

# On the Lovász Theta function for Independent Sets in Sparse Graphs

Nikhil Bansal\*

Anupam Gupta†

Guru Guruganesh†

## Abstract

We consider the maximum independent set problem on sparse graphs with maximum degree  $d$ . We show that the integrality gap of the Lovász  $\vartheta$ -function based SDP is

$$\tilde{O}(d/\log^{3/2} d).$$

This improves on the previous best result of  $\tilde{O}(d/\log d)$ , and almost matches the integrality gap of  $\tilde{O}(d/\log^2 d)$  recently shown for stronger SDPs, namely those obtained using poly  $\log(d)$  levels of the  $SA^+$  semidefinite hierarchy. The improvement comes from an improved Ramsey-theoretic bound on the independence number of  $K_r$ -free graphs for large values of  $r$ .

We also show how to obtain an algorithmic version of the above-mentioned  $SA^+$ -based integrality gap result, via a coloring algorithm of Johansson. The resulting approximation guarantee of  $\tilde{O}(d/\log^2 d)$  matches the best unique-games-based hardness result up to lower-order poly( $\log \log d$ ) factors.

---

\*Eindhoven University of Technology. Email: n.bansal@tue.nl. Supported by NWO grant 639.022.211 and an ERC consolidator grant 617951.

†Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Research partly supported by NSF awards CCF-1016799 and CCF-1319811.

# 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 graph theory, and its study has led to various remarkable developments such as the seminal result of Lovász [Lov79] in which he introduced the  $\vartheta$ -function based on semidefinite programming, as well as several surprising results in Ramsey theory and extremal combinatorics.

In general graphs, the problem is notoriously hard to approximate. Given a graph on  $n$  vertices, the best known algorithm is due to Feige [Fei04], and achieves an approximation ratio of  $\tilde{O}(n/\log^3 n)$ ; here  $\tilde{O}(\cdot)$  suppresses some  $\log \log n$  factors. On the hardness side, a result of Håstad [Hås96] shows that no  $n^{1-\varepsilon}$  approximation exists for any constant  $\varepsilon > 0$ , assuming  $\text{NP} \not\subseteq \text{ZPP}$ . The hardness has been improved more recently to  $n/\exp((\log n)^{3/4+\varepsilon})$  by Khot and Ponnuswami [KP06].

In this paper, we focus on the case of bounded-degree graphs, with maximum degree  $d$ . Recall that the naïve 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 [HR94], who gave a  $O(d/\log \log d)$  guarantee, based on a Ramsey theoretic result of Ajtai et al. [AEKS81]. Subsequently, an  $O(d \frac{\log \log d}{\log d})$ -approximation was obtained independently by several researchers [AK98, Hal02, Hal00] using the ideas of Karger, Motwani and Sudan [KMS98] to round the natural SDP for the problem, which was itself based on the Lovász  $\vartheta$ -function.

On the negative side, Austrin, Khot and Safra [AKS11] showed an  $\Omega(d/\log^2 d)$  hardness of approximation, assuming the Unique Games Conjecture. Assuming  $\text{P} \neq \text{NP}$ , a hardness of  $d/\log^4 d$  was recently shown by Chan [Cha13]. We remark that these hardness results 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 of [AKS11] is inconsistent with the known  $O(n/\log^3 n)$  approximation [Fei04]. Hence throughout this paper, it will be convenient to view  $d$  as being a sufficiently large but fixed constant.

Roughly speaking, the gap between the  $\Omega(d/\log^2 d)$ -hardness and the  $\tilde{O}(d/\log d)$ -approximation arises for the following fundamental reason. Approaches based on the SDP work extremely well if the  $\vartheta$ -function has value more than  $\tilde{O}(n/\log d)$ , but not below this threshold. In order to show an  $\Omega(d/\log d)$ -hardness result, at the very least, one needs an instance with SDP value around  $n/\log d$ , but optimum integral value about  $n/d$ . While graphs with the latter property clearly exist (e.g., a graph consisting of  $n/(d+1)$  disjoint cliques  $K_{d+1}$ ), the SDP value for such graphs seems to be low. In particular, having a large SDP value imposes various constraints on the graph (for example, they cannot contain many large cliques) which might allow the optimum to be non-trivially larger than  $n/d$ , for example due to Ramsey-theoretic reasons.

Recently, Bansal [Ban15] leveraged some of these ideas to improve the approximation guarantee by a modest  $O(\log \log d)$  factor to  $d/\log d$  using  $\text{polylog}(d)$  levels of the  $SA^+$  hierarchy. His improvement was based on combining properties of the  $SA^+$  hierarchies together with the ideas of [AEKS81]. He also showed that the  $O(\log^4 d)$ -level  $SA^+$  relaxation has an *integrality gap* of  $\tilde{O}(d/\log^2 d)$ , where  $\tilde{O}(\cdot)$  suppresses some  $\log \log d$  factors. The main observation was that as the  $SA^+$  relaxation specifies a local distribution on independent sets, and if the relaxation has high objective value then it must be that any  $\text{polylog}(d)$  size subset of vertices  $X$  must contain a large independent subset. One can then use a result of Alon [Alo96], in turn based on an elegant entropy-based approach of Shearer [She95], to show that such graphs have non-trivially large independent sets. However,

this argument is non-algorithmic; it shows that the lifted SDP has a small integrality gap, but does not give a corresponding approximation algorithm with running time sub-exponential in  $n$ . This leads to the question whether this approach can be converted into an approximation algorithm that outputs a set of size  $\tilde{\Omega}(\log^2 d/d)$  times the optimal independent set, or if there is a gap between the approximability and estimability of this problem (as recently shown for an NP problem by Feige and Jozeph [FJ14]).

## 1.1 Our Results.

Our results resolve some of these questions. For our first result, we consider the standard SDP relaxation for independent set (without applying any lift-and-project steps) and show that it is surprisingly more powerful than the guarantee given by Alon and Kahale [AK98] and Halperin [Hal02].

**Theorem 1.1** *On graphs with maximum degree  $d$ , the standard  $\vartheta$ -function-based SDP formulation for the independent set problem has an integrality gap of  $\tilde{O}(d/\log^{3/2} d)$ .<sup>1</sup>*

The proof of Theorem 1.1 is non-constructive; while it shows that the SDP value is within the claimed factor of the optimal IS size, it does not give an efficient algorithm to find such an approximate solution. Finding such an algorithm remains an open question.

The main technical ingredient behind Theorem 1.1 is the following new Ramsey-type result about the existence of large independent sets in  $K_r$ -free graphs. This builds on a long line of previous results in Ramsey theory (some of which we discuss in Section 2), and is of independent interest. (Recall that  $\alpha(G)$  is the maximum independent set size in  $G$ .)

**Theorem 1.2** *For any  $r > 0$ , if  $G$  is a  $K_r$ -free graph with maximum degree  $d$  then*

$$\alpha(G) = \Omega\left(\frac{n}{d} \cdot \max\left(\frac{\log d}{r \log \log d}, \left(\frac{\log d}{\log r}\right)^{1/2}\right)\right). \quad (1)$$

Previously, the best known bound for  $K_r$ -free graphs was  $\Omega(\frac{n}{d} \frac{\log d}{r \log \log d})$  given by Shearer [She95]. Observe the dependence on  $r$ : when  $r \geq \frac{\log d}{\log \log d}$ , i.e., when we are only guaranteed to exclude very large cliques, this result does not give anything better than the trivial  $n/d$  bound. It is in this range of  $r \geq \log d$  that the second term in the maximization in (1) starts to perform better and give a non-trivial improvement. In particular, if  $G$  does not contain cliques of size  $r = O(\log^{3/2} d)$  (which will be the interesting case for Theorem 1.1), Theorem 1.2 gives a bound of  $\tilde{\Omega}(\frac{n}{d} (\log d)^{1/2})$ . Even for substantially larger values such as  $r = \exp(\log^{1-2\varepsilon} d)$ , this gives a non-trivial bound of  $\tilde{O}(\frac{n}{d} \log^\varepsilon d)$ .

Improving on Shearer's bound has been a long-standing open problem in the area, and it is conceivable that the right answer for  $K_r$ -free graphs of maximum degree  $d$  is  $\alpha(G) \geq \frac{n \log d}{d \log r}$ . This would be best possible, since in Section 3.1 we give a simple construction showing an upper bound of  $\alpha(G) = O(\frac{n \log d}{d \log r})$  for  $r \geq \log d$ , which to the best of our knowledge is the smallest upper bound currently known. The gap between our lower bound and this upper bound remains an intriguing

---

<sup>1</sup>Here and subsequently,  $\tilde{O}(\cdot)$  suppresses poly( $\log \log d$ ) factors.

one to close; in fact it follows from our proof of Theorem 1.1 that such a lower bound would imply an  $\tilde{O}(d/\log^2 d)$  integrality gap for the standard SDP. Alon [Alo96] shows that this bound is achievable under the stronger condition that the neighborhood of each vertex is  $(r - 1)$ -colorable.

