that, with probability $1 - o(1)$, $\Phi(S) \leqslant 0$ for all $S \subseteq V$ of size at most $\tau(n)$, which imply that $y_\emptyset = 1$:

▶ **Corollary 13.** *With high probability, $y_\emptyset = 1$.*

More importantly, Lemma 12 implies that the value of the cutoff $\tau(n)$ in fact does not matter! Specifically, as shown below, all sets S' that attain the maximum value of S in fact has size at most $\tau'(n) = \tau(n)/10$. Note that in the proof of the following lemma we need a separation between α' and α, which is the main reason behind our choice of α'.

▶ **Lemma 14.** *With high probability, $\mathcal{S}'(S) \subseteq \binom{V}{\leqslant \tau'(n)}$ for all $S \in \binom{V}{\leqslant r(n)}$.*

Proof. Let us assume that the high probability event from Lemma 12 occurs. Suppose for the sake of contradiction that there is some $S' \in \mathcal{S}'(S)$ of size $|S'| \in [\tau(n)/10, \tau(n)]$. Then in particular, S' is at least as good a set for S as S itself. That is, we have

$$
\begin{aligned}
0 \leqslant \Phi(S') - \Phi(S) &= \alpha'(|E(S')| - |E(S)|) - (|S'| - |S|) \\
&\leqslant \alpha'|E(S')| - (1 - |S|/|S'|)|S'| \\
&\leqslant \alpha'|E(S')| - (1 - 10r(n)/\tau(n))|S'|.
\end{aligned}
$$

But this implies that

$$
|E(S')| \geqslant |S'| \cdot (1 - 10r(n)/\tau(n))/\alpha' \geqslant |S'| \cdot (1 + 3c\tau/\log n)/\alpha,
$$

where the second inequality follows from our choice of α'. This contradicts our assumption that the high probability event in Lemma 12 occurs.   ◀

2.3   Proof of The Density Constraint

We will next prove constraint (7). Assume that the high probability event in Lemma 14 holds. Let $S' \in \mathcal{S}'(S)$. By our assumption, $|S'| \leqslant \tau'(n)$, which is at most $\tau(n) - c$ when n is sufficiently large. Hence, for every hyperedge $e \in E$, the set $S' \cup e$ is a candidate set for $S \cup e$, meaning that if $e \not\subseteq S'$, then

$$
y_{S \cup e} \geqslant L^{-|S|-c} n^{(1-\beta)\Phi(S' \cup e)} \geqslant L^{-|S|-c} n^{(1-\beta)(\Phi(S')+\alpha'-c)} = L^{-c} n^{(1-\beta)(\alpha'-c)} y_S.
$$

Moreover, observe that w.h.p. the total number of hyperedges in G (and not contained in S') is at least $(1 - o(1)) \cdot n^{c-\alpha}$. As a result, summing the above inequality over all $e \in E$ gives

$$
\sum_{e \in E} y_{S \cup e} \geqslant (1 - o(1)) L^{-c} n^{(c-\alpha)+(1-\beta)(\alpha'-c)} y_S = n^{\beta(c-\alpha)-o(1)} y_S.
$$

Degree constraints.   Algorithms in the log-density framework often rely on a bound on the minimum degree in (part of) an optimal solution. We note that when $\alpha < c - 1 - \varepsilon$ (that is, when G is dense enough not to have isolated vertices), our solution also satisfies the corresponding lifted constraints for every vertex. As before, due to the dampening factors, these constraints are equivalent to the following:

$$
\sum_{e \in E: e \ni i} y_{S \cup e} \geqslant n^{\beta(c-1-\alpha)-o(1)} y_S \qquad \forall S \subseteq V \text{ such that } |S| \leqslant r, \forall i \in S \tag{9}
$$

The proof follows much the same argument. As above, for $S' \in \mathcal{S}'(S)$, node $i \in S(\subseteq S')$, and hyperedge $e \not\subseteq S'$ that contains i, it follows by a similar calculation that $y_{S \cup e} \geqslant L^{-c+1} n^{\alpha'-c+1}$. Combining this bound with the observation that w.h.p. the number of hyperedges containing i is at least $(1 - o(1))n^{c-1-\alpha}$, we obtain the above constraint.

2.4 Proof of The Size Constraint

We now turn our attention to the only remaining constraint: the size constraint (6). The proof of the size constraint is more complicated than the previous constraints and, before we get to the main argument, we will prove a couple more structural properties of the optimal candidate sets.

2.4.1 Structural Properties of Optimal Candidate Sets

Here, we make a number of useful observations regarding the structure of $\mathcal{S}'(S)$ and the sets it contains. First, observe that $\Phi(\cdot)$ is supermodular; this will be useful in the sequel.

▶ **Observation 15.** *For every $S_1, S_2 \subseteq V$, $\Phi(S_1 \cup S_2) + \Phi(S_1 \cap S_2) \geqslant \Phi(S_1) + \Phi(S_2)$.*

Proof. Observe that $|E(S_1 \cup S_2)| + |E(S_1 \cap S_2)| \geqslant |E(S_1)| + |E(S_2)|$ and $|S_1 \cup S_2| + |S_1 \cap S_2| = |S_1| + |S_2|$. Subtracting the two yields the above bound. ◀

From this point on, we will assume that the high probability event in Lemma 14 occurs, i.e., that all optimal candidate sets (for all $S \subseteq V$ of size at most $r(n)$) are of size at most $\tau'(n)$.

We call a collection $\mathcal{S} \subseteq \mathcal{P}(V)$ *union closed*, if, for every $S_1, S_2 \in \mathcal{S}$, $S_1 \cup S_2 \in \mathcal{S}$.

