

Approximating independent sets in sparse graphs*

Nikhil Bansal[†]

Abstract

We consider the maximum independent set problem on sparse graphs with maximum degree d . The best known result for the problem is an SDP based $O(d \log \log d / \log d)$ approximation due to Halperin. It is also known that no $o(d / \log^2 d)$ approximation exists assuming the Unique Games Conjecture. We show the following two results:

(i) The natural LP formulation for the problem strengthened by $O(\log^4 d)$ levels of the mixed-hierarchy has an integrality gap of $\tilde{O}(d / \log^2 d)$, where $\tilde{O}(\cdot)$ ignores some $\log \log d$ factors. However, our proof is non-constructive, in particular it uses an entropy based approach due to Shearer, and does not give a $\tilde{O}(d / \log^2 d)$ approximation algorithm with sub-exponential running time.

(ii) We give an $O(d / \log d)$ approximation based on $\text{polylog}(d)$ -levels of the mixed hierarchy that runs in $n^{O(1)} \exp(\log^{O(1)} d)$ time, improving upon Halperin's bound by a modest $\log \log d$ factor. Our algorithm is based on combining Halperin's approach together with an idea used by Ajtai, Erdős, Komlós and Szemerédi to show that K_r -free, degree- d graphs have independent sets of size $\Omega_r(n \log \log d / d)$.

1 Introduction

Given a graph $G = (V, E)$, an independent set is a subset of vertices S such that no two vertices in S are adjacent. The maximum independent set problem is one of the most well-studied problems in algorithms, and is known to be notoriously hard to approximate. For a graph on n vertices, the best known algorithm, due to Feige [10], achieves an approximation ratio of $\tilde{O}(n / \log^3 n)$, where $\tilde{O}(\cdot)$ suppresses some $\log \log n$ factors, whereas Hastad [16] showed (assuming $\text{NP} \not\subseteq \text{ZPP}$) that no $n^{1-\epsilon}$ approximation exists for any constant $\epsilon > 0$. The hardness has been improved more recently to $n / \exp((\log n)^{3/4+\epsilon})$ [18].

In this paper we focus on the case of bounded degree graphs with maximum degree d . Recall that the naive algorithm (that repeatedly picks an arbitrary vertex v and deletes its neighborhood) produces an independent set of size at least $n / (d + 1)$, and hence is a

$d + 1$ approximation. The first $o(d)$ approximation was obtained by Halldórsson and Radhakrishnan [14], who gave a $O(d / \log \log d)$ guarantee. Subsequently, Vishwanathan observed (see [12] for more details) that an $O(d \log \log d / \log d)$ approximation follows from the results of Alon and Kahale [4] and Karger, Motwani and Sudan [17]. This is the currently best known guarantee. Later, a simpler and direct $O(d \log \log d / \log d)$ was obtained independently by Halperin [15] and Halldórsson [13].

On the negative side, Austrin, Khot and Safra showed an $\Omega(d / \log^2 d)$ hardness of approximation, assuming the Unique Games Conjecture [6]. Assuming $\text{P} \neq \text{NP}$, a hardness of $d / \log^4 d$ was recently shown by Chan [8]. Closing the gap between $\Omega(d / \log^2 d)$ and $\tilde{O}(d / \log d)$ has been an intriguing open question. In particular, it is unclear what the right answer should be: there is a natural Ramsey theoretic barrier to improving the hardness beyond $d / \log^2 d$ (roughly speaking the gadgets used in the reductions are random-like and thus have large independent sets themselves). On the other hand, the natural SDP by itself does not seem to help beyond a $d \log \log / \log d$ approximation (more precisely, the negative correlation between vectors seems to become negligible when the SDP objective is less than $n \log \log d / \log d$).

We remark that while the algorithmic results mentioned above hold for the entire regime of $1 \leq d \leq n$, the hardness results of [6] only seem to hold when d is a constant or a very mildly increasing function of n . In fact for $d = n$, the $\Omega(d / \log^2 d)$ hardness [6] is inconsistent with the known $O(n / \log^3 n)$ approximation [10]. Hence throughout this paper, we will view d as being substantially smaller than n .

1.1 Our Results. Consider the standard LP formulation for the independent set problem strengthened by ℓ levels of the Sherali-Adams hierarchy, together with semidefinite constraints at the first level (see section 2 for details). We will refer to this as ℓ levels of the mixed hierarchy (this is also referred to as the SA+ hierarchy) and denote this relaxation by $M(\ell)$. Our first result is the following.

THEOREM 1.1. *The mixed hierarchy formulation using $O(\log^4 d)$ levels has an integrality gap of*

*Supported by NWO grant 639.022.211 and ERC consolidator grant 617951.

[†]Eindhoven University of Technology. Email: n.bansal@tue.nl

$O(d(\log \log d)^2 / \log^2 d)$.

Our proof of theorem 1.1 is non-algorithmic and does not give an $\tilde{O}(d/\log^2 d)$ approximation algorithm, even if we allow sub-exponential in n running time. In particular, it is based on a result of Alon [3] on the existence of large independent sets in locally sparse graphs, which in turn is based on an entropy-based approach of Shearer [22]. Theorem 1.1 suggests that $\tilde{\Omega}(d/\log^2 d)$ might be the right approximation threshold for the problem, at least when d is constant.

Our second result gives an algorithm that improves the previously known approximation by a modest $\Omega(\log \log d)$ factor, although at the expense of somewhat higher running time.