We then turn to the approximation question. Our third result shows how to make Bansal’s result algorithmic, thereby resolving the approximability of the problem (up to lower order  $\text{poly}(\log \log d)$  factors), at least for moderate values of  $d$ .

**Theorem 1.3** *There is an  $\tilde{O}(d/\log^2 d)$ -approximation algorithm with running time<sup>2</sup>.  $\text{poly}(n) \cdot 2^{O(d)}$ , based on rounding a  $d$ -level  $SA^+$  semidefinite relaxation.*

The improvement is simple, and is based on bringing the right tool to bear on the problem. As in [Ban15], the starting point is the observation that if the  $d$ -level  $SA^+$  relaxation has objective value at least  $n/s$  (and for Theorem 1.3 the value  $s = \log^2 d$  suffices), then the neighborhood of every vertex in the graph is  $k$ -colorable for  $k = s \cdot \text{polylog}(d)$  — they are “locally colorable”. By Alon’s result mentioned above, such graphs have  $\alpha(G) = \Omega(\frac{n}{d} \frac{\log d}{\log k})$ . However, instead of using [Alo96] which relies on Shearer’s entropy based approach, and is not known to be constructive, we use an ingenious and remarkable (and stronger) result of Johansson [Joh96b], who shows that the list-chromatic number of such locally-colorable graphs is  $\chi_\ell(G) = O(d \frac{\log k}{\log d})$ . His result is based on a very clever application of the Rödl “nibble” method, together with Lovász Local Lemma to tightly control the various parameters of the process at every vertex in the graph. Applying Johansson’s result to our problem gives us the desired algorithm.

Unfortunately, Johansson’s preprint (back from 1996) was never published, and cannot be found on the Internet.<sup>3</sup> For completeness (and to facilitate verification), we give the proof in its entirety in the Appendix. We essentially follow his presentation, but streamline some arguments based on recent developments such as concentration bounds for low-degree polynomials of random variables, and the algorithmic version of Local Lemma. His manuscript contains many other results that build upon and make substantial progress on a long line of work (we give more details in Section 2). We hope that this will make Johansson’s ideas and results accessible to a wider audience. (Johansson’s previous preprint [Joh96a] showing the analogous list-coloring result for triangle-free graphs is also unavailable publicly, but is presented in the graph coloring book by Molloy and Reed [MR02], and has received considerable attention since, both in the math [AKS99, Vu02, FM13] and computer science communities [GP00, CPS14].)

The proof of Theorem 1.3 also implies the following new results about the LP-based Sherali-Adams ( $SA$ ) hierarchies, without any SDP constraints.

**Corollary 1.4** *The LP relaxation with clique constraints on sets of size up to  $\log d$  (and hence the relaxation  $SA_{(\log d)}$ ) has an integrality gap of  $\tilde{O}(d/\log d)$ . Moreover, the relaxation  $SA_{(d)}$  can be used to find an independent set achieving an  $\tilde{O}(d/\log d)$  approximation in time  $\text{poly}(n) \cdot 2^{O(d)}$ .*

Since LP-based relaxations have traditionally been found to be very weak for the independent set problem, it may be somewhat surprising that a few rounds of the  $SA$ -hierarchy improves the integrality gap by a non-trivial amount.

<sup>2</sup> While a  $d$ -level  $SA^+$  relaxation has size  $n^{O(d)}$  in general, our relaxation only uses variables corresponding to subsets of vertices that lie in the neighborhood of some vertex  $v$ , and thus has  $n \cdot 2^{O(d)}$  variables

<sup>3</sup>We thank Alan Frieze for sharing a copy with us.

All our results extend to the case when  $d$  is the average degree of the graph; by first deleting the (at most  $n/2$ ) vertices with degree more than  $2\bar{d}$  and then applying the results.

## 2 Preliminaries

Given the input graph  $G = (V, E)$ , we will denote the vertex set  $V$  by  $[n] = \{1, \dots, n\}$ . Let  $\alpha(G)$  denote the size of a maximum independent set in  $G$ , and  $d$  denote the maximum degree in  $G$ . The naive greedy algorithm implies  $\alpha(G) \geq n/(d+1)$  for every  $G$ . 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  $d+1$  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,  $\alpha(G) \leq \bar{\chi}(G)$ .

**Standard LP/ SDP Relaxations.** 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:

$$\max \sum_i x_i, \quad \text{s.t.} \quad x_i + x_j \leq 1 \quad \forall (i, j) \in E, \quad \text{and} \quad x_i \in [0, 1] \quad \forall i \in [n]. \quad (2)$$

Observe that this linear program is very weak, and cannot give an approximation better than  $(d+1)/2$ : even if the graph consists of  $n/(d+1)$  copies of  $K_{d+1}$ , the solution  $x_i = 1/2$  for each  $i$  is a feasible one.

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 relaxation:

$$\max \sum_i v_i \cdot v_0, \quad \text{s.t.} \quad v_0 \cdot v_0 = 1, \quad v_0 \cdot v_i = v_i \cdot v_i \quad \forall i \in [n], \quad \text{and} \quad v_i \cdot v_j = 0 \quad \forall (i, j) \in E. \quad (3)$$

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 relaxation

$$\max \sum_i y_{0i}, \quad \text{s.t.} \quad y_{00} = 1, \quad y_{0i} = y_{ii} \quad \forall i \in [n], \quad y_{ij} = 0 \quad \forall (i, j) \in E \quad \text{and} \quad Y \succeq 0. \quad (4)$$

The above SDP which is equivalent to the well-known  $\vartheta$ -function of Lovász [Lau](Lemma 3.4.4), satisfies  $\alpha(G) \leq \vartheta(G) \leq \bar{\chi}(G)$ . The  $O(d \frac{\log \log d}{\log d})$  approximations due to [AK98, Hal02, Hal00] are all based on SDPs.

We will use the following important result due to Halperin [Hal02] about the performance of the SDP. The form below differs slightly from the one in [Hal02] as he works with a  $\{-1, 1\}$  formulation. A proof for the form below can be found in [Ban15, Theorem 3.1].

**Theorem 2.1 (Halperin [Hal02], 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  $\Omega\left(\frac{d^{2\eta}}{d\sqrt{\ln d}}|Z|\right)$ .*

Note that if  $\eta = c \log \log d / \log d$ , then for  $c \leq 1/4$  Theorem 2.1 does not return any non-trivial independent set. On the other hand, for  $c \geq 1/4$  the size of the independent set returned rises exponentially fast with  $c$ .

For more details on SDPs, and the Lovász  $\vartheta$ -function, we refer the reader to [GLS88, GM12].

**Lower Bounds on the Independence Number.** As SDPs can handle cliques, looking at  $\vartheta(G)$  naturally leads to Ramsey theoretic considerations. In particular, if  $\vartheta(G)$  is small then the trivial  $n/(d+1)$  solution already gives a good approximation. Otherwise, if  $\vartheta(G)$  is large, then this essentially means that there are no 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  $\alpha(G) = \Omega(n \frac{\log d}{d})$  for triangle-free graphs [AKS80, She83] (i.e. if there are no cliques of size 3). A particularly elegant proof (based on an idea due to Shearer [She95]) is in [AS92]. 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. [AEKS81] 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 [HR94]. Shearer [She95] improved this result substantially and showed that  $\alpha(G) \geq \Omega(\frac{1}{r} n \frac{\log d}{d \log \log d})$  for  $K_r$ -free graphs. His result is based on an elegant entropy based approach that has subsequently found many applications. However, it is not known how to make this method algorithmic. Removing the  $\log \log d$  factor above is a major open question, even for  $r = 4$ . Also, note that his bound is trivial when  $r \geq \frac{\log d}{\log \log d}$ .

Interestingly, this result also implies another (non-algorithmic) proof that the SDP has integrality gap  $d \frac{\log \log d}{\log d}$ . In particular, if the SDP objective is about  $n/r$  this essentially implies that the graph is  $K_r$ -free (as roughly each vertex contributes about  $x_i = 1/r$ ). Thus, by Shearer's bound the integrality gap is  $(n/r)/\alpha(G) \leq d \frac{\log \log d}{\log d}$ . It is interesting to note that both Halperin's approach and Shearer's bound seem to get stuck at the same point.

Alon [Alo96] generalized the triangle-free result in a different direction, also using the entropy method. He considered locally  $k$ -colorable graphs, where the neighborhood of every vertex is  $k$ -colorable and showed that  $\alpha(G) = \Omega\left(\frac{n}{d} \frac{\log d}{\log^{k+1} d}\right)$ . Note that triangle-free graphs are locally 1-colorable. This result also holds under weaker conditions, and plays a key role in the results of [Ban15] on bounding the integrality gap of  $SA^+$  relaxations.

**Bounds on the chromatic number.** Most of the above results generalize to the much more demanding setting of list coloring. All of them are based on “nibble” method, but require increasingly sophisticated ideas. The intuition for why  $O(d/\log d)$  arises can be seen via a coupon-collector argument: if each vertex in the neighborhood  $N(v)$  chooses a color from  $s$  colors independently and u.a.r., they will use up all  $s$  colors unless  $d \leq O(s \log s)$ , or  $s \geq \Omega(d/\log d)$ . (Of course, the colors at the neighbors are not chosen uniformly or independently.) Kim showed that  $\chi_\ell(G) = O(d/\log d)$  for graphs with girth at least 5 [Kim95]. His idea was that for any  $v$ , and  $u, w \in N(v)$ ,  $N(u) \cap N(w) = \{v\}$  because of the girth, and hence the available colors at  $u, w$  evolve essentially independently, and hence conform to the intuition.