▶ **Lemma 16.** *For all $S \subseteq V$ of size at most $r(n)$, $\mathcal{S}'(S)$ is union closed.*

Proof. Consider any set $S_1, S_2 \in \mathcal{S}'(S)$. Since $S \subseteq S_1, S_2$ and $|S_1|, |S_2| \leqslant \tau'(n)$, $S_1 \cap S_2$ must also be a candidate set for S, which implies that $\Phi(S_1 \cap S_2) \leqslant \Phi_{\mathrm{OPT}}(S) = \Phi(S_1) = \Phi(S_2)$. As a result, from Observation 15, we have $\Phi(S_1 \cup S_2) \geqslant \Phi(S_1) + \Phi(S_2) - \Phi(S_1 \cap S_2) \geqslant \Phi_{\mathrm{OPT}}(S)$. Furthermore, $|S_1 \cup S_2| \leqslant |S_1| + |S_2| \leqslant 2\tau'(n) \leqslant \tau(n)$, meaning that $S_1 \cup S_2$ is a candidate set for S as well. Hence, $S_1 \cup S_2$ must also be in $\mathcal{S}'(S)$. ◀

Union closed collection of sets always contains a maximal set, which is simply the union of all the sets. When $\mathcal{S}'(S)$ is union closed, we denote its maximal set by $S'_{\max}(S)$. The following lemma relates the maximal optimal candidate set of $S \cup \{i\}$ to that of S:

▶ **Lemma 17.** *For all $S \subseteq V$ of size at most $r(n)$ and for all $i \in V$, $S'_{\max}(S) \subseteq S'_{\max}(S \cup \{i\})$.*

Proof. The proof is essentially the same as that of Lemma 16. Consider any $S'' \in \mathcal{S}'(S \cup \{i\})$. Since $S'_{\max}(S)$ is optimal for S and $S'_{\max}(S) \cap S''$ is also a candidate set for S, we have $\Phi(S'_{\max}(S)) \geqslant \Phi(S'_{\max}(S) \cap S'')$. Similar to the calculation in the previous lemma, the supermodularity of Φ implies that $\Phi(S'_{\max}(S) \cup S'') \geqslant \Phi(S'')$. Moreover, since $|S'_{\max}(S) \cup S''| \leqslant 2\tau'(n) \leqslant \tau(n)$, $S'_{\max}(S) \cup S''$ is a valid candidate set for S, meaning that $(S'_{\max}(S) \cup S'') \in \mathcal{S}'(S \cup \{i\})$ as desired. ◀

2.4.2 Setting up the Argument

We are now ready to bound $\sum_{i \in V} y_{S \cup \{i\}}$. For brevity, we write $S_{\max}$ to denote $S'_{\max}(S)$ and $S^i_{\max}$ to denote $S'_{\max}(S \cup \{i\})$ for every $i \in V$. Let $\mathcal{I}_{\mathrm{iso}}$ be the set of $i \in V$ such that i is an isolated vertex in the subhypergraph $E(S^i_{\max})$. It is not hard to show the following:

▶ **Proposition 18.**
1. *For every $i \in \mathcal{I}_{iso} \setminus S_{\max}$, we have $S_{\max} \cup \{i\} \in \mathcal{S}'(S \cup \{i\})$.*
2. *For every $i \notin \mathcal{I}_{iso}$, no vertex outside of S is isolated in the subhypergraph $E(S_{\max}^i)$.*

Proof.
1. Suppose for the sake of contradiction that $i \in \mathcal{I}_{iso} \setminus S_{\max}$ but $(S_{\max} \cup \{i\}) \notin \mathcal{S}'(S \cup \{i\})$. This implies

$$\Phi_{\mathrm{OPT}}(S) = \Phi(S_{\max}) \leqslant \Phi(S_{\max} \cup \{i\}) - 1 < \Phi(S_{\max}^i) - 1 = \Phi(S_{\max}^i \setminus \{i\}) \leqslant \Phi_{\mathrm{OPT}}(S),$$

where the second inequality comes from $S_{\max} \cup \{i\} \notin \mathcal{S}'(S \cup \{i\})$ and the next equality comes from our assumption that i is an isolated vertex in $E(S_{\max}^i)$ and the last inequality follows from the fact that $S_{\max}^i \setminus \{i\}$ is a candidate set for S. Thus, we have arrived at a contradiction.
2. Suppose for the sake of contradiction that $i \notin \mathcal{I}_{iso}$ but there exists $j \in S_{\max}^i \setminus S$ such that j is isolated in $E(S_{\max}^i)$. From $i \notin \mathcal{I}_{iso}$, we have $j \neq i$; this means that $S_{\max}^i \setminus \{j\}$ is a candidate set for $S \cup \{i\}$ but this implies that

$$\Phi_{\mathrm{OPT}}(S \cup \{i\}) = \Phi(S_{\max}^i) = \Phi(S_{\max}^i \setminus \{j\}) - 1 \leqslant \Phi_{\mathrm{OPT}}(S \cup \{i\}) - 1,$$

which is a contradiction. ◀

Using property 1 of Proposition 18, we can bound $\sum_{i \in V} y_{S \cup \{i\}}$ as follows.