THEOREM 1.2. *There is an algorithm based on $\log^{O(1)} d$ levels of the mixed-hierarchy, that has approximation ratio $O(d/\log d)$ and runs in time $n^{O(1)} \exp(\log^{O(1)} d)$.*

While the $\log \log d$ improvement is perhaps not so interesting by itself, our techniques may be more interesting. The improvement in Theorem 1.2 is based on combining Halperin’s approach together with an idea used by Ajtai, Erdős, Komlós and Szemerédi [1] to show that K_r -free, degree- d graphs have independence number $\Omega_r(n \log \log d/d)$ (i.e. an $\Omega(\log \log d)$ factor more than the naive bound). Specifically, we use the properties of hierarchies to simulate the approach of Ajtai et al. [1] on top of Halperin’s algorithm and combine both their improvements.

Both Theorems 1.1 and 1.2 extend to the case when d is the average degree (instead of maximum) using standard techniques. However, for simplicity we only focus on maximum degree setting here.

1.2 Our Techniques. We first give a brief overview of previous techniques, and then describe our main ideas.

Let $\alpha(G)$ denote the size of a maximum independent set in a graph G . Let d and \bar{d} denote the maximum and average degree of G . The naive greedy algorithm implies $\alpha(G) \geq n/(d+1)$ for every G . In fact, it implies that $\alpha(G) = \Omega(n/\bar{d})$, since we can delete the vertices with degree more than $2\bar{d}$ (there are at most $n/2$ of them) and then apply the greedy algorithm.

As the greedy guarantee is tight in general (e.g. if the graph is a disjoint union of $n/(d+1)$ copies of the clique K_{d+1}), the trivial upper bound of $\alpha(G) \leq n$ cannot give an approximation better than $\bar{d}+1$ and hence stronger upper bounds are needed. A natural bound is the clique cover number $\bar{\chi}(G)$, defined as the minimum number of vertex disjoint cliques needed to cover V . As any independent set can contain at most one vertex from any clique, $\alpha(G) \leq \bar{\chi}(G)$.

Ramsey-theoretic approaches. Looking at $\bar{\chi}(G)$ naturally leads to Ramsey theoretic considerations. One must either show that the graph can be covered by large cliques, in which case $\alpha(G)$ is small and the trivial $n/(d+1)$ solution gives a good approximation. Otherwise, if $\bar{\chi}(G)$ is large, then this essentially means that there are not many large cliques and one must argue that a large independent set exists (and can be found efficiently).

For bounded degree graphs, a well-known result of this type is that $\alpha(G) = \Omega(n \log d/d)$ for triangle-free graphs [2, 21] (i.e. if there are no cliques of size 3). A particularly elegant proof (based on an idea due to Shearer [22] is in [5]). Moreover this bound is tight, and simple probabilistic constructions show that this bound cannot be improved even for graphs with large girth.

For the case of K_r -free graphs with $r \geq 4$, the situation is less clear. Ajtai et al. [1] showed that K_r -free graphs have $\alpha(G) = \Omega(n(\log(\log d/r))/d)$, which implies that $\alpha(G) = \Omega(n \log \log d/d)$ for $r \ll \log d$. This result was the basis of the $O(d/\log \log d)$ approximation due to [14]. Shearer [22] improved this result substantially and showed that $\alpha(G) \geq c_r(\log d/\log \log d)(n/d)$ for K_r -free graphs. However his argument is existential (and uses an elegant entropy based approach) and we are not aware of any algorithmic variants or applications of this result. Removing the $\log \log d$ factor above is a major open question, even for $r = 4$.

Semidefinite Programming. SDPs provide at least as strong upper bounds on α as $\bar{\chi}(G)$, as the well-known ϑ -function of Lovász satisfies $\alpha(G) \leq \vartheta(G) \leq \bar{\chi}(G)$. The $O(d \log \log d/\log d)$ approximations due to [15, 13] are both based on SDPs. For details on SDPs, and the Lovász ϑ -function, we refer the reader to [11]

Bounding the integrality gap. We describe the main ideas behind Theorem 1.1. For simplicity, suppose that we have a d -level mixed hierarchy formulation $M(d)$ (instead of an $O(\log^4 d)$ -levels).

The first observation is that if $M(d)$ has a solution of value say $\geq 4n \log \log d/\log d$, then Halperin’s algorithm already returns a sufficiently large independent set (details in Theorem 3.1 below). So we can assume that the objective value is $o(n \log \log d/\log d)$. For concreteness, let us suppose that each vertex contributes exactly $1/\log d$ to the objective.

Observe that a d -level Sherali-Adams solution specifies a valid local distribution over independent sets for every subset of vertices of size d . Fix a vertex v and consider the local-distribution μ_v over independent sets in the subgraph $G_{|N(v)}$ of G induced by the neighborhood $N(v)$ of v . As $x_u = 1/\log d$ for every vertex u , each such $u \in N(v)$ must lie in $1/\log d$ fraction (ac-

ording to μ_v) of the independent sets in $G_{|N(v)}$. Scaling by $\log d$, this implies that $N(v)$ has fractional chromatic number $\log d$, and hence by randomized rounding has chromatic number $\chi(N(v)) = O(\log^2 d)$. By a result of Alon [3] (Theorem 2.3 below), about independent sets in graphs where vertex neighborhoods have low chromatic number, this implies that $\alpha(G) = \Omega((n/d)(\log d/\log \log d))$. As the relaxation $M(d)$ had objective value $O(n \log \log d/\log d)$, this implies that the integrality gap is $\tilde{O}(d/\log^2 d)$.