These ideas fail for triangle-free graphs (of girth 4): we could have a vertex  $v$ , with  $u, w \in N(v)$ ,



and  $N(u) = N(w)$  (i.e., all their neighbors are common). In this case the lists of available colors at  $u$  and  $w$  are far from independent: they would be *completely identical*. Johansson [Joh96a] had the crucial insight that this positive correlation is not a problem, since there is no edge between  $u$  and  $w$  (because of triangle-freeness!). His clever proof introduced the crucial notions of entropy and energy to capture and control the positive correlation along edges in such  $K_3$ -free graphs.

If there are triangles, say if the graphs are only locally  $k$ -colorable, then using these ideas naïvely fails. A next key new idea, also introduced by Johansson [Joh96b], is to actually modify the standard nibble process by introducing a *probability reshuffling* step at each vertex depending on its local graph structure, which makes it more complicated. In Section 6, we give his result for locally-colorable graphs in its entirety.

**Lift-and-project Hierarchies.** An excellent introduction to hierarchies and their algorithmic uses can be found in [CT12, Lau03]. Here, we only describe here the most basic facts that we need.

The Sherali-Adams (SA) 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. Such a solution can be viewed as specifying a local distribution over valid  $\{0, 1\}$ -solutions for each set  $S$  of size at most  $t + 1$ . A formal description of the  $t$ -round Sherali-Adams LP  $SA_{(t)}$  for the independent set problem can be found in [CT12, Lemma 1].

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 (4) on  $y_{ij}$  variables). We will call this the  $t$ -level  $SA^+$  formulation and denote it by  $SA_{(t)}^+$ . To keep the notation consistent with the LP (2), we will use  $x_i$  to denote the marginals  $y_{ii}$  on singleton vertices.

### 3 Integrality Gap

In this section, we show Theorem 1.1, that the integrality gap of the standard Lovász  $\vartheta$ -function based SDP relaxation is

$$O\left(d\left(\frac{\log \log d}{\log d}\right)^{3/2}\right) = \tilde{O}\left(d/\log^{3/2} d\right).$$

To show this we prove the following result (which is Theorem 1.2, restated):

**Theorem 3.1** *Let  $G$  be a  $K_r$ -free graph with maximum degree  $d$ . Then*

$$\alpha(G) = \Omega\left(\frac{n}{d} \max\left(\frac{\log d}{r \log \log d}, \left(\frac{\log d}{\log r}\right)^{1/2}\right)\right).$$

*In particular, for  $r = \log^c d$  with  $c \geq 1$ , we get  $\alpha(G) = \Omega\left(\frac{n}{d} \left(\frac{\log d}{c \log \log d}\right)^{1/2}\right)$ .*

We need the following basic facts. The first follows from a simple counting argument (see [Alo96, Lemma 2.2] for a proof).

**Lemma 3.2** *Let  $F$  be a family of  $2^{\varepsilon x}$  distinct subsets of an  $x$ -element set  $X$ . Then the average size of a member of  $F$  is at least  $\varepsilon x / (10 \log(1 + 1/\varepsilon))$ .*

**Fact 3.3** *Let  $G$  be a  $K_r$ -free graph on  $x$  vertices, then*

$$\alpha(G) \geq \max\left(\frac{x^{1/r}}{2}, \frac{\log x}{\log(2r)}\right).$$

*Note that the latter bound is stronger when  $r$  is large, i.e., roughly when  $r \geq \log x / \log \log x$ .*

**Proof:** Let  $R(s, t)$  denote the off-diagonal  $(s, t)$ -Ramsey number, defined as the smallest number  $n$  such that any graph on  $n$  vertices contains either an independent set of size  $s$  or a clique of size  $t$ .

It is well known that  $R(s, t) \leq \binom{s+t-2}{s-1}$  [ES35]. Approximating the binomial gives us the bounds  $R(s, t) \leq (2s)^t$  and  $R(s, t) \leq (2t)^s$ ; the former is useful for  $t \leq s$  and the latter for  $s \leq t$ . If we set  $R(s, t) = x$  and  $t = r$ , the first bound gives  $s \geq (1/2)x^{1/r}$  and the second bound gives  $s \geq \log x / \log(2r)$ . ■

We will be interested in lower bounding the number of independent sets  $\mathcal{I}$  in a  $K_r$ -free graph. Clearly,  $\mathcal{I} \geq 2^{\alpha(G)}$  (consider every subset of maximum independent set). However the following improved estimate will play a key role in Theorem 3.1. Roughly speaking it says that if  $\alpha(G)$  is small, in particular of size logarithmic in  $x$ , then the independent sets are spread all over  $G$ , and hence their number is close to  $x^{\Omega(\alpha(G))}$ .

**Theorem 3.4** *Let  $G$  be a  $K_r$ -free graph on  $x$  vertices, and let  $\mathcal{I}$  denote the number of independent sets in  $G$ . Then we have*

$$\log \mathcal{I} \geq \max\left(\frac{x^{1/r}}{2}, \frac{\log^2 x}{18 \log 2r}\right).$$

**Proof:** The first bound follows trivially from Fact 3.3, and hence we focus on the second bound. Also, assume  $r \geq 3$  and  $x \geq 64$  else the second bound is trivial.

Define  $s := \log x / \log(2r)$ . Let  $G'$  be the graph obtained by sampling each vertex of  $G$  independently with probability  $p := 2/x^{1/2}$ . The expected number of vertices in  $G'$  is  $px = 2x^{1/2}$ . Let  $\mathcal{G}$  denote the good event that  $G'$  has at least  $x^{1/2}$  vertices. Clearly,  $\Pr[\mathcal{G}] \geq 1/2$  (in fact it is exponentially close to 1). Since the graph  $G'$  is also  $K_r$ -free, conditioned on the event  $\mathcal{G}$ , it has an independent set of size at least  $\log(x^{1/2}) / \log(2r) = s/2$ . Thus the expected number of independent sets of size  $s/2$  in  $G'$  is at least  $1/2$ .

Now consider some independent set  $Y$  of size  $s/2$  in  $G$ . The probability that  $Y$  survives in  $G'$  is exactly  $p^{s/2}$ . As the expected number of independent sets of size  $s/2$  in  $G'$  is at least  $1/2$ , it follows that  $G$  must contain at least  $(1/2)(1/p^{s/2})$  independent sets of size  $s/2$ . This gives us that

$$\log \mathcal{I} \geq \frac{s}{2} \log\left(\frac{1}{p}\right) - 1 \geq \frac{s}{2} \log x^{1/2} - \frac{s}{2} - 1 \geq \frac{s}{18} \log x,$$

where the last inequality assumes that  $x$  is large enough. ■

We are now ready to prove Theorem 3.1.

**Proof:** We can assume that  $d \geq 16$ , else the claim is trivial. Our arguments follow the probabilistic approach of [She95, Alo96]. 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  $X_v = d|v \cap W| + |N(v) \cap W|$ .

Observe that  $|W|$  can be written as  $\sum_v |v \cap W|$ ; moreover, it satisfies  $|W| \geq (1/d) \sum_v |N(v) \cap W|$ , since a vertex in  $W$  can be in at most  $d$  sets  $N(v)$ . Hence we have that

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

Let  $\gamma = \max\left(\frac{\log d}{r \log \log d}, \left(\frac{\log d}{\log r}\right)^{1/2}\right)$  denote the improvement factor in Theorem 3.1 over the trivial bound of  $n/d$ . Thus to show that  $\alpha(G)$  is large, it suffices to show that

$$\mathbb{E}[X_v] \geq c\gamma \tag{5}$$

for each vertex  $v$  and some fixed constant  $c$ .

In fact, we show that (5) 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

$$\mathbb{E}[X_v \mid W \cap V(H) = S] \geq c\gamma.$$

Fix a choice of  $S$ . Let  $X$  denote the non-neighbors of  $S$  in  $N(v)$ , and let  $x = |X|$ . Let  $\varepsilon$  be such that  $2^{\varepsilon x}$  denotes the number of independent sets in the induced subgraph  $G[X]$ . Now, conditioning on the intersection  $W \cap V(H) = S$ , there are precisely  $2^{\varepsilon x} + 1$  possibilities for  $W$ : one in which  $W = S \cup \{v\}$ , and  $2^{\varepsilon x}$  possibilities in which  $v \notin W$  and  $W$  is the union of  $S$  with an independent set in  $G[X]$ .