▶ **Claim 19.** *The following bound holds:*
$$\sum_{i \in V} y_{S \cup \{i\}} \leqslant y_S \left(1 + n^\beta + \sum_{i \notin S_{\max} \cup \mathcal{I}_{iso}} \frac{n^{(1-\beta)\alpha'|E(S_{\max}^i) \setminus E(S_{\max})|}}{n^{(1-\beta)|S_{\max}^i \setminus S_{\max}|}} \right).$$

Proof.

$$\sum_{i \in V} y_{S \cup \{i\}} = \sum_{i \in S_{\max}} y_{S \cup \{i\}} + \sum_{i \notin S_{\max}} y_{S \cup \{i\}}$$

$$= |S_{\max}| \cdot \frac{y_S}{L} + \sum_{i \notin S_{\max}} \frac{1}{L^{|S| \cup \{i\}}} \cdot \frac{n^{(1-\beta)\alpha'|E(S_{\max}^i)|}}{n^{(1-\beta)|S_{\max}^i|}}$$

$$\leqslant y_S \left(\frac{\tau(n)}{L} + \sum_{i \notin S_{\max}} \frac{n^{(1-\beta)\alpha'|E(S_{\max}^i) \setminus E(S_{\max})|}}{n^{(1-\beta)|S_{\max}^i \setminus S_{\max}|}} \right)$$

$$\leqslant y_S \left(1 + \sum_{i \notin S_{\max}} \frac{n^{(1-\beta)\alpha'|E(S_{\max}^i) \setminus E(S_{\max})|}}{n^{(1-\beta)|S_{\max}^i \setminus S_{\max}|}} \right)$$

$$\leqslant y_S \left(1 + |\mathcal{I}_{iso}| \cdot \frac{1}{n^{1-\beta}} + \sum_{i \notin S_{\max} \cup \mathcal{I}_{iso}} \frac{n^{(1-\beta)\alpha'|E(S_{\max}^i) \setminus E(S_{\max})|}}{n^{(1-\beta)|S_{\max}^i \setminus S_{\max}|}} \right) \quad \text{by Proposition 18}$$

$$\leqslant y_S \left(1 + n^\beta + \sum_{i \notin S_{\max} \cup \mathcal{I}_{iso}} \frac{n^{(1-\beta)\alpha'|E(S_{\max}^i) \setminus E(S_{\max})|}}{n^{(1-\beta)|S_{\max}^i \setminus S_{\max}|}} \right)$$

◀

To bound the remaining sum in the above right hand side expression, we will need some additional notations and observations.

First, observe that, from the optimality of $S_{\max}^i$ for $S \cup \{i\}$, $\Phi(S_{\max}^i)$ must be at least $\Phi(S_{\max} \cup \{i\}) \geqslant \Phi(S_{\max}) - 1$, which implies the following:

▶ **Observation 20.** *For every* $i \notin S_{\max}$, $\alpha'|E(S_{\max}^i) \setminus E(S_{\max})| - |S_{\max}^i \setminus S_{\max}| \geqslant -1$.

Next, we define an α'-strictly balanced (α'-s.b.) sets of hyperedges with respect to $S_{\max}$. Intuitively speaking, a set T of hyperedges is α'-s.b. with respect to $S_{\max}$ if we cannot select a subset $T' \subseteq S_{\max}$ such that we can add T' into $S_{\max}$ and create a new set with higher Φ than $\Phi(S_{\max})$. This is formalized below.

▶ **Definition 21.** For each $T \subseteq \binom{V}{c}$ such that $T \cap \binom{S_{\max}}{c} = \emptyset$, T is said to be α'-*strictly balanced* (α'-s.b.) with respect to $S_{\max} \subseteq V$ if for any $T' \subseteq T$, we have $|v(T') \setminus S_{\max}| > \alpha'|T'|$.

The optimality of $S_{\max}$ for S implies the following:

▶ **Observation 22.** *For every* $T \subseteq E_G$ *such that* $T \cap E(S_{\max}) = \emptyset$ *and* $|v(T) \setminus S_{\max}| \leqslant \tau(n) - \tau'(n)$, T *is* α'-*strictly balanced with respect to* $S_{\max}$.

For convenience, we say that for $S^* \subseteq V$, a hypergraph H with vertices S^* satisfies condition (*) if the following conditions hold:

- $S^* \supseteq S_{\max}$,
- $E(H) \setminus E(S_{\max})$ is α'-strictly balanced with respect to $S_{\max}$,
- $\alpha'|E_H(S^*) \setminus E(S_{\max})| - |S^* \setminus S_{\max}| \geqslant -1$
- Every $v \in S^* \setminus S_{\max}$ is not isolated in $E_H(S^*)$.

With the above two observations and the second property of Proposition 18, we can further bound the remaining sum in Claim 19 as follows.

$$\sum_{i \notin S_{\max} \cup \mathcal{I}_{\mathrm{iso}}} \frac{n^{(1-\beta)\alpha'|E(S_{\max}^i) \setminus E(S_{\max})|}}{n^{(1-\beta)|S_{\max}^i \setminus S_{\max}|}} \leqslant \sum_{\substack{S^* \in \binom{V}{\leqslant \tau'(n)} \\ G|_{S^*} \text{ satisfies } (*)}} \sum_{\substack{i \notin S_{\max} \cup \mathcal{I}_{\mathrm{iso}} \\ S_{\max}^i = S^*}} \frac{n^{(1-\beta)\alpha'|E(S^*) \setminus E(S_{\max})|}}{n^{(1-\beta)|S^* \setminus S_{\max}|}}$$

$$\leqslant \tau'(n) \cdot \sum_{\substack{S^* \in \binom{V}{\leqslant \tau'(n)} \\ G|_{S^*} \text{ satisfies } (*)}} \frac{n^{(1-\beta)\alpha'|E(S^*) \setminus E(S_{\max})|}}{n^{(1-\beta)|S^* \setminus S_{\max}|}}$$

Since $\tau'(n) = n^{o(1)}$, we will focus on bounding the final sum

$$\sum_{\substack{S^* \in \binom{V}{\leqslant \tau'(n)} \\ G|_{S^*} \text{ satisfies } (*)}} \frac{n^{(1-\beta)\alpha'|E(S^*) \setminus E(S_{\max})|}}{n^{(1-\beta)|S^* \setminus S_{\max}|}} \tag{10}$$

As we shall see, in expectation, this expression is bounded from above by $n^{\beta + o(1)}$, as required. However, this is not enough for us: we want this sum to be bounded by $n^{\beta + o(1)}$ for all choices of S's simultaneously. To do so, we will first prove a concentration for a fixed S and then use a union bound over all $n^{O(r(n))}$ choices of S's. We begin with an intuitive overview of the proof, before giving a more formal proof in Appendix C.1.