To reduce the number of levels from d to $O(\log^4 d)$, we use some observations from [3] about the proof of Theorem 2.3. In particular, the result of Theorem 2.3 also holds under much weaker conditions (that we state in Theorem 2.4 below). As these observations only appear as a remark in [3], we will give a proof of Theorem 2.4 here for completeness.

Improved approximation. We now describe the idea for our algorithmic $O(d/\log d)$ approximation. For simplicity, let us first assume that we have a solution to a d -level formulation $M(d)$, and that for each vertex i , the relaxation sets $x_i = \eta$ where $\eta = \log \log d/(4 \log d)$ (Halperin’s rounding does not give anything better than the trivial n/d size independent set in this case).

We will show that we find an independent set of size $\Omega(n \log \log d/d)$, using an idea similar to the one used by Ajtai et al.[1] to show that $\alpha(G) = \Omega(n \log \log d/d)$ for K_r -free graphs, when $r \ll \log^{1-\epsilon} d$.

Roughly speaking, [1] show that for a K_r -free graph either (i) many vertices are involved in few triangles (in which case Theorem 2.2 below gives a large independent set), or else (ii) one can find a large enough subgraph with slightly smaller edge density. Iterating this process eventually produces a graph with edge density low enough that the naive greedy algorithm gives the desired approximation. We show that this argument can be implemented using the properties of the local-distribution on independent sets given by the Sherali-Adams solution.

In section 4, we first describe the algorithm based on a d -level solution to the mixed hierarchy. Later we sketch how this can also be achieved using $O(\text{polylog}(d))$ -levels of the hierarchy, using the local-global correlation approach of [7].

2 Preliminaries

Given the input graph $G = (V, E)$, we will denote the vertex set V by $[n] = \{1, \dots, n\}$.

Standard LP/SDP Formulations. In the standard LP relaxation for the independent set problem, there is variable x_i for each vertex i that is intended to be 1 if i

lies in the independent set and 0 otherwise. The LP is the following:

$$(2.1) \quad \begin{aligned} \max \quad & \sum_i x_i, \\ \text{s.t.} \quad & x_i + x_j \leq 1 \quad \forall (i, j) \in E \\ & x_i \in [0, 1] \quad \forall i \in [n]. \end{aligned}$$

In the standard SDP relaxation, there is a special unit vector v_0 (intended to indicate 1) and a vector v_i for each vertex i . The vector v_i is intended to be v_0 if i lies in the independent set and be $\mathbf{0}$ otherwise. This gives the following formulation:

$$(2.2) \quad \begin{aligned} \max \quad & \sum_i v_i \cdot v_0 \\ \text{s.t.} \quad & v_0 \cdot v_0 = 1 \\ & v_0 \cdot v_i = v_i \cdot v_i \quad \forall i \in [n] \\ & v_i \cdot v_j = 0 \quad \forall (i, j) \in E. \end{aligned}$$

Let Y denote the $(n+1) \times (n+1)$ Gram matrix with entries $y_{ij} = v_i \cdot v_j$, for $i, j \in \{0, \dots, n\}$. Then we have the equivalent formulation

$$(2.3) \quad \begin{aligned} \max \quad & \sum_i y_{0i} \\ \text{s.t.} \quad & y_{00} = 1 \\ & y_{0i} = y_{ii} \quad \forall i \in [n] \\ & y_{ij} = 0 \quad \forall (i, j) \in E \\ & Y \succeq 0. \end{aligned}$$

Lift-and-project Hierarchies. An excellent introduction to hierarchies and their algorithmic uses can be found in [9, 19] (in fact both these surveys discuss the independent set problem as an example). Here, we only describe here the most basic facts that we need.

The Sherali-Adams hierarchy defines a hierarchy of linear programs with increasingly tighter relaxations. At level t , there is a variable Y_S for each subset $S \subseteq [n]$ with $|S| \leq t+1$. Intuitively, one views Y_S as the probability that all the variables in S are set to 1. A solution to the Sherali-Adams hierarchy can be viewed as a local distribution over valid $\{0, 1\}$ -solutions involving variables from a set S of size at most $t+1$. More formally, for the independent set problem we have the following theorem from [9].

THEOREM 2.1. ([9], LEMMA 1) *Consider a family of distributions $\{\mathcal{D}(S)\}_{S \subseteq [n]: |S| \leq t+2}$, where each $\mathcal{D}(S)$ is defined over $\{0, 1\}^S$. If the distributions satisfy*

1. *For all $(i, j) \in E$ and $S \supseteq \{i, j\}$, it holds that $\Pr_{\mathcal{D}(S)}[(y_i = 1) \cap (y_j = 1)] = 0$, and*

2. For all $S' \subseteq S \subset [n]$ with $|S| \leq t + 1$, the distribution $\mathcal{D}(S')$, $\mathcal{D}(S)$ agree on S' .

Then $Y_S = \mathbb{P}_{\mathcal{D}(S)}[\bigwedge_{i \in S} (y_i = 1)]$ is a feasible solution for the level- t Sherali Adams relaxation.

Conversely, for any feasible solution $\{Y'_S\}$ for the level- $(t + 1)$ Sherali-Adams relaxation, there exists a family of distributions satisfying the above properties, as well as $Y_S = \mathbb{P}_{\mathcal{D}(S)}[\bigwedge_{i \in S} (y_i = 1)] = Y'_S$ for all $S' \subseteq S \subset [n]$ such that $|S| \leq t + 1$.