By Lemma 3.2, the average size of an independent set in  $X$  is at least  $\frac{\varepsilon x}{10 \log(1/\varepsilon + 1)}$  and thus we have that

$$\mathbb{E}[X_v \mid W \cap V(H) = S] \geq d \frac{1}{2^{\varepsilon x} + 1} + \frac{\varepsilon x}{10 \log(1/\varepsilon + 1)} \frac{2^{\varepsilon x}}{2^{\varepsilon x} + 1} \tag{6}$$

Now, if  $2^{\varepsilon x} + 1 \leq \sqrt{d}$ , then the first term is at least  $\sqrt{d}$ , and we've shown (5) with room to spare. So we can assume that  $\varepsilon x \geq (1/2) \log d$ . Moreover, by Theorem 3.4,

$$\varepsilon x \geq \max\left(\frac{x^{1/r}}{2}, \frac{\log^2 x}{18 \log(2r)}\right)$$

and hence the right hand side in (6) is at least

$$\begin{aligned} & \frac{1}{40 \log(1/\varepsilon + 1)} \max\left(\frac{\log d}{2}, \frac{x^{1/r}}{2}, \frac{\log^2 x}{18 \log 2r}\right) \\ & \geq \frac{1}{40 \log(x + 1)} \max\left(\frac{\log d}{2}, \frac{x^{1/r}}{2}, \frac{\log^2 x}{18 \log 2r}\right), \end{aligned} \tag{7}$$

where the inequality uses  $\varepsilon \geq 1/x$  (since  $\varepsilon x \geq (1/2) \log d \geq 1$ ).

First, let's consider the first two expressions in (7). If  $x \geq \log^r d$ , then as  $x^{1/r}/\log(x + 1)$  is increasing in  $x$ , it follows that the right hand side of (7) is at least

$$\frac{x^{1/r}}{80 \log(x + 1)} = \Omega\left(\frac{\log d}{r \log \log d}\right).$$

On the other hand if  $x \leq \log^r d$ , then we have that the right hand side is again at least

$$\frac{1}{40 \log(x+1)} \frac{\log d}{2} = \Omega\left(\frac{\log d}{r \log \log d}\right).$$

Now, consider the first and third expressions in in (7). Using the fact that  $\max(a, b) \geq \sqrt{ab}$  with  $a = (\log d)/2$  and  $b = (\log^2 x)/(18 \log 2r)$ , we get that (7) is at least  $\Omega\left(\frac{\log d}{\log r}\right)^{1/2}$ . Hence, for every value of  $x$  we get that (7) is at least  $\Omega(\gamma)$  as desired in (5); this completes the proof of Theorem 3.1. ■

We can now show the main result of this section.

**Theorem 3.5** *The standard SDP for independent set has an integrality gap of*

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

**Proof:** Given a graph  $G$  on  $n$  vertices, let  $\beta \in [0, 1]$  be such that the SDP on  $G$  has objective value  $\beta n$ . If  $\beta \leq 2/\log^{3/2} d$ , the naïve greedy algorithm already implies a  $d/\log^{3/2} d$  approximation. Thus, we will assume that  $\beta \geq 2/\log^{3/2} d$ .

Let us delete all the vertices that contribute  $x_i \leq \beta/2$  to the objective. The residual graph has objective value at least  $\beta n - (\beta/2)n = \beta n/2$ .

Let  $\eta = 2 \log \log d / \log d$ . If there are more than  $n/\log^2 d$  vertices with  $x_i \geq \eta$ , applying Theorem 2.1 to the collection of these vertices already gives independent set of size at least

$$\Omega\left(\frac{d^{2n}}{d\sqrt{\ln d}} \cdot \frac{n}{\log^2 d}\right) = \Omega\left(\frac{n \log^{3/2} d}{d}\right),$$

and hence a  $O(d/\log^{3/2} d)$  approximation.

Thus we can assume that fewer than  $n/\log^2 d$  vertices have  $x_i \geq \eta$ . As each vertex can contribute at most 1 to the objective, the SDP objective on the residual graph obtained by deleting the vertices with  $x_i \geq \eta$  is at least  $\beta n/2 - n/(\log^2 d)$  which is at least  $\beta n/3$ , since  $\beta \geq 2/\log^{3/2} d$ .

So we have a feasible SDP solution on a subgraph  $G'$  of  $G$ , where the objective is at least  $\beta n/3$  (here  $n$  is the number of vertices in  $G$  and not  $G'$ ) and each surviving vertex  $i$  has value  $x_i$  in the range  $[\beta/2, \eta]$ .

As  $x_i \leq \eta$  for each  $i$ , and the SDP objective is at least  $\beta n/3$ , the number of vertices  $n'$  in  $G'$  satisfies  $n' \geq (\beta n/3)/\eta = \Omega(n\beta/\eta)$ . Moreover, as  $x_i \geq \beta/2$  for each vertex  $i \in G'$ , and the SDP does not put more than one unit of probability mass on any clique, it follows that  $G'$  is  $K_r$ -free for  $r = 2/\beta = \log^{3/2} d$ . Applying Theorem 3.1 to  $G'$  with parameter  $r = \log^{3/2} d$ , we obtain that  $G'$  has an independent set of size

$$\Omega\left(\frac{n'}{d} \sqrt{\frac{\log d}{\log r}}\right) = \Omega\left(\frac{n'}{d} \sqrt{\frac{\log d}{\log \log d}}\right) = \Omega\left(\frac{n\beta}{d\eta} \sqrt{1/\eta}\right) = \Omega\left(\frac{\beta n}{d} \cdot \eta^{-3/2}\right).$$

The SDP objective for  $G$  was  $\beta n$ , so the integrality gap is  $O(d\eta^{3/2}) = O(d(\frac{\log \log d}{\log d})^{3/2})$ . ■

### 3.1 An upper bound

We give a simple construction that  $\alpha(G) \leq \frac{n \log d}{d \log r}$  for  $r \geq \log d$ . We use the standard lower bound  $R(s, t) = \Omega(t^{s/2})$  for off-diagonal Ramsey numbers for  $t \geq s$ . Setting  $t = r$  with  $r \geq \log d$ , it follows that there exist  $K_r$ -free graphs  $H$  on  $d$  vertices such that  $\alpha(H) = O(\log d / \log r)$ . Now set  $G$  to be  $n/d$  disjoint copies of  $H$ .

## 4 An Algorithm using Lift-and-Project

In this section, we briefly illustrate how to make Bansal's argument about the integrality gap of the lifted SDP [Ban15] algorithmic. Consider the  $SA_{(d)}^+$  relaxation on  $G$ , and let  $\text{sdp}(G)$  denote its value. We can assume that

$$\text{sdp}(G) \geq n / \log^2 d, \quad (8)$$

otherwise the naive algorithm already gives a  $d / \log^2 d$  approximation.

Let  $\eta = 3 \log \log d / \log d$ , and  $Z$  denote the set of vertices  $i$  with  $x_i \geq \eta$ . We can assume that  $|Z| \leq n / (4 \log^2 d)$ , otherwise applying Theorem 2.1 gives an independent set of size  $\Omega(|Z| \cdot d^{2\eta} / (d \sqrt{\log d})) = \Omega(n \log^2 d / d)$ . Applying Theorem 2.1 is fine, since our solution belongs to  $SA^+$  and hence is a valid SDP solution. Hence,

$$\text{sdp}(G) \leq |Z| \cdot 1 + (n - |Z|) \cdot \eta \leq (n / (4 \log^2 d)) \cdot 1 + n \cdot \eta = 2\eta n.$$

Let  $V'$  denote the set of vertices  $i$  with  $x_i \in [1 / (4 \log^2 d), \eta]$ .

**Claim 4.1**  $|V'| \geq \text{sdp}(G) / (2\eta)$ .

**Proof:** The total contribution to  $\text{sdp}(G)$  of vertices  $i$  with  $x_i \leq 1 / (4 \log^2 d)$  can be at most  $n / (4 \log^2 d)$ , which by (8) is at most  $\text{sdp}(G) / 4$ . Similarly, the contribution of vertices in  $Z$  is at most  $|Z|$ , which is again at most  $\text{sdp}(G) / 4$ . Together this gives  $\text{sdp}(G') \geq \text{sdp}(G) / 2$ . As each vertex in  $V'$  has  $x_i \leq \eta$ , the claim follows.  $\blacksquare$

**Lemma 4.2** *The graph  $G' = G[V']$  induced on  $V'$  is locally  $k$ -colorable for  $k = O(\log^3 d)$ .*

**Proof:** Consider the solution  $SA_{(d)}^+$  restricted to  $G'$ . For a vertex  $v \in V'$ , let  $N(v)$  denote its neighborhood in  $G'$ . As  $|N(v)| \leq d$  and  $x_i \geq 1 / (4 \log^2 d)$  for all  $i \in N(v)$ , the  $SA_{(d)}^+$  solution defines a ‘‘local distribution’’  $\{X_S\}_{S \subseteq N(v)}$  over subsets of each neighborhood with the following properties:

- (i)  $X_S \geq 0$  and  $\sum_{S \subseteq N(v)} X_S = 1$ ,
- (ii)  $X_S > 0$  only if  $S$  is independent in the subgraph induced on  $N(v)$ , and
- (iii) for each vertex  $i \in N(v)$ , it holds that

$$x_i = \sum_{S \subseteq N(v): i \in S} x_S \geq 1 / (4 \log^2 d).$$

Scaling up the solution  $X_S$  by  $4 \log^2 d$  thus gives a valid fractional coloring of  $N(v)$  using  $4 \log^2 d$  colors, which by a set-covering argument implies that  $\chi(N(v)) = O(\log^2 d \cdot \log N(v)) = O(\log^3 d)$ .  
 ■

Using Johansson's coloring algorithm for locally  $k$ -colorable graphs (Theorem 6.1) we can find an independent set of  $G'$  with size

$$\text{alg}(G') = \Omega \left( \frac{|V'|}{d} \cdot \frac{\log d}{\log(k+1)} \right).$$

Using  $k = O(\log^3 d)$  and Claim 4.1 this implies an algorithm to find independent sets in degree  $d$  graphs, with an integrality gap of

$$\frac{\text{sdp}(G)}{\text{alg}(G)} \leq \frac{\text{sdp}(G)}{\text{alg}(G')} \leq O \left( \frac{d \eta \log(k+1)}{\log d} \right) = \tilde{O} \left( \frac{d}{\log^2 d} \right).$$

Our algorithm only required a fractional coloring on the neighborhood of vertices. Since they are at most  $2^d$  independent sets in each neighborhood, there are at most  $n \cdot 2^d$  relevant variables in our SDP. Hence, we can compute the relevant fractional coloring in time  $\text{poly}(n) \cdot 2^{O(d)}$ .

## 5 LP-based guarantees

We prove Corollary 1.4. Consider the standard LP (2) strengthened by the clique inequalities  $\sum_{i \in C} x_i \leq 1$  for each clique  $C$  with  $|C| \leq \log d$ . As each clique lies in the neighborhood of some vertex, the number of such cliques is at most  $n \cdot \binom{d}{\log d}$ . Let  $\beta n$  denote the objective value of this LP relaxation. We assume that  $\beta \geq 2/\log d$ , otherwise the naïve algorithm already gives a  $d/\log d$  approximation.

Let  $B_0$  denote the set of vertices with  $x_i \leq 1/\log d = \beta/2$ . For  $j = 1, \dots, k$ , where  $k = \log \log d$ , let  $B_j$  denote the set of vertices with  $x_i \in (2^{j-1}/\log d, 2^j/\log d]$ . Note that  $\sum_{j \geq 1} \sum_{i \in B_j} x_i = \beta n - \sum_{i \in B_0} x_i \geq \beta n/2$ , and thus there exists some index  $j$  such that  $\sum_{i \in B_j} x_i \geq \beta n/(2k)$ .

Let  $\gamma = 2^{j-1}/\log d$ ; for each  $i \in B_j$ ,  $x_i \in (\gamma, 2\gamma]$ . Since  $x_i > \gamma$  for each  $i \in B_j$ , the clique constraints ensure that the graph induced on  $B_j$  is  $K_r$ -free for  $r = 1/\gamma$ . Moreover, since  $x_i \leq 2\gamma$  for each  $i \in B_j$ ,  $|B_j| \geq \frac{1}{2\gamma} \cdot \frac{\beta n}{2k}$ . By Shearer's result for  $K_r$ -free graphs we obtain

$$\alpha(B_j) = \Omega \left( |B_j| \cdot \frac{\gamma \log d}{d \log \log d} \right) = \Omega \left( \frac{\beta n \log d}{d (\log \log d)^2} \right).$$

This implies the claim about the integrality gap.

A similar argument implies the constructive result. Let  $\beta n$  denote the value of the  $SA_{(d)}$  relaxation. As before, we assume that  $\beta \geq 2/\log d$  and divide the vertices into  $1 + \log \log d$  classes. Consider the class  $B_j$  with  $j \geq 1$  that contributes most to the objective, and use the fact that the graph induced on  $B_j$  is locally  $k$ -colorable for  $k = (\log d / 2^{j-1} \cdot \log d) = O(\log^2 d)$ . As in Section 4, we can now use Johansson's coloring algorithm Theorem 6.1 to find a large independent set.

## Acknowledgments

We thank Alan Frieze for sharing a copy of the manuscript of Johansson with us. We thank Noga Alon, Tom Bohman, Alan Frieze, and Venkatesan Guruswami for enlightening discussions.

## References

- [AEKS81] Miklós Ajtai, Paul Erdős, János Komlós, and Endre Szemerédi. On Turán’s theorem for sparse graphs. *Combinatorica*, 1(4):313–317, 1981. [1](#), [2](#)
- [AK98] Noga Alon and Nabil Kahale. Approximating the independence number via the  $\vartheta$ -function. *Math. Programming*, 80(3, Ser. A):253–264, 1998. [1](#), [1.1](#), [2](#)
- [AKS80] Miklós Ajtai, János Komlós, and Endre Szemerédi. A note on Ramsey numbers. *J. Comb. Theory, Ser. A*, 29(3):354–360, 1980. [2](#)
- [AKS99] Noga Alon, Michael Krivelevich, and Benny Sudakov. Coloring graphs with sparse neighborhoods. *J. Comb. Theory, Ser. B*, 77(1):73–82, 1999. [1.1](#)
- [AKS11] 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. [1](#)
- [Alo96] Noga Alon. Independence numbers of locally sparse graphs and a Ramsey type problem. *Random Struct. Algorithms*, 9(3):271–278, 1996. [1](#), [1.1](#), [1.1](#), [2](#), [3](#), [3](#)
- [AS92] Noga Alon and Joel Spencer. *The Probabilistic Method*. Wiley Interscience, New York, 1992. [2](#), [A.1](#)
- [Ban15] Nikhil Bansal. Approximating independent sets in sparse graphs. In *SODA*, 2015. [1](#), [1.1](#), [2](#), [2](#), [4](#)
- [BHK09] Andreas Björklund, Thore Husfeldt, and Mikko Koivisto. Set partitioning via inclusion-exclusion. *SIAM J. Comput.*, 39(2):546–563, 2009. [6.3](#)
- [Cha13] Siu On Chan. Approximation resistance from pairwise independent subgroups. In *STOC*, pages 447–456, 2013. [1](#)
- [CPS14] Kai-Min Chung, Seth Pettie, and Hsin-Hao Su. Distributed algorithms for the Lovász local lemma and graph coloring. In *ACM Symposium on Principles of Distributed Computing, PODC ’14, Paris, France, July 15-18, 2014*, pages 134–143, 2014. [1.1](#)
- [CT12] 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. [2](#)
- [ES35] Paul Erdős and George Szekeres. A combinatorial problem in geometry. *Compositio Mathematica*, 2:463–470, 1935. [3](#)
- [Fei04] Uriel Feige. Approximating maximum clique by removing subgraphs. *SIAM J. Discrete Math.*, 18(2):219–225, 2004. [1](#)
- [FJ14] Uriel Feige and Shlomo Jozeph. Separation between estimation and approximation. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:110, 2014. [1](#)
- [FM13] Alan Frieze and Dhruv Mubayi. Coloring simple hypergraphs. *Journal of Combinatorial Theory, Series B*, 103(6):767 – 794, 2013. [1.1](#)
- [GKPS06] Rajiv Gandhi, Samir Khuller, Srinivasan Parthasarathy, and Aravind Srinivasan. Dependent rounding and its applications to approximation algorithms. *J. ACM*, 53(3):324–360, 2006. [1](#)
- [GLS88] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*. Springer-Verlag, Berlin, 1988. [2](#)
- [GM12] Bernd Gärtner and Jiří Matoušek. *Approximation algorithms and semidefinite programming*. Springer, Heidelberg, 2012. [2](#)
- [GP00] David A. Grable and Alessandro Panconesi. Fast distributed algorithms for Brooks-Vizing colorings. *J. Algorithms*, 37(1):85–120, 2000. [1.1](#)