Bounding the Expectation. To bound (10) in expectation, we consider every possible hypergraph of size $\tau'(n)$ which could appear as a subhypergraph containing $S_{\max}$. Consider a hypergraph "template" $H_{X,T} = (S_{\max} \cup X, T \cup E(S_{\max}))$, where X are (at most) $\tau'(n) - |S_{\max}|$ dummy vertices, and $T \subseteq \binom{S_{\max} \cup X}{c} \setminus \binom{S_{\max}}{c}$. Then (10) can be bounded from above by

$$\sum_{\substack{X, T \text{ as above} \\ H_{X,T} \text{ satisfies } (*)}} \frac{n^{(1-\beta)\alpha'|T|} \cdot |\{S^* \supseteq S_{\max} : H_{X,T} \text{ is isomorphic to a subgraph of } G|_{S^*}\}|}{n^{(1-\beta)|X|}} \tag{11}$$

To bound the expectation (under the random choices of $\mathcal{G}_c(n, n^{-\alpha})$), it suffices to bound the expected number of copies of each $H_{X,T}$ in the above expression. This expectation is easily seen to be $n^{|X|-\alpha|T|} \leq n^{|X|-\alpha'|T|}$, allowing us to bound the expectation of (11) by

$$\sum_{\substack{X,T \text{ as above} \\ H_{X,T} \text{ satisfies } (*)}} \frac{n^{(1-\beta)\alpha'|T|}}{n^{(1-\beta)|X|}} \cdot \frac{n^{|X|}}{n^{\alpha'|T|}} = \sum_{\substack{X,T \text{ as above} \\ H_{X,T} \text{ satisfies } (*)}} n^{\beta(|X|-\alpha'|T|)}$$

$$\leqslant \sum_{\substack{X,T \text{ as above} \\ H_{X,T} \text{ satisfies } (*)}} n^{\beta}. \qquad \text{by the third condition in } (*)$$

Since the number of possible template hypergraphs $H_{X,T}$ of size $\tau'(n)$ is again subpolynomial, in expectation, we have our bound of $n^{\beta+o(1)}$.

Proving Concentration around the Expectation.   As mentioned earlier, we will formally prove that concentration holds by expressing (10) as a low-degree polynomial, and applying a concentration bound of Kim and Vu [27]. However, this application is not straightforward. Not only is a large expectation not enough for our analysis, it is not true for general sets that we would even necessarily have any meaningful concentration. We must use the fact that $S_{\max}$ is an optimal set. Specifically, that we only consider α-strictly balanced subgraphs containing $S_{\max}$.

To see why this is important, let us consider a template hypergraph $H_{X,T}$ for $S_{\max}$. Recall that the expected number of copies of $H_{X,T}$ in G is $n^{|X|-\alpha|T|}$. While it may be that $|X| > \alpha|T|$, giving us a polynomially large expectation, it is still possible (from a purely graph-theoretic perspective) that for some set X' s.t. $S_{\max} \subseteq X' \subset X$, the induced subtemplate $H_{X',T'}$ (where $T' = T \cap E(H|_{X' \cup S_{\max}})$) does not satisfy this, but rather has $|X'| < \alpha|T'|$. In this case, rather than concentration, we would have a case where with high probability there is no copy of $H_{X,T}$ (since in particular w.h.p. there is no copy of $H_{X',T'}$). This does not seem bad for us, since there are even far fewer copies than expected, and we would like to get an upper bound. However, the flip side of this situation is that with $n^{-\Omega(1)}$ probability, there are actually polynomially many times *more* copies of $H_{X,T}$ than the expectation, and a polynomially small probability is not good enough, since our bound must hold simultaneously for all $n^{\Omega(r(n))}$ small sets S.

Fortunately, the α-balanced condition precisely states that for any subtemplate $H_{X',T'}$ we in fact have $|X'| \geqslant \alpha|T'|$, so this situation will not arise. Thus concentration is not ruled out, and in fact can be formally proven using the Kim-Vu bound; due to space constraint, we defer this to Appendix C.

3   Discussions and Open Questions

Having shown a general technique for proving Sherali-Adams gaps at the log-density threshold, it would now be interesting to see if our approach can be applied to more rigidly structured problems that have been considered in the log-density framework, such as Label Cover.

Furthermore, an exciting challenge that could further bolster the conjectured hardness would be to give matching lower-bounds in the Sum-of-Squares SDP hierarchy. However, note that for some parameter regimes, such an integrality gap does not hold even for a simple SDP! As shown in [7], for $1 > \alpha > 1/2$, a simple SDP relaxation for DkS applied to $G = \mathcal{G}(n, n^{-\alpha})$ can already witness the fact that G does not contain a $k = n^{\alpha}$-subgraph with more than $n^{(1+\alpha)/2} \ll k^{2-\alpha}$ edges. On the other hand, in the worst case, the best currently-known approximation for this value of k is still $k^{1-\alpha}$ (matching the log-density

threshold). Thus, at least for certain parameter regimes, it would be very interesting to prove either of the following two possibilities, both of which are consistent with the current state of our knowledge:

1. There are algorithms for certain parameter regimes which give (worst-case) approximation guarantees strictly better than the log-density gap.
2. There is some other family of instances (other than Erdős-Rényi (hyper)graphs) for which there is a hardness of approximation, or at least Sum-of-Squares integrality gap, matching the log-density threshold.

References

1. Noga Alon, Sanjeev Arora, Rajsekar Manokaran, Dana Moshkovitz, and Omri Weinstein. Inapproximabilty of densest k-subgraph from average case hardness. Unpublished Manuscript, 2011.

2. Benny Applebaum, Boaz Barak, and Avi Wigderson. Public-key cryptography from different assumptions. In *STOC*, pages 171–180, 2010. `doi:10.1145/1806689.1806714`.