Here, Condition 1 implies that for a subset of vertices S with $|S| \leq t + 1$, the local-distribution $\mathcal{D}(S)$ has support on the valid independent sets in the graph induced on S , and Condition 2 guarantees that different local distributions induce a consistent distribution on the common elements.

For our purposes, we will also impose the PSD constraint on the variables y_{ij} at the first level (i.e. we add the constraints in (2.3) on y_{ij} variables). We will call this the level t mixed hierarchy formulation and denote it by $M(t)$.

A solution to the ℓ -level mixed relaxation above specifies values y_S for multi-sets S with $|S| \leq \ell + 1$. However to keep the notation consistent with the LP 2.1, we will use x_i to denote the marginals on the sets corresponding to single vertices.

Ramsey-theoretic lower bounds on the independence number. We will use the following result on independence number of graphs with few triangles.

THEOREM 2.2. ([2, 21]) *Let G be a graph with average degree \bar{d} , and at most $\epsilon \bar{d}^2 n$ triangles. Then, $\alpha(G) = \Omega(\log(1/\epsilon) \cdot (n/\bar{d}))$.*

Remark: This result is algorithmic. In fact it follows directly from the classic fact that $\alpha(G) = \Omega(n \log \bar{d}/\bar{d})$ for triangle free graphs [2, 21]. Indeed, sample each vertex with probability $p = 1/(2\sqrt{\epsilon d})$. This gives a graph with $n' = pn$ vertices in expectation (and tightly concentrated around this value). Moreover the average degree is $d' = p\bar{d} = 1/2\sqrt{\epsilon}$ and the number of triangles is $p^3 \epsilon \bar{d}^2 n = pn/4$ in expectation. Removing every vertex involved in a triangle gives a triangle free graph G' with $\Omega(n')$ vertices, and $\alpha(G') = \Omega(n' \log(d')/d') = \Omega(n \log(1/\epsilon)/\bar{d})$.

The following result will be crucial in our arguments.

THEOREM 2.3. (ALON [3], THEOREM 1.1) *Let $G = (V, E)$ be a graph on n vertices with maximum degree $d \geq 1$ in which for every vertex $v \in V$ the induced subgraph on the set of all neighbors of v is k -colorable. Then, the independence number of G is at*

least $\frac{c}{\log(k+1)} \frac{n}{d} \log d$, for some absolute positive constant c .

The result above also holds under the following weaker condition.

THEOREM 2.4. (ALON [3]) *Let G be a graph with maximum degree d , and let $k \geq 1$ be an integer. If for every vertex v and every subset $S \subset N(v)$ with $|S| \leq k \log^2 d$, it holds that the subgraph induced on S has an independent set of size at least $|S|/k$, then $\alpha(G) = \Omega(n \log d / (d \log k))$.*

The proof of theorem 2.4 can be found in the Appendix.

3 Integrality Gap

In this section we show that the integrality gap of the relaxation $M(\log^4 d)$ is $O(d(\log \log d)^2 / \log^2 d) = \tilde{O}(d/\log^2 d)$. Given a graph G on n vertices, let Y denote some optimum solution to the relaxation $M(\log^4 d)$. We will first preprocess the solution Y slightly to have some desirable properties.

3.1 Preprocessing Let s denote the objective value of the solution Y . We can assume that $s \geq 4n/\log^2 d$, otherwise the naive greedy algorithm trivially gives a $O(d/\log^2 d)$ approximation.

More importantly, we can also assume that s is not too large, otherwise Halperin's algorithm already gives a good approximation. In particular, we need the following result about the SDP formulation (2.2).

THEOREM 3.1. (HALPERIN [15], LEMMA 5.2) *Let $\eta \in [0, \frac{1}{2}]$ be a parameter and let Z be the collection of vectors v_i satisfying $\|v_i\|^2 \geq \eta$ in the SDP solution. Then there is an algorithm that returns an independent set of size $\Omega\left(\frac{d^{2\eta}}{d\sqrt{\ln d}}|Z|\right)$.*

As Halperin uses an SDP which looks different from ours (2.2) (though they are equivalent), we sketch the proof of Theorem 3.1 in the Appendix for completeness.

Note that in the formulation $M(\log^4 d)$, we have $\|v_i\|^2 = y_{ii} = y_{0i} = x_i$.

Let $\eta := 3 \log \log d / \log d$ and let Z denote the set of vertices with $x_i \geq \eta$. Then, theorem 3.1 returns an independent set of size $\Omega(|Z| \log^5 d/d)$. If $|Z| \geq n/\log^3 d$ then this gives an independent of size $\Omega(n \log^2 d/d)$ and hence an $O(d/\log^2 d)$ approximate solution. So we can assume that $|Z| \leq n/\log^3 d$. As the contribution of vertices in Z to the objective s can be at most $|Z|$, this contribution is $o(s)$ as $s \geq 4n/\log^2 d$.

This allows us to upper bound s as follows.

$$(3.4) \quad s \leq n \cdot \eta + |Z| \cdot 1 = (3 + o(1))n \frac{\log \log d}{\log d}.$$

Let A denote the set of vertices with $x_i \leq 1/\log^2 d$. Vertices in A contribute at most $|A|/(\log^2 d) \leq n/\log^2 d \leq s/4$ to s .