- [Hal00] Magnús M. Halldórsson. Approximations of weighted independent set and hereditary subset problems. *J. Graph Algorithms Appl.*, 4:no. 1, 16 pp., 2000. [1](#), [2](#)
- [Hal02] Eran Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *SIAM J. Comput.*, 31(5):1608–1623, 2002. [1](#), [1.1](#), [2](#), [2.1](#)
- [Hås96] Johan Håstad. Clique is hard to approximate within  $n^{1-\epsilon}$ . In *FOCS*, pages 627–636, 1996. [1](#)
- [HR94] 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. [1](#), [2](#)
- [Joh96a] Anders Johansson. Asymptotic choice number for triangle-free graphs. preprint, 1996. [1.1](#), [2](#), [6.1](#), [7](#)
- [Joh96b] Anders Johansson. The choice number of sparse graphs. preprint, August 1996. [1.1](#), [2](#), [6](#)
- [Kim95] Jeong Han Kim. On Brooks’ theorem for sparse graphs. *Combinatorics, Probability & Computing*, 4:97–132, 1995. [2](#)
- [KMS98] David R. Karger, Rajeev Motwani, and Madhu Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, 1998. [1](#)
- [KP06] Subhash Khot and Ashok Kumar Ponnuswami. Better inapproximability results for MaxClique, chromatic number and Min-3Lin-Deletion. In *ICALP (1)*, pages 226–237, 2006. [1](#)
- [Lau] Monique Laurent. Networks and semidefinite programming (lecture notes 2014). <http://homepages.cwi.nl/~monique/lnmb14/lnmb14.pdf>. [2](#)
- [Lau03] 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. [2](#)
- [Lov79] László Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, 1979. [1](#)
- [MR02] Michael Molloy and Bruce Reed. *Graph colouring and the probabilistic method*, volume 23 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2002. [1.1](#), [6.1](#), [7](#)
- [MT10] Robin A. Moser and Gábor Tardos. A constructive proof of the general Lovász local lemma. *J. ACM*, 57(2):Art. 11, 15, 2010. [8.1.1](#), [A.2](#)
- [She83] James B. Shearer. A note on the independence number of triangle-free graphs. *Discrete Mathematics*, 46(1):83–87, 1983. [2](#)
- [She95] James B. Shearer. On the independence number of sparse graphs. *Random Struct. Algorithms*, 7(3):269–272, 1995. [1](#), [1.1](#), [2](#), [3](#)
- [SS12] Warren Schudy and Maxim Sviridenko. Concentration and moment inequalities for polynomials of independent random variables. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 437–446. ACM, New York, 2012. [A.1](#), [A.1](#)
- [Vu02] V.H. Vu. A general upper bound on the list chromatic number of locally sparse graphs. *Combinatorics Probability and Computing*, 11(1):103–111, 2002. [1.1](#)

## 6 Johansson’s Algorithm for Coloring Sparse Graphs

For completeness, we give proofs for two results of Johansson [Joh96b] on coloring degree- $d$  graphs: one about graphs where vertex neighborhoods can be colored using few colors (“locally-colorable” graphs), and another about  $K_r$ -free graphs.



**Theorem 6.1** *For any  $r, \Delta$ , there exists a randomized algorithm that, given a graph  $G$  with maximum degree  $\Delta$  such that the neighborhood of each vertex is  $r$ -colorable, outputs a proper coloring of  $V(G)$  using  $O(\frac{\Delta}{\ln \Delta} \ln r)$  colors in expected  $\text{poly}(n2^\Delta)$  time.*

**Theorem 6.2** *For any  $r, \Delta$ , there exists a randomized algorithm that, given a graph  $G$  with maximum degree  $\Delta$  which excludes  $K_r$  as a subgraph, outputs a proper coloring of  $V(G)$  using  $O(\frac{\Delta}{\ln \Delta}(r^2 + r \ln \ln \Delta))$  colors in expected  $\text{poly}(n)$  time.*

We emphasize that Johansson’s manuscript contains proofs of other results and extensions, such as colorability under weaker conditions than above, and extensions to list-coloring; we omit these extensions for now. Our presentation largely follows his, but streamlines some of the proofs using techniques that have developed since, such as concentration bounds for low-degree polynomials of variables, and dependent rounding techniques. Roadmap: we first give the intuition in Section 6.1. We give the proof of Theorem 6.1 in §6.3–§8.5, and then show how to extend it to  $K_r$ -free graphs in §9.

## 6.1 Overview and Ideas

Johansson’s algorithm for locally-colorable graphs uses the “nibble” approach: in each round, some  $\theta > 0$  fraction of vertices get colored from their currently-allowable colors. The goal is to argue (using concentration of measure, and the Local Lemma) that the degree of each surviving vertex goes down exponentially like  $(1 - O(\theta))^t$ , whereas the number of colors does not decrease very fast. This means that after  $\approx (\varepsilon/\theta) \ln \Delta$  rounds the degree of the remaining vertices is smaller than  $\Delta^{1-\varepsilon}$  before running out of the prescribed number of colors, at which point even the naïve greedy algorithm can color the remaining vertices with a few more colors. The proof for the degree reduction uses concentration bounds for quadratic polynomials of random variables. The real challenge is to lower-bound the number of remaining colors. Johansson’s argument shows that the entropy of the probability distribution of a vertex over its colors remains high throughout the process, and hence there must be many colors available. This requires a carefully orchestrated process, which we describe next.

In more detail (but still at a high level): in each round, some  $\theta \approx \Delta^{-1/4}$  fraction of the vertices get activated, and each tentatively chooses a color from its own probability distribution. (This per-vertex distribution is initially the uniform distribution.) Any vertex that gets the same color as its neighbor rejects its color; since the number of these is small, we can ignore these for now. Then each tentatively colored vertex (say  $v$  with color  $\gamma$ ), with probability  $\frac{1}{2}$  accepts color  $\gamma$  permanently and deletes the probability mass corresponding to  $\gamma$  from its neighbors (so that they cannot take color  $\gamma$ ); with the remaining probability  $\frac{1}{2}$ ,  $v$  rejects color  $\gamma$  for this round and waits for another round. In order to ensure the total probability mass at each vertex remains about 1, since the first option caused the probability mass for color  $\gamma$  to decrease at the neighbors, the second option must increase color  $\gamma$ ’s mass at the neighbors. If two of these neighbors  $u, w$  are connected by an edge, this means that we’re increasing the chance that both these will get color  $\gamma$ ; this is potentially worrisome.

This problem does not arise if the graph is triangle-free, because there are no edges in the neighborhood of any vertex. In this case Johansson’s previous preprint [Joh96a] argued that the entropy of each vertex’s distribution remains high—itsself a clever and delicate argument (see [MR02, Chapters 12-13]). However, if we just assume that the graph is locally  $r$ -colorable, the existence of edges

in node  $v$ 's neighborhood means that the probability mass for a color at both endpoints of an edge may become higher, creating undesirable positive correlations. What Johansson's new proof does is simple but ingenious: it "reshuffles" the measure for the color randomly to some independent set in the neighborhood. This is where the  $r$ -colorability condition kicks in: since there are large independent sets (of size  $\Delta/r$ ) in each neighborhood, the reshuffling does not change the probabilities too suddenly. Now carefully applied concentration bounds and LLL show a similar behavior as in the triangle-free case, and proves Theorem 6.1.

The argument for  $K_r$ -free graphs requires a more involved recursive reshuffling operation: in this case the size of the independent sets may be too small (if we just use Ramsey's bound, for instance), so the idea is to move the measure (on average) to sets that avoid  $K_t$  for  $t$  smaller than  $r$ . This process (which Johansson calls a "trimming modifier") creates a very slight negative correlation on the edges, but this suffices to show Theorem 6.2.

## 6.2 Notation and Preliminaries

We now define some notation and concepts, and give properties useful for the following proofs. We will interchangeably use  $u \sim v$  and  $u \in N(v)$  to denote that  $u$  and  $v$  are adjacent.

### 6.2.1 Mean-One Random Variables

An r.v.  $X$  is a *mean-one random variable* (m.o. r.v.) if  $X$  only takes on values in  $\{0\} \cup [1, \infty)$ , and  $\mathbb{E}[X] = 1$ . One simple class of m.o. r.v.s take on some value  $c \geq 1$  w.p.  $1/c$ , and 0 w.p.  $1 - 1/c$ .

### 6.2.2 The Stopped Product

Given a sequence of non-negative random variables  $Y_1, Y_2, \dots, Y_m$ , and a "threshold" value  $a \geq 0$ , define a stopping time  $\tau_a$  as

$$\tau_a := \min \left\{ t \mid \prod_{i \leq t} Y_i \geq a \right\}$$

Then the *stopped product*  $\overline{\prod}_i Y_i$  is defined as

$$\overline{\prod}_i Y_i := \prod_{i \leq \min(m, \tau_a)} Y_i$$

### 6.2.3 The $\kappa$ and $\hat{\kappa}$ Functions

For a random variable  $X$ , define the function

$$\kappa(X) := \mathbb{E}[X \ln X]. \tag{9}$$

The following facts are easy to verify, and will be useful in calculations.

- (a) If  $X, Y$  are independent, then  $\kappa(XY) = \kappa(X) \mathbb{E}[Y] + \kappa(Y) \mathbb{E}[X]$ .
- (b) Hence if  $X, Y$  are independent m.o. r.v.s, then  $\kappa(XY) = \kappa(X) + \kappa(Y)$ .
- (c) Also,  $\kappa(aX) = \mathbb{E}[X](a \ln a) + a\kappa(X)$ .
- (d) For an event  $\mathcal{E}$  and the associated m.o. r.v.  $X = \frac{\mathbf{1}(\mathcal{E})}{\Pr[\mathcal{E}]}$ ,  $\kappa(X) = \ln(1/\Pr[\mathcal{E}])$ .
- (e) For a stopped product  $X = \overline{\prod}_i X_i$  (as defined in §6.2.2) of independent m.o. r.v.s with respect to some threshold  $a$ ,

$$\kappa(X) \leq \sum_i \kappa(X_i).$$

(f) If  $X = (1 - \mathbf{1}_{(\mathcal{E})}) + \mathbf{1}_{(\mathcal{E})} \cdot Y$ , and  $Y$  is independent of the event  $\mathcal{E}$ , then  $\kappa(X) = \Pr[\mathcal{E}] \kappa(Y)$ .