3. Sanjeev Arora, Boaz Barak, Markus Brunnermeier, and Rong Ge. Computational complexity and information asymmetry in financial products. *Commun. ACM*, 54(5):101–107, 2011. `doi:10.1145/1941487.1941511`.

4. Pranjal Awasthi, Moses Charikar, Kevin A. Lai, and Andrej Risteski. Label optimal regret bounds for online local learning. In *COLT*, pages 150–166, 2015. URL: `http://jmlr.org/proceedings/papers/v40/Awasthi15a.html`.

5. Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series*, 19(3):357–367, 1967.

6. Boaz Barak, Samuel B. Hopkins, Jonathan Kelner, P. Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *FOCS*, pages 428–437, 2016.

7. Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an $O(n^{1/4})$ approximation for densest k-subgraph. In *STOC*, pages 201–210, 2010. `doi:10.1145/1806689.1806718`.

8. Aditya Bhaskara, Moses Charikar, Aravindan Vijayaraghavan, Venkatesan Guruswami, and Yuan Zhou. Polynomial integrality gaps for strong SDP relaxations of densest k-subgraph. In *SODA*, pages 388–405, 2012. URL: `http://dl.acm.org/citation.cfm?id=2095116.2095150`.

9. Mark Braverman, Young Kun-Ko, Aviad Rubinstein, and Omri Weinstein. ETH hardness for densest-k-subgraph with perfect completeness. In *SODA*, pages 1326–1341, 2017.

10. Moses Charikar, MohammadTaghi Hajiaghayi, and Howard J. Karloff. Improved approximation algorithms for label cover problems. *Algorithmica*, 61(1):190–206, 2011. `doi:10.1007/s00453-010-9464-3`.

11. Moses Charikar, Yonatan Naamad, and Anthony Wirth. On approximating target set selection. In *APPROX*, pages 4:1–4:16, 2016. `doi:10.4230/LIPIcs.APPROX-RANDOM.2016.4`.

12. Stephen R. Chestnut and Rico Zenklusen. Hardness and approximation for network flow interdiction. *Networks*, 69(4):378–387, 2017. `doi:10.1002/net.21739`.

13. Eden Chlamtac, Michael Dinitz, and Robert Krauthgamer. Everywhere-sparse spanners via dense subgraphs. In *FOCS*, pages 758–767, 2012. `doi:10.1109/FOCS.2012.61`.

14. Eden Chlamtác, Michael Dinitz, and Yury Makarychev. Minimizing the union: Tight approximations for small set bipartite vertex expansion. In *SODA*, pages 881–899, 2017. `doi:10.1137/1.9781611974782.56`.

15   Eden Chlamtác, Pasin Manurangsi, Dana Moshkovitz, and Aravindan Vijayaraghavan. Approximation algorithms for label cover and the log-density threshold. In *SODA*, pages 900–919, 2017. `doi:10.1137/1.9781611974782.57`.

16   Eden Chlamtáč and Pasin Manurangsi. Tight approximations in the semirandom log-density framework via label reduction. In preparation.

17   Julia Chuzhoy, Yury Makarychev, Aravindan Vijayaraghavan, and Yuan Zhou. Approximation algorithms and hardness of the k-route cut problem. *ACM Trans. Algorithms*, 12(1):2:1–2:40, 2016. `doi:10.1145/2644814`.

18   Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *COLT*, pages 523–562, 2015. URL: `http://jmlr.org/proceedings/papers/v40/Deshpande15.html`.

19   Uriel Feige. Relations between average case complexity and approximation complexity. In *STOC*, pages 534–543, 2002. `doi:10.1145/509907.509985`.

20   Uriel Feige and Yael Hitron. The ordered covering problem. *Algorithmica*, Aug 2017. `doi:10.1007/s00453-017-0357-6`.

21   Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001. `doi:10.1016/S0304-3975(00)00157-2`.

22   Mohammad Taghi Hajiaghayi and Kamal Jain. The prize-collecting generalized steiner tree problem via a new approach of primal-dual schema. In *SODA*, pages 631–640, 2006. URL: `http://dl.acm.org/citation.cfm?id=1109557.1109626`.

23   Mohammad Taghi Hajiaghayi, Kamal Jain, Lap Chi Lau, Ion I. Mandoiu, Alexander Russell, and Vijay V. Vazirani. Minimum multicolored subgraph problem in multiplex PCR primer set selection and population haplotyping. In *ICCS*, pages 758–766, 2006. `doi:10.1007/11758525_102`.

24   Samuel B. Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm. On the integrality gap of degree-4 sum of squares for planted clique. In *SODA*, pages 1079–1095, 2016. `doi:10.1137/1.9781611974331.ch76`.

25   Jonah Kallenbach, Robert Kleinberg, and Scott Duke Kominers. Orienteering for electioneering. *Operations Research Letters*, 46(2):205–210, 2018. `doi:10.1016/j.orl.2017.10.013`.

26   Subhash Khot. Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J. Comput.*, 36(4):1025–1071, 2006. `doi:10.1137/S0097539705447037`.

27   Jeong Han Kim and Van H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000. `doi:10.1007/s004930070014`.

28   Euiwoong Lee. Partitioning a graph into small pieces with applications to path transversal. In *SODA*, pages 1546–1558, 2017. `doi:10.1137/1.9781611974782.101`.

29   Pasin Manurangsi. Almost-polynomial ratio ETH-hardness of approximating densest k-subgraph. In *STOC*, pages 954–961, 2017. `doi:10.1145/3055399.3055412`.

30   Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. In *STOC*, pages 87–96, 2015. `doi:10.1145/2746539.2746600`.

31   Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *STOC*, pages 755–764, 2010. `doi:10.1145/1806689.1806792`.