Let G' denote the graph obtained by deleting Z and A from G , and let Y' denote the solution Y restricted to G' . Note that Y' is a feasible solution for maximum independent set problem on G' and the objective value s' of Y' satisfies $s' \leq s$ and $s' \geq s - s/4 - o(s) \geq s/2$. Moreover, $1/\log^2 d \leq x_i \leq \eta$ for each vertex $i \in G'$. This implies that $s/2 \leq s' \leq \eta|V(G')|$ and hence that

$$(3.5) \quad |V(G')| \geq s/(2\eta)$$

We will show that

$$(3.6) \quad \alpha(G') \geq c \frac{|V(G')| \log d}{d \log(\log^3 d)}$$

for some absolute constant c . Together with (3.5), this will give the desired result that the integrality gap is at most

$$\frac{s}{\alpha(G)} \leq \frac{s}{\alpha(G')} \leq \frac{2\eta \log d}{cd \log(\log^3 d)} = O\left(\frac{d(\log \log d)^2}{\log^2 d}\right).$$

To show (3.6) we use Theorem 2.4 together with following observation.

LEMMA 3.1. *For the graph G' defined above, for every vertex $v \in V(G')$ and every subset $S \subset N(v)$ with $|S| \leq \log^4 d$, S contains an independent set of size at least $|S|/\log^2 d$.*

Proof. Fix a vertex v and $S \subset N(v)$ with $|S| \leq \log^4 d$. As $|S| \leq \log^4 d$, let us consider the local-distribution $\mathcal{D}(S)$ defined on the subsets $A \subseteq S$ given by the solution Y_A to $M(\log^4 d)$. By inclusion-exclusion, for each subset $A \subset S$, let $X_A = \sum_{T \subseteq S \setminus A} (-1)^{|T|} Y_{A \cup T}$ denote the ‘‘probability’’ that the vertices in A are exactly the ones picked in the independent set among the ones in S . By the properties of the Sherali-Adams hierarchy, $X_A \geq 0$ for each A , and moreover $X_A > 0$ only if A forms an independent set in the subgraph induced on S . Moreover, for a vertex $i \in S$, it holds that $x_i = \sum_{A \subseteq S: i \in A} x_A$ and that $1 = Y_\emptyset = \sum_{A \subseteq S} X_A$. As $x_i \geq 1/\log^2 d$ for each i , scaling the solution X_A by $\log^2 d$ defines a valid fractional set cover solution with objective value $\log^2 d$, and hence there must exist at least one independent set $A \subset S$ with $|A| \geq |S|/\log^2 d$.

Thus the graph G' satisfies the requirement in Theorem 2.4 with $k = \log^2 d$, and applying Theorem 2.4 to G' implies (3.6).

4 Algorithm

We now prove theorem 1.2. To keep the main idea clear, we first assume that we have a solution to a d -level relaxation. Later in section 4.3 we show how to reduce the number of levels to $\text{polylog}(d)$.

4.1 Preprocessing. Let Y denote a solution to the relaxation $M(d)$ on the instance I , and let s denote the objective value. We can assume that $s \geq 2n/\log d$, otherwise the naive algorithm trivially gives a $O(d/\log d)$ approximation.

Let $\gamma = 2 \log \log d / \log d$. By theorem 3.1 we can assume that at most $n/\log^2 d$ vertices have $x_i \geq \gamma$, otherwise we already have an $O(d/\log d)$ approximate solution. Let V' denote the set of vertices with $x_i \in [1/\log d, \gamma]$, and let G' denote the graph G restricted to vertices in V' .

CLAIM 4.1. $|V'| \geq \frac{s}{4\gamma}$.

Proof. The contribution of the vertices with $x_i \geq \gamma$ to s is at most the number of such vertices, which is at most $n/\log^2 d$. Similarly, the vertices with $x_i < 1/\log d$ contribute at most $n/\log d$ to s . Thus, vertices in V' contribute at least $s - (n/\log^2 d) - (n/\log d) \geq s - s/(2 \log d) - s/2 \geq s/4$ to the objective. As each $x_i \leq \gamma$ for $i \in V'$, it follows that $|V'| \geq s/(4\gamma)$.

4.2 Iterative Thinning Procedure. We will give an algorithm that finds an independent set H of G' of size at least $\Omega(|V'| \cdot \log \log d / d)$. By claim 4.1, $|H| = \Omega(s \log \log d / (\gamma d)) = \Omega(s \log d / d)$, which implies an $O(d/\log d)$ approximation.

To find such an independent set we will give an iterative procedure inspired by the approach in [1]. In particular, we will show the following result.

LEMMA 4.1. *Let G be a graph on n vertices with maximum degree d and average degree \bar{d} . Moreover, suppose there is a feasible d -level mixed hierarchy solution $M(d)$ satisfying $x_i \geq 1/\log d$, for each vertex i . Then, there is an efficient polynomial time algorithm that outputs one of the following:*

1. An independent set of size $\Omega((n \log \log d)/\bar{d})$.
2. A subgraph G' with at least $n' \geq n/(8 \log d)$ vertices and average degree \bar{d}' satisfying

$$(4.7) \quad n'/\bar{d}' \geq (1 + \beta)(n/\bar{d})$$

where $\beta = 1/(\log d)^{1/2}$.

Before proving Lemma 4.1, we show how this gives the claimed large independent set. We start

with the graph $G_0 = G$ and apply the algorithm in Lemma 4.1 repeatedly until the case (1) holds or until $\ell = (4 \log \log d)/\beta$ iterations are completed. Let G_1, G_2, \dots, G_k be the sequence of the subgraphs produced for some $k \leq \ell$. Let n_i and \bar{d}_i denote the number of vertices and average degree of G_i . Note that by (4.7), n_i/\bar{d}_i forms an increasing sequence.