It is also useful to define  $\widehat{\kappa}(X)$  as an absolute upper bound on  $X$ :

$$\widehat{\kappa}(X) := \inf\{c \mid X \leq c \text{ a.s.}\}. \quad (10)$$

### 6.3 The Algorithm for Locally-Colorable Graphs

Let us present the algorithm for Theorem 6.1 about finding colorings of  $r$ -locally-colorable graphs. The proof follows in §7.

Let  $s = O(\Delta \frac{\ln \Delta}{\ln r})$  be the number of colors we are aiming for, and  $L$  is the set of  $s$  colors. For vertex  $v$ , let  $\mathcal{P}_v = \{v\} \times L$  be a collection of tuples indicating which colors are still permissible for  $v$ . The term **whp** $_{\Delta}$  denotes “with probability at least  $1 - 1/\text{poly}(\Delta)$ ”.

We follow the algorithmic outline from the overview in §6.1. The algorithm starts with each vertex  $v$  having a uniform probability distribution  $p^0(v, \gamma) = 1/s$  over the colors  $\gamma \in L$ . In each stage  $t$  we pick some vertices from the current graph  $G^t$  and color them based on the current values of  $p^t(v, \gamma)$ , then update the probability distributions of the other vertices to get  $p^{t+1}(v, \gamma)$ , drop the colored vertices to get  $G^{t+1}$ , and proceed to the next stage. (This is the so-called “nibble”.) The goal is to show that after sufficiently many stages we have a partial proper coloring using at most  $s$  colors, and the degree of the graph induced by the yet-uncolored vertices is  $\Delta^{1-\varepsilon}$ . We can then use a greedy algorithm to color the remaining vertices.

The process in a generic stage  $t$  is as follows (we drop superscripts of  $t$  to avoid visual clutter).

1. Let  $\widehat{p} \in (0, 1)$  be a threshold to be defined later. Define

$$p_a(v, \gamma) := p(v, \gamma) \cdot \mathbf{1}_{(p(v, \gamma) \leq \widehat{p})} \quad (11)$$

$$p_c(v, \gamma) := p_a(v, \gamma) \cdot \mathbf{1}_{(\sum_{u \sim v} p(u, \gamma) \leq 100 \ln \Delta)}. \quad (12)$$

Hence  $p_a$  zeroes out any (vertex, color) tuple  $(v, \gamma)$  which has a high value, and  $p_c$  additionally zeroes out  $(v, \gamma)$  when  $v$ 's neighbors have a lot of probability mass on color  $\gamma$ .<sup>4</sup>

2. For each vertex  $v$  in the current graph  $G$  and each color  $\gamma$  in  $L$ , independently flip a coin with probability  $\theta p_c(v, \gamma)$ . (The parameter  $\theta \in (0, 1)$  is defined later.) Let  $A_{(v, \gamma)}$  be this indicator variable. If  $A_{(v, \gamma)} = 1$  then color  $\gamma$  is *tentatively assigned* to  $v$ . Many colors may be tentatively assigned to  $v$ .

Also, let  $\eta_{(v, \gamma)} \sim \text{Bin}(1/2)$  be an unbiased coin-flip independent of all else.

3. For  $v \in V$ , let  $\mathcal{T}_v = \{(v, \gamma) \mid A_{(v, \gamma)} = 1\}$ , and let  $\mathcal{T} = \cup_v \mathcal{T}_v$  be all the tentatively assigned tuples. Let

$$\mathcal{C}_v = \{(v, \gamma) \mid A_{(v, \gamma)} = 1 \wedge \eta_{(v, \gamma)} = 1 \wedge A_{(u, \gamma)} = 0 \forall u \sim v\},$$

and let  $\mathcal{C} = \cup_v \mathcal{C}_v$  similarly. Note that the pair  $(v, \gamma) \in \mathcal{T}$  is dropped from  $\mathcal{C}$  if any neighbor of  $v$  is tentatively assigned color  $\gamma$  (this ensures proper colorings), or if its own coin-flip  $\eta_{(v, \gamma)}$  comes up tails (this gives us a “damping” that is useful for the rest of the process).

---

<sup>4</sup>The definition of  $p_c$  is not required to prove Theorem 6.1, but is useful in extending the result to  $K_r$ -free graphs. The reader only interested in the former result should think of  $p_c = p_a$  for this discussion.

4. For  $v \in G$ , if there exists some  $\gamma$  such that  $(v, \gamma) \in \mathcal{C}$ , then color  $v$  with an arbitrary such  $\gamma$  and remove  $v$  from  $G'$ . So the events  $\{v \in G'\} = \{\mathcal{C}_v = \emptyset\}$ .

Now comes the changing of the probabilities via the “modifiers” (which are simply mean-one r.v.s, as defined in §6.2.1).

5. For each pair  $(w, \gamma)$ , generate modifiers  $M_{(v, \gamma)}^w$  for all  $v \in N(w)$ :
- If  $(w, \gamma) \notin \mathcal{T}$  (i.e.,  $A_{(w, \gamma)} = 0$ ) then  $M_{(v, \gamma)}^w = 1$  for all  $v \in N(w)$ .
  - Else, if  $(w, \gamma) \in \mathcal{T}$ , then for all  $v \in N(w)$ ,

$$M_{(v, \gamma)}^w = \underbrace{2(1 - \eta_{(w, \gamma)})}_{m.o.r.v.} \cdot \underbrace{r \mathbf{1}_{(v \in S(w, \gamma))}}_{m.o.r.v.} \quad (13)$$

where  $S(w, \gamma) \subseteq N(w)$  is a random color class from an  $r$ -coloring of  $N(w)$ , this randomness is independent of all other  $(w, \gamma')$ , and  $(w', \cdot)$ .

6. For each pair  $(v, \gamma)$ , collect modifiers  $M_{(v, \gamma)}^w$  from its neighbors  $w \sim v$ , and define  $M_{(v, \gamma)} := \overline{\prod}_{w \sim v} M_{(v, \gamma)}^w$ . Here  $\overline{\prod}$  is the stopped product (as in §6.2.2) w.r.t. threshold  $\widehat{p}/p(v, \gamma)$ .<sup>5</sup> Finally set the probability values for the next stage to be

$$p'(v, \gamma) := p(v, \gamma) \cdot M_{(v, \gamma)} \quad (14)$$

Recalling the  $\widehat{\kappa}$  function from (10), observe that  $\widehat{\kappa} := \max_{w, v, \gamma} \widehat{\kappa}(M_{(v, \gamma)}^w) = 2r$ ; hence the stopped product ensures  $p'(v, \gamma) < \widehat{\kappa} \cdot \widehat{p} = 2r\widehat{p}$ ; define  $p^* = \widehat{\kappa}\widehat{p}$ .

We’ve now finished defining the new probabilities  $p^{t+1}(v, \gamma) := p'(v, \gamma)$  and the new graph  $G^{t+1} = G'$ , and proceed to the next stage. This is done for some  $T = \Theta(\frac{1}{\theta} \ln \Delta)$  many stages, after which we claim that the degree of  $G^T$  becomes  $\ll s$ , and a naïve coloring suffices to color the rest of the vertices.

The run time: the most expensive step is to find an  $r$ -coloring of the neighborhood of the vertices. For each vertex this can be done in time  $O(r \cdot \Delta \cdot 2^\Delta)$  using  $O(2^\Delta)$  space (see [BHK09]). That is followed by  $T = O(\frac{\ln \Delta}{\theta})$  rounds of partial colorings, each taking  $\text{poly}(n\Delta)$  time. Hence the total runtime is  $O(nr \cdot 2^\Delta) + \text{poly}(n)$ .

## 7 The Proof of Theorem 6.1: I. The Setup

The analysis of this coloring algorithm is similar in spirit (but more technical than) Johansson’s previous result for coloring triangle-free graphs of maximum degree  $\Delta$ . Although that result also appears only as an unpublished manuscript [Joh96a], a lucid presentation appears in the book by Molloy and Reed [MR02].