32   Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *FOCS*, pages 593–602, 2008. `doi:10.1109/FOCS.2008.74`.

33   Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3(3):411–430, 1990. `doi:10.1137/0403036`.

A   Pseudo-Calibration for Densest k-Subhypergraph

Let $n, k \in \mathbb{N}$ and $p, q \in [0, 1]$ be parameters to be chosen later.

- Let $\mathcal{G}_{\text{random}} = \mathcal{G}_c(n, p)$ denote the distribution of Erdős-Rényi random c-uniform hypergraph on n vertices where each c-size set of vertices is a hyperedge w.p. p.
- Let $\mathcal{G}_{\text{planted}}$ denote the "planted" distribution where each of the n vertices is marked as in the planted solution independently with probability k/n. For each c-size set of vertices, if all of its elements are marked as in the planted solution, then they are joined by a hyperedge with probability q. Otherwise, they are joined with probability p.

We represent a graph G as $\left\{ \frac{1-p}{\sqrt{p(1-p)}}, \frac{-p}{\sqrt{p(1-p)}} \right\}^{\binom{[n]}{c}}$ where $G_e = \frac{1-p}{\sqrt{p(1-p)}}$ if the hyperedge e exists and in the graph and $G_e = \frac{-p}{\sqrt{p(1-p)}}$ otherwise. As usual, for each $T \subseteq \binom{[n]}{c}$, let $\chi_T(G) = \prod_{e \in T} G_e$. Observe that

$$\mathbb{E}_{G \sim \mathcal{G}_{\text{random}}} [\chi_T(G)\chi_{T'}(G)] = \begin{cases} 1 & \text{if } T = T', \\ 0 & \text{otherwise.} \end{cases}$$

Finally, we use $x_u \in \{0, 1\}$ to denote the indicator variable of whether $u \in [n]$ is marked as planted. As usual, we define $x_S = \prod_{u \in S} x_u$

The Pseudo-Calibration heuristic suggests the following pseudo-distribution as a candidate solution.

$$\tilde{\mathbb{E}}[x_S](G) = \sum_{\substack{T \subseteq \binom{[n]}{c} \\ |v(T) \cup S| \leqslant \tau}} \widehat{\tilde{\mathbb{E}}[x_S]}(T)\chi_T(G)$$

where

$$\widehat{\tilde{\mathbb{E}}[x_S]}(T) = \mathbb{E}_{(x,G) \sim \mathcal{G}_{\text{planted}}} [x_S \chi_T(G)] = \left(\frac{k}{n} \right)^{|v(T) \cup S|} \mathbb{E}_{(x,G) \sim \mathcal{G}_{\text{planted}}} [\chi_T(G) \mid x_{v(T) \cup S} = 1]$$

$$= \left(\frac{k}{n} \right)^{|v(T) \cup S|} \left(\frac{q - p}{\sqrt{p(1-p)}} \right)^{|T|}$$

$$\approx \left(\frac{k}{n} \right)^{|v(T) \cup S|} \left(\frac{q}{\sqrt{p}} \right)^{|T|}.$$

In other words, the suggested solution is

$$\tilde{\mathbb{E}}[x_S](G) = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |v(T) \cup S| \leqslant \tau}} \left(\frac{k}{n} \right)^{|v(T) \cup S|} \left(\frac{q}{\sqrt{p}} \right)^{|T|} \chi_T(G).$$

We will further approximate this by replacing $\chi_T(G)$ with $(1/\sqrt{p})^{|T|} \mathbb{1}[T \subseteq E_G]$. This gives the solution

$$\tilde{\mathbb{E}}[x_S](G) = \sum_{\substack{T \subseteq E_G \\ |v(T) \cup S| \leqslant \tau}} \left(\frac{k}{n} \right)^{|v(T) \cup S|} \left(\frac{q}{p} \right)^{|T|}.$$

Our solution (specified in (5)) is the same as the above expression except that we change the summation to maximization, and we add the dampening factor and some $o(1)$ slack in the exponent; note that the parameters there are $p = n^\alpha, k = n^\beta, q = k^\alpha = n^{\alpha\beta}$.

B   Proof of Lemma 12

Proof. Note that this bound holds trivially for any S of size at most $c - 1$ since $|E(S)| = 0$. For succinctness, let $\tilde{A} = (1 + 3c\tau/\log n)/\alpha$. For every $k \in \{c, \dots, \tau\}$, the number of labeled k-vertex hypergraphs H with $t(k) = \lceil k\tilde{A} \rceil$ hyperedges is at most $\binom{k^c}{t(k)} \leqslant k^{c \cdot t(k)}$. Since the expected number of copies of such an H in G is $n^{|V(H)| - \alpha|E(H)|} = n^{k - \alpha t(k)} \leqslant 2^{-3c\tau k}$, we have

$$\Pr[\exists H \text{ as above} : H \text{ is a subhypergraph of } G] < \sum_{H \text{ as above}} \mathbb{E}[\#\text{copies of } H \text{ in G}]$$

$$\leqslant \sum_{k=c}^{\tau} \sum_{\substack{H \text{ as above} \\ |V(H)|=k}} 2^{-3c\tau k}$$

$$\leqslant \sum_{k=c}^{\tau} k^{c \cdot t(k)} 2^{-3c\tau k}$$

$$= \sum_{k=c}^{\tau} 2^{c(t(k) \log k - 3\tau k)}$$

$$\leqslant \sum_{k=c}^{\tau} 2^{c((1 + k\tilde{A}) \log k - 3\tau k)}$$

$$\leqslant \sum_{k=c}^{\tau} 2^{ck(\tilde{A} \log k - 2\tau)}$$

$$\leqslant \sum_{k=c}^{\tau} 2^{ck(\tilde{A} \log \tau - 2\tau)}.$$

Now, observe that, for sufficiently large n, we have $\tilde{A} \log \tau \leqslant \tau$. As a result, we have

$$\Pr[\exists H \text{ as above} : H \text{ is a subgraph of } G] \leqslant \sum_{k=c}^{\tau} 2^{-ck\tau} \leqslant 2^{1 - c^2\tau} \leqslant 2^{-c^2\tau/2} = o(1),$$

which concludes our proof.   ◀

C   Proof of The Size Constraint

C.1   Concentration of Low Degree Multilinear Polynomials.