If the process terminates after ℓ steps, then

$$\frac{n_\ell}{\bar{d}_\ell} \geq \left(1 + \frac{1}{\beta}\right)^\ell \frac{n}{\bar{d}} = \frac{n}{\bar{d}} \log^{\Omega(1)} d$$

and hence the greedy algorithm applied to G_ℓ gives the desired independent set.

On the other hand, if the process terminates at step k for $k < \ell$, then we get an independent set (in case 1) of size $\Omega((n_k/\bar{d}_k) \log \log d)$. This is at least $\Omega((n_0/\bar{d}_0) \log \log d)$, as n_i/\bar{d}_i is increasing. As $n_0 = n$ and $\bar{d}_0 = \bar{d}$, the result follows.

It remains to prove Lemma 4.1.

Proof. (Lemma 4.1). Start with the graph G and let $V = V(G)$. If there is a vertex v such that the graph induced on its neighborhood $N(v)$ has at least $\beta\bar{d}|N(v)|$ edges, remove $S = \{v\} \cup N(v)$ from G . Continue the process on the remaining graph with vertex set $V \setminus S$ until no such vertex exists or when the number of remaining vertices first falls below $n/2$. We consider two cases depending on when the process terminates.

1. If more than $n/2$ vertices are left, then in the remaining graph G , the neighborhood $N(v)$ of each vertex v contains at most $\beta\bar{d}|N(v)|$ edges, and thus the total number of triangles is at most $(1/3)\beta\bar{d}\sum_v N(v) = \beta n \bar{d}^2/3$. Moreover as G has at least $n/2$ vertices, its average degree is at most $2\bar{d}$. Thus by theorem 2.2, G contains an independent set of size $\Omega(\log(1/\beta) \cdot (n/\bar{d})) = \Omega(n \log \log d/\bar{d})$. This gives the desired set.
2. If $n/2$ or fewer vertices were left (there must be at least $n/2 - d$), consider the sets S_1, \dots, S_T that were removed from G during the process, where $S_i = v_i \cup N(v_i)$. As the number of edges in each subgraph S_i is at least

$$\beta\bar{d}|N(v_i)| = \beta\bar{d}|S_i - 1| \geq \beta\bar{d}|S_i|/2.$$

The number of edges in G that have both endpoints in the set S_i for some i , is at least

$$\sum_{i=1}^T \beta\bar{d}|S_i|/2 \geq \beta\bar{d}/2(n/2) \geq \beta|E|/2,$$

where $|E|$ denotes the number of edges in G .

Now consider the distribution over the independent sets in S_i determined by the local-distribution $\mathcal{D}(S_i)$ given by the Sherali-Adams solution.

For each $i = 1, \dots, T$ we do the following: Sample an independent set $T_i \subset S_i$ according to the probabilities determined by $\mathcal{D}(S_i)$. Let $p = 1/\log d$. For each vertex $v \in T_i$, sample it independently with probability p/x_v , and reject it otherwise. This step is well-defined as $x_v \geq p$ for each vertex $v \in G$. Let $U_i \subset T_i$ denote the set of vertices that are sampled. Clearly, each U_i is an independent sets, and each vertex $v \in S_i$ lies in U_i with probability exactly p .

For vertices $v \in (V \setminus (S_1 \cup \dots \cup S_T))$, we sample each of them independently with probability p .

Let G' be the graph induced on the sampled vertices. The expected number of vertices in G' is np . Moreover, as $|S_i| \leq d+1$ for each i , the number of vertices is tightly concentrated around np (with standard deviation $O(\sqrt{npd})$). The crucial point is that an edge $e \in G$ with both its end-points in some S_i has 0 probability of lying in G' and exactly p^2 probability otherwise. Thus the expected number of edges in G' is $|E|(1 - \beta/2)p^2 = n\bar{d}(1 - \beta/2)p^2/2$ (and tightly concentrated). This gives us that $d' = dp(1 - \beta/2)$ plus lower order terms with high probability. Thus n'/\bar{d}' is at least $(1 + \beta/4)n/d$ with high probability, as desired.

We note that we only require the Sherali-Adams local distribution on subsets of vertices that lie in some neighborhood $N(v)$. Thus, the running time to compute the solution to the relaxation is $n^{O(1)} \cdot \exp(d)$. We now show how the number of levels required can be reduced.

4.3 Reducing the number of levels. We sketch the idea to reduce the number of levels to $\text{polylog}(d)$, and defer the relatively standard details to the full version of the paper.

Observe that in the proof of lemma 4.1, we do not necessarily require a distribution over the independent sets T_i of S_i . In fact, any distribution over subsets T_i with edge density a constant factor less than that of S_i suffices to argue that the graph G' has $n'/d' \geq (1 + c\beta)n/d$ for some constant c .

Now the crucial point is that the graph induced on S_i is quite dense. It has at most d vertices and at least $\beta\bar{d}|S_i|$ edges. Moreover, we can assume that $\bar{d} \geq d/\log \log d$, otherwise the greedy algorithm (applied after removing vertices of degree more than $2\bar{d}$) already returns an independent set of size $\Omega(n \log \log d/d)$. Thus the edge density of S_i is at least $\beta\bar{d}|S_i|/|S_i|^2 \geq$

$\beta d/(|S_i| \log \log d) \geq 1/\log d$, and one can use additive approximations for dense constraint satisfaction problems.

In particular, we can use the global correlation rounding technique introduced by Barak, Raghavendra and Steurer [7], to find a large sparse subset T_i of S_i with relatively few levels. More precisely, we use the following result (while originally stated for the Lasserre hierarchy, but it also works for the Sherali Adams hierarchy).

THEOREM 4.1. ([20], THM 17, SLIGHTLY RESTATED)
Let $K = \{x \in \mathbb{R}^n | Ax \geq b\}$ be any polytope and Y be some solution to a t -level Sherali-Adams lifting of K with $t \geq 1/\epsilon^3 + 2$. For any $R \subseteq [n]$, one can condition on at most $1/\epsilon^3$ randomly chosen variables in R to obtain a $t - 1/\epsilon^3$ level Sherali Adams solution y' such that $\Pr_{i,j \in R}[|y_i y_j - y_{ij}| \geq \epsilon] \leq \epsilon$.

Thus for each S_i separately, we can do the following. We set $\epsilon \leq \beta/(10 \log^2 d)$, and pick $1/\epsilon^3$ variables at random from S_i , and condition upon them to obtain an essentially uncorrelated solution as implied by Theorem 4.1. If we sample each vertex independently from this solution, we obtain a subset T_i of size $\Omega(|S_i|/\log d)$, as each vertex lies in T_i with probability x_v in expectation (where the expectation is over the conditionings in Theorem 4.1) and the number of edges in T_i is at most $\epsilon |S_i|^2$ which is $\leq \beta |T_i|^2/10$ as desired.

Acknowledgment

I thank Anand Louis, Konstantin Makarychev and Prasad Raghavendra for some interesting discussions. I would also like to thank an anonymous referee for pointing an error in a previous version of the paper.

References

- [1] Miklós Ajtai, Paul Erdős, János Komlós, and Endre Szemerédi. On turán's theorem for sparse graphs. *Combinatorica*, 1(4):313–317, 1981.
- [2] Miklós Ajtai, János Komlós, and Endre Szemerédi. A note on ramsey numbers. *J. Comb. Theory, Ser. A*, 29(3):354–360, 1980.
- [3] Noga Alon. Independence numbers of locally sparse graphs and a ramsey type problem. *Random Struct. Algorithms*, 9(3):271–278, 1996.
- [4] Noga Alon and Nabil Kahale. Approximating the independence number via the theta-function. *Math. Program.*, 80:253–264, 1998.
- [5] Noga Alon and Joel Spencer. *The Probabilistic Method*. John Wiley, 2008.
- [6] Per Austrin, Subhash Khot, and Muli Safra. Inapproximability of vertex cover and independent set in bounded degree graphs. *Theory of Computing*, 7(1):27–43, 2011.
- [7] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *FOCS*, pages 472–481, 2011.
- [8] Siu On Chan. Approximation resistance from pairwise independent subgroups. In *STOC*, pages 447–456, 2013.
- [9] Eden Chlamtac and Madhur Tulsiani. Convex relaxations and integrality gaps. In Miguel F. Anjos and Jean B. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*. Springer, 2012.
- [10] Uriel Feige. Approximating maximum clique by removing subgraphs. *SIAM J. Discrete Math.*, 18(2):219–225, 2004.
- [11] Bernd Gärtner and Jiri Matousek. *Approximation Algorithms and Semidefinite Programming*. Springer, 2012.
- [12] Magnús M. Halldórsson. Approximations of independent sets in graphs. In *APPROX*, pages 1–13, 1998.
- [13] Magnús M. Halldórsson. Approximations of weighted independent set and hereditary subset problems. In *COCOON*, pages 261–270, 1999.
- [14] Magnús M. Halldórsson and Jaikumar Radhakrishnan. Improved approximations of independent sets in bounded-degree graphs via subgraph removal. *Nord. J. Comput.*, 1(4):475–492, 1994.
- [15] Eran Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *SIAM J. Comput.*, 31(5):1608–1623, 2002.
- [16] Johan Hastad. Clique is hard to approximate within $n^{1-\epsilon}$. In *FOCS*, pages 627–636, 1996.
- [17] David R. Karger, Rajeev Motwani, and Madhu Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, 1998.
- [18] Subhash Khot and Ashok Kumar Ponnuswami. Better inapproximability results for maxclique, chromatic number and min-3lin-deletion. In *ICALP (1)*, pages 226–237, 2006.
- [19] Monique Laurent. A comparison of the sherali-adams, lovász-schrijver, and lasserre relaxations for 0-1 programming. *Math. Oper. Res.*, 28(3):470–496, 2003.
- [20] Thomas Rothvoss. The lasserre hierarchy in approximation algorithms, 2013. Lecture notes for the MAPSP 2013 Tutorial.
- [21] James B. Shearer. A note on the independence number of triangle-free graphs. *Discrete Mathematics*, 46(1):83–87, 1983.
- [22] James B. Shearer. On the independence number of sparse graphs. *Random Struct. Algorithms*, 7(3):269–272, 1995.

5 Appendix

5.1 Proof of Theorem 2.4. Here we prove the following result.

Theorem 2.4 [Alon [3]] Let G be a graph with maximum degree d , and let $k \geq 1$ be an integer. If

for every vertex v and every subset $S \subset N(v)$ with $|S| \leq k \log^2 d$, it holds that the subgraph induced on S has an independent set of size at least $|S|/k$, then $\alpha(G) = \Omega(n \log d / (d \log k))$.

We first need the following basic fact (see Lemma 2.2 in [3] for a proof).