Recall that we would like to bound (10). For each $e \in \binom{V}{c}$, let X_e denote the indicator variable $\mathbb{1}[e \in E_G]$ for $G = \mathcal{G}_c(n, n^{-\alpha})$. Then (10) can easily be bounded by a multilinear polynomial in these random variables X_e's as follows:

$$\sum_{\substack{S^* \in \binom{V}{\leqslant \tau'(n)} \\ G|_{S^*} \text{ satisfies } (*)}} \frac{n^{(1-\beta)\alpha'|E(S^*) \setminus E(S_{\max})|}}{n^{(1-\beta)|S^* \setminus S_{\max}|}}$$

$$\leqslant \sum_{S^* \in \binom{V}{\leqslant \tau'(n)}} \frac{1}{n^{(1-\beta)|S^* \setminus S_{\max}|}} \cdot \left(\sum_{\substack{T \subseteq \binom{S^*}{c} \setminus \binom{S_{\max}}{c} \\ (S^*, T) \text{ satisfies } (*)}} \mathbb{1}[T \subseteq E_G] \cdot n^{(1-\beta)\alpha'|T|} \right)$$

$$= \sum_{\substack{T \subseteq \binom{S^*_c}{c} \setminus \binom{S_{\max}}{c} \\ (S^*, T) \text{ satisfies } (*)}} \frac{n^{(1-\beta)\alpha'|T|}}{n^{(1-\beta)|v(T)\setminus S_{\max}|}} \prod_{e \in T} X_e \tag{12}$$

Denote the polynomial in (12) by $F((X_e)_{e \in \binom{V}{c}})$. Note that F has degree $t \leqslant \tau'(n)/\alpha'$ since $|v(T) \cup S_{\max}| \leqslant \tau'(n)$ and T being strictly balanced implies that $|T| \leqslant |v(T) \setminus S_{\max}|/\alpha' \leqslant \tau'(n)/\alpha'$.

To show that this quantity is not too large, we will resort to the following concentration bound of low degree multilinear polynomial of Kim and Vu [27], which applies for any low degree polynomials whose *expectation* of every partial derivative is small. We note that this condition is weaker than those of some other similar inequalities, such as Azuma's [5], which requires the *maximum* (of partial derivative or some other measures of effect) to be small.

▶ **Theorem 23** (Kim-Vu Concentration Bound [27]). *Let $f(x_1, \ldots, x_N)$ be any degree-t multilinear polynomial, i.e.,*

$$f(x_1, \ldots, x_N) = \sum_{T \in \mathcal{T}} w_T \prod_{i \in T} x_i$$

where $\mathcal{T}$ is a collection of subsets of $[N]$ of size at most t and $w_T \geqslant 0$ is a non-negative weight of each $T \in \mathcal{T}$.

For each $A \subseteq [N]$, let the truncated polynomial $f_A(x_1, \ldots, x_N)$ be defined as

$$f_A(x_1, \ldots, x_N) = \sum_{\substack{T \in \mathcal{T} \\ A \subseteq T}} w_T \prod_{i \in T \setminus A} x_i.$$

(In other words, for every monomial containing A, we substitute $x_i = 1$ for all $i \in A$. And other monomials are deleted entirely from the sum. Note that f_A does not depend on x_i for any $i \in A$.)

For any independent Bernoulli random variables $X_1, \ldots, X_N$, let E_A denote $\mathbb{E}[f_A(X_1, \ldots, X_N)]$ for every $A \subseteq [N]$. Moreover, let E denote $\max_{A \subseteq [N]} E_A$. Then, for any $\lambda > 1$, we have

$$\Pr[|f(X_1, \ldots, X_N) - E_\emptyset| > E(8\lambda)^t \sqrt{t!}] < 2e^2 e^{-\lambda} n^{t-1}.$$

To apply the Kim-Vu bound, we only need to show that E is also $n^{\beta+o(1)}$ as the term that is multiplied with E is negligible. The bound on E is stated below; we remark that both the α'-strictly balancedness and the constraint $\alpha'|T| - |v(T) \setminus S_{\max}| \geqslant -1$ are needed. The latter is needed even when bounding the expectation of F (i.e. $E_\emptyset$). On the other hand, roughly speaking, the former is used to ensure that the set T' that we take the partial derivative on is not too dense; otherwise, it could increase the expectation significantly. This intuition is formalized in the following proof.

▶ **Lemma 24.** *For $F((X_e)_{e \in \binom{V}{c}})$ as defined above and for any (possibility empty) subset $T' \subseteq \binom{V}{c}$, we have $E_{T'} \leqslant \left(\tau'(n)^{c\tau'(n)/\alpha'} \tau'(n)^2/\alpha' \right) k$. In particular, this also implies that $E \leqslant \left(\tau'(n)^{c\tau'(n)/\alpha'} \tau'(n)^2/\alpha' \right) k$.*

Proof. Observe that $F_{T'}((X_e)_{e \in \binom{V}{c}})$ is exactly

$$\sum_{\substack{T' \subseteq T \subseteq \binom{V}{c} \setminus \binom{S_{\max}}{c} \\ (S^*, T) \text{ satisfies } (*)}} \frac{n^{(1-\beta)\alpha'|T|}}{n^{(1-\beta)|v(T)\setminus S_{\max}|}} \prod_{e \in T \setminus T'} X_e.$$

Hence, if $|v(T') \cup S_{\max}| \geqslant \tau'(n)$, then this is simply zero. Furthermore, since the sum is only over T that is α'-strictly balanced, the expression is also zero unless $|v(T') \setminus S_{\max}| > \alpha'|T'|$. As a result, from now on, we may assume that $|v(T') \setminus S_{\max}| > \alpha'|T'|$. For brevity, let $y = |T'|$ and $z = |v(T') \setminus S_{\max}|$.