LEMMA 5.1. *Let F be a family of $2^{\epsilon x}$ distinct subsets of an x -element set X . Then the average size of a member of F is at least $\epsilon x / (10 \log(1 + 1/\epsilon))$.*

Proof. Let W be a random independent set of vertices in G , chosen uniformly among all independent sets in G . For each vertex v , let X_v be a random variable defined as $X_v = d|v \cap W| + |N(v) \cap W|$.

Since $|W|$ can be written as $\sum_v |v \cap W|$ and as it satisfies $|W| \geq (1/d) \sum_v |N(v) \cap W|$ (a vertex in W can be in at most d sets $N(v)$), we have that

$$|W| \geq \frac{1}{2d} \sum_v X_v.$$

Thus to show that $\alpha(G)$ is large, it suffices to show that

$$(5.8) \quad E[X_v] \geq \log d / (80 \log k + 1).$$

To show (5.8), we prove that in fact it holds for every conditioning of the choice of the independent set in $V - (N(v) \cup \{v\})$. In particular, let H denote the subgraph of G induced on $V - (N(v) \cup \{v\})$. For each possible independent set S in H , we will show that

$$E[X_v | W \cap V(H) = S] \geq \log d / (80 \log k + 1).$$

Fix a choice of S . Let X denote the non-neighbors of S in $N(v)$, and let $x = |X|$. Let ϵ be such that $2^{\epsilon x}$ denotes the number of independent sets in the induced subgraph $G|_X$ of G on X . Now, conditioning on the intersection $W \cap V(H) = S$, there are precisely $2^{\epsilon x} + 1$ possibilities for W : one in which $W = S \cup \{v\}$ and $2^{\epsilon x}$ in which $v \notin W$ and W is the union of S with an independent set in $G|_X$. By lemma 5.1, the average size of an independent set in X is at least $\frac{\epsilon x}{10 \log(1/\epsilon + 1)}$ and thus we have that

$$(5.9) \quad \begin{aligned} & E[X_v | W \cap V(H) = S] \\ & \geq d \frac{1}{2^{\epsilon x} + 1} + \frac{\epsilon x}{10 \log(1/\epsilon + 1)} \frac{2^{\epsilon x}}{2^{\epsilon x} + 1}. \end{aligned}$$

Now, if $2^{\epsilon x} + 1 \leq \sqrt{d}$, then the first term is at least \sqrt{d} and we are done. Otherwise, it must be that

$\epsilon x \geq (1/2) \log d$ and hence the right hand side in (5.9) is at least

$$(5.10) \quad \frac{\log d}{40 \log(1/\epsilon + 1)}.$$

We now consider two cases depending on the value of x . If $x \leq k \log^2 d$, then by our assumption, X contains an independent set of size at least $|X|/k$ and hence $2^{\epsilon x} \geq 2^{x/k}$ and hence $\epsilon \geq 1/k$. This gives that (5.10) is at least $\frac{\log d}{40 \log(k+1)}$.

On the other hand if $x \geq k \log^2 d$, then by our assumption, X contains at least $2^{\log^2 d}$ independent sets, and hence $\epsilon x \geq \log^2 d$. As $x \leq d$, it follows that $\epsilon \geq \log^2 d / d \geq 1/d$ and hence $\log(1/\epsilon + 1) \leq 2 \log d$. Thus the right hand side of (5.9) is at least

$$\frac{\epsilon x}{20 \log(1/\epsilon + 1)} \geq \frac{\log^2 d}{40 \log d}.$$

5.2 Proof of Theorem 3.1. We prove the following result.

Theorem 3.1 [Halperin [15]] Let $\eta \in [0, \frac{1}{2}]$ be a parameter and let Z be the collection of vectors v_i satisfying $\|v_i\|^2 \geq \eta$ in the SDP solution. Then there is an algorithm that returns an independent set of size $\Omega\left(\frac{d^{2\eta}}{d\sqrt{\ln d}} |Z|\right)$.

Proof. Let $a_i = v_i \cdot v_0 = \|v_i\|^2$, and let $w_i = v_i - \langle v_i, v_0 \rangle v_0$ denote the projection of v_i to v_0^\perp , the hyperplane orthogonal to v_0 . As $\|w_i\|^2 + \langle v_i, v_0 \rangle^2 = \|v_i\|^2$, we obtain $\|w_i\|^2 = a_i - a_i^2$. Let $u_i = w_i / \|w_i\|$.

Now for any pair of vertices i, j , we have that $w_i \cdot w_j = v_i \cdot v_j - \langle v_i, v_0 \rangle \langle v_j, v_0 \rangle$. As $v_i \cdot v_j = 0$ if $(i, j) \in E$, we have that

$$w_i \cdot w_j = v_i \cdot v_j - \langle v_i, v_0 \rangle \langle v_j, v_0 \rangle = -a_i a_j$$

and hence

$$u_i \cdot u_j = -\frac{\sqrt{a_i a_j}}{\sqrt{(1-a_i)(1-a_j)}} \leq -\frac{\eta}{(1-\eta)}.$$

The last step follows as $a_i, a_j \geq \eta$ and as $x/1 - x$ is increasing for $x \in [0, 1]$.

Thus the unit vectors u_i can be viewed as a feasible solution to a vector k -coloring (in the sense of [17]) where k is such that $1/(k-1) = \eta/(1-\eta)$. This gives $k = 1/\eta$, and now we can use the result of [17] (Lemma 7.1) that such graphs have independent sets of size $\Omega(|Z|/d^{1-2/k} \sqrt{\ln d}) = \Omega(|Z|d^{2\eta}/(d\sqrt{\ln d}))$.