Now, observe that $\mathbb{E}[\prod_{e \in T \setminus T'} X_e]$ is simply $n^{-\alpha|T \setminus T'|}$. In other words, we have

$$
E_{T'} = \sum_{\substack{T' \subseteq T \subseteq \binom{V}{c} \setminus \binom{S_{\max}}{c} \\ (S^*,T) \text{ satisfies } (*)}} \frac{n^{(1-\beta)\alpha'|T|}}{n^{(1-\beta)|v(T) \setminus S_{\max}| + \alpha|T \setminus T'|}}
$$

$$
\leqslant \sum_{\substack{a,b \in \mathbb{N}_0 \\ b+z+|S_{\max}| \leqslant \tau'(n) \\ a \leqslant \tau'(n)/\alpha' \\ \alpha'(a+y)-(b+z) \geqslant -1}} \sum_{\substack{T' \subseteq T \subseteq \binom{V}{c} \setminus \binom{S_{\max}}{c} \\ |v(T) \setminus (v(T') \cup S_{\max})|=b \\ |T|=a+y}} \frac{n^{(1-\beta)\alpha'(a+y)}}{n^{(1-\beta)(b+z)+\alpha a}}
$$

$$
\leqslant \sum_{\substack{a,b \in \mathbb{N}_0 \\ b+z+|S_{\max}| \leqslant \tau'(n) \\ a \leqslant \tau'(n)/\alpha' \\ \alpha'(a+y)-(b+z) \geqslant -1}} n^b \cdot \binom{\tau'(n)^c}{a} \cdot \frac{n^{(1-\beta)\alpha'(a+y)}}{n^{(1-\beta)(b+z)+\alpha a}}
$$

$$
\leqslant \tau'(n)^{c\tau'(n)/\alpha'} \cdot \sum_{\substack{a,b \in \mathbb{N}_0 \\ b+z+|S_{\max}| \leqslant \tau'(n) \\ a \leqslant \tau'(n)/\alpha' \\ \alpha'(a+y)-(b+z) \geqslant -1}} n^b \cdot \frac{n^{(1-\beta)\alpha'(a+y)}}{n^{(1-\beta)(b+z)+\alpha'a}}
$$

$$
= \tau'(n)^{c\tau'(n)/\alpha'} \cdot \sum_{\substack{a,b \in \mathbb{N}_0 \\ b+z+|S_{\max}| \leqslant \tau'(n) \\ a \leqslant \tau'(n)/\alpha' \\ \alpha'(a+y)-(b+z) \geqslant -1}} n^{\beta((b+z)-\alpha'(a+y))} \cdot n^{\alpha'y-z}
$$

$$
< \tau'(n)^{c\tau'(n)/\alpha'} \cdot \sum_{\substack{a,b \in \mathbb{N}_0 \\ b+z+|S_{\max}| \leqslant \tau'(n) \\ a \leqslant \tau'(n)/\alpha' \\ \alpha'(a+y)-(b+z) \geqslant -1}} n^{\beta} \cdot n^0
$$

$$
\leqslant \left(\tau'(n)^{c\tau'(n)/\alpha'} \tau'(n)^2/\alpha' \right) k. \qquad \blacktriangleleft
$$

C.2   Putting Things Together

Note that for sufficiently large n, we have $\tau'(n) \geqslant 1/\alpha'$. Applying Theorem 23 with the above bound and with $\lambda = (\log n)^{10}$ and $t = \tau'(n)/\alpha'$, we arrive at the following bound:

$$
\Pr\left[F((X_e)_{e \in \binom{V}{2}}) > \left(2(8\log n)^{10\tau'(n)/\alpha'} (\tau'(n))^{3+(c+1)\tau'(n)/\alpha'} \right) \cdot k \right]
$$
$$
\leqslant O(e^{-\log^{10} n + (\tau'(n)/\alpha')\log n})
$$

Together with Claim 19, the above equation implies

$$
\Pr\left[\frac{\sum_{i \in V} y_{S \cup \{i\}}}{y_S} > \left(2 + 2(8\log n)^{10\tau'(n)/\alpha'} (\tau'(n))^{3+(c+1)\tau'(n)/\alpha'} \right) \cdot k \right]
$$
$$
\leqslant O(e^{-\log^{10} n + (\tau'(n)/\alpha')\log n}).
$$

Hence, by taking union bound over all S of size at most $r(n) \leqslant \tau'(n)$ we have

$$\Pr\left[\exists S, \frac{\sum_{i \in V} y_{S \cup \{i\}}}{y_S} > \left(2 + 2(8 \log n)^{10\tau'(n)/\alpha'}(\tau'(n))^{3+(c+1)\tau'(n)/\alpha'}\right) \cdot k\right]$$
$$\leqslant O(e^{-\log^{10} n + (2\tau'(n)/\alpha') \log n}).$$

Finally, observe that in our parameter selection $c\tau'(n) = o(\log n / \log \log n)$; hence, the term $\left(2 + 2(8 \log n)^{10\tau'(n)/\alpha'}(\tau'(n))^{3+(c+1)\tau'(n)/\alpha'}\right)$ is $n^{o(1)}$ as desired, and $O(e^{-\log^{10} n + (2\tau'(n)/\alpha') \log n}) = e^{-\log^{10} n + o(\log^2 n)} = o(1)$. In other words, have shown that the size constraint holds for all $S \subseteq V$ of size at most $r(n)$ with high probability, which concludes our proof.