The idea is clever, but also natural in hindsight: as the process goes on and probability values  $p(v, \gamma)$  increase, we want to show that for each surviving vertex, its degree goes down rapidly, whereas it has many colors still remaining. Showing the former, that many vertices in each neighborhood

<sup>5</sup>The stopped product is with respect to a sequence, so let us assume a total order on the vertices, and the variables  $\{M_{(v, \gamma)}^w\}_{w \in N(v)}$  are considered in this order.

are colored at each step, proceeds by showing that for each  $v$ , there is not too much positive correlation between the colors of its neighbors, and they behave somewhat independently. To show the latter, that many colors remain valid for each vertex  $v$ , we show that the entropy of the  $p(v, \cdot)$  “probability distribution” remains high. (The quotes are because we only have  $\sum_{\gamma} p(v, \gamma) \approx 1$ , and not equal to one, but this approximate equality suffices.) And lower-bounding the entropy, as in [Joh96a, MR02], relies on upper-bounding the “energy” of the edges which captures the positive correlation between the colors of its endpoints, and is defined as follows:

**Definition 7.1 (Energy)** *For an edge  $uv$  the energy with respect to the  $p$  values is*

$$\xi(uv, p) := \sum_{\gamma} p(u, \gamma)p(v, \gamma).$$

*For a non-edge  $uv$ , define  $\xi(uv, p) := 0$ . For a graph  $G$ , the energy of a vertex  $u$  is  $\xi_G(u, p) := \sum_{v \in G} \xi(uv, p)$ ; when the graph is clear from context we drop the subscript and use just  $\xi(u, p)$ .*

## 7.1 The Parameters

For ease of reference, we present the parameters used in the proof here. Some of these have already been used in the algorithm description, the others will be introduced in due course.

$$\begin{array}{lll} \varepsilon := 1/100 & \theta := \Delta^{-1/4+2\varepsilon} & s := |L| := \frac{\Delta}{K} \\ \widehat{\kappa} := \max_{w, v, \gamma} \widehat{\kappa}(M_{(v, \gamma)}^w) = 2r & \widehat{p} := \Delta^{-3/4-5\varepsilon} & p^* := \widehat{\kappa}\widehat{p} = 2r\widehat{p} \\ a := 1/2 - 3\varepsilon & \hat{c} := 0 & b := a - \hat{c} \\ C := \max(1, \ln(\widehat{\kappa}(M_{(v, \gamma)}^w))) = \ln(2r) & T := \frac{2\varepsilon}{a \cdot \theta} \ln \Delta & K := \frac{(b/4)\varepsilon \ln \Delta}{C} \end{array}$$

We will assume that  $r \leq \Delta^\varepsilon = \Delta^{1/100}$ , else the desired coloring number of  $O(\frac{\Delta}{\ln \Delta} \ln r)$  will just be  $O(\Delta)$ , which is trivial to achieve. Finally, we will assume that  $\Delta$  is large enough whenever necessary. In particular,

$$\ln \Delta \geq 10000 \tag{15}$$

suffices for the rest of the analysis, however no attempt has been made to optimize any constants.

## 7.2 Initial Values

At the beginning, the probability values are  $p^0(v, \gamma) = 1/s$  for all  $(v, \gamma)$ , and the degrees are at most  $\Delta$ . Hence

$$p^0(\mathcal{P}_v) = 1 \tag{P0}$$

$$d(v, G^t) \leq \Delta \tag{D0}$$

$$\xi_{G_0}(v, p^0) = \sum_{\gamma, u \sim v} p^0(v, \gamma)p^0(u, \gamma) \leq \Delta \cdot s \cdot 1/s^2 = K \tag{E0}$$

$$h(v, p^0) = \sum_{\gamma} p^0(v, \gamma) \ln 1/p^0(v, \gamma) = \ln s = \ln \Delta - \ln K \tag{H0}$$

### 7.3 The Invariants

The proof is by induction over the stages. We maintain the invariants that for all  $t \leq T$ , for all  $v \in G^t$ , the following hold true:

$$\begin{aligned}
p^t(\mathcal{P}_v) &\in 1 \pm t\lambda_P \subseteq 1 \pm \sqrt{\theta} \subseteq 1 \pm \varepsilon \subseteq [1/2, 2] && ((\text{InvP})) \\
d(v, G^t) &\leq \Delta e^{-a\theta t} + t\lambda_D && ((\text{InvD})) \\
\xi_{G^t}(v, p^t) &\leq K e^{-b\theta t} + \frac{\lambda_E}{b\theta} \leq 2K && ((\text{InvE})) \\
h(v, p^t) &\geq h(v, p^0) - \theta C \sum_{t' \leq t} \xi_{G_{t'}}(v, p_{t'}^t) - t\lambda_H \\
&\geq (1 - \varepsilon) \ln \Delta && ((\text{InvH}))
\end{aligned}$$

where

$$\begin{aligned}
\lambda_P &= O(\sqrt{p^* \ln \Delta}) & \lambda_D &= O(K\theta^2 \Delta) \\
\lambda_E &= O(\theta \Delta^{-2\varepsilon} \ln \Delta) & \lambda_H &= O(\sqrt{p^* \ln^3 \Delta})
\end{aligned} \tag{16}$$

In the following, we will show that the tightest bounds hold for each  $t$ . The weaker bounds given above are just implications useful in our proofs, and in all cases these follow by algebra. Observe the initial values from §7.2 satisfy these invariants.

## 8 The Proof of Theorem 6.1: II. The Inductive Step

We assume the invariants hold for all times upto and including time  $t$ , and now want to show these are satisfied at the end of stage  $t + 1$ . As usual, we use  $p$  to denote  $p^t$ , and  $p'$  to denote  $p^{t+1}$ . The plan is to show that for each fixed vertex  $v$ , each of the invariants hold with high probability. Since we cannot take a union bound without losing terms dependent on  $n$ , we show that the failure events depend only on a small number of other failure events, whereupon we can apply the Lovász Local Lemma to complete the argument.

In the following arguments, we assume that  $d(v) \geq \frac{\Delta}{\ln \Delta}$  for all  $v \in G_t$ . Indeed, at the beginning of any stage we can repeatedly remove vertices with degree less than  $\frac{\Delta}{\ln \Delta}$ , and having found a coloring for the rest of the vertices, we can color these removed vertices greedily at the end.

### 8.1 The Total Probability

By construction, the probability values  $p^t(v, \gamma)$  form a martingale, and hence it is not surprising that their sum remains concentrated around 1. Here is the formal proof.

**Lemma 8.1** *For vertex  $v$ , if  $p(\mathcal{P}_v) \in 1 \pm \varepsilon$ , then  $p'(\mathcal{P}_v) = p(\mathcal{P}_v) \pm O(\sqrt{p^* \ln \Delta})$  **whp** $_{\Delta}$ .*

**Proof:** By construction of the modifiers as mean-one r.v.s,  $\mathbb{E}[p'(v, \gamma)] = p(v, \gamma) \cdot \mathbb{E}[M_{(v, \gamma)}] = p(v, \gamma)$  for all  $(v, \gamma)$ . Moreover,  $p'(\mathcal{P}_v)$  is the sum of independent  $p^*$ -bounded random variables, with  $\mathbb{E}[p'(\mathcal{P}_v)] \leq (1 + \varepsilon) \leq 2$ , Theorem A.1 implies that the deviation  $|p'(\mathcal{P}_v) - p(\mathcal{P}_v)|$  is at most  $\lambda_P := O(\sqrt{p^* \ln \Delta} + p^* \ln \Delta) = O(\sqrt{p^* \ln \Delta})$  **whp** $_{\Delta}$ . ■



### 8.1.1 The Lovász Local Lemma Argument

To get the property of Lemma 8.1 for all vertices  $v$  simultaneously requires the LLL, since we cannot take a union bound over all the  $n$  vertices. For this, define the bad event  $\mathcal{B}_v^p = \{\omega : p'(\mathcal{P}_v) > p(\mathcal{P}_v) + \lambda_P\}$ . Note that this event depends only on a subset of the variables  $\{A_{(u,\gamma)}, \eta_{(u,\gamma)}, M_{(u,\gamma)}\}$  for  $u \in \{v\} \cup N(v)$  and  $\gamma \in L$ , and all these  $A, \eta, M$  variables are independent. Hence,  $\mathcal{B}_v^p$  and  $\mathcal{B}_w^p$  are clearly independent if  $N(u) \cap N(v) = \emptyset$ , and the dependency graph has degree at most  $\Delta^2 s \leq \Delta^3$ . Using the Moser-Tardos framework [MT10] we can get a coloring where none of these bad events happen.

For each of the the other invariants, we use a similar approach using the LLL: we define “local” bad events — i.e., the bad event at a vertex will depend only on r.v.s for vertices at some constant distance from it — and hence the degree of the dependency graph over the bad events will be bounded by  $\text{poly}(\Delta)$ . Moreover, the probability of each bad event will be at most  $1/\text{poly}(\Delta)$ . Using the Moser-Tardos framework will allow us to find an outcome that will avoid all the bad events at all vertices simultaneously. Since the arguments will be very similar, we will henceforth just explain what the bad events are, and omit the details.

## 8.2 The Degree

**Lemma 8.2** *For vertex  $v$ , whp $_{\Delta}$  the new degree is*

$$d'(v) \leq (1 - a\theta) d(v) \pm O(K\theta^2 \Delta).$$

It is more convenient to study  $X_u$ , the indicator of whether  $u$  was colored in this round; let  $X = \sum_{u \sim v} X_u$ . Then  $d'(v) = d(v) - X$ . We will first show that  $\mathbb{E}[X]$  is large, and then that  $X$  is concentrated around its mean. Let

$$Y_u := 1 - \prod_{\gamma} (1 - A_{(u,\gamma)} \eta_{(u,\gamma)}) \tag{17}$$

indicate whether  $u$  was tentatively assigned at least one color that was not dropped due to the  $\eta$  coin flip; clearly  $X_u \leq Y_u$ . Moreover, let

$$Y'_u := \sum_{\gamma} \sum_{w \sim u} A_{(u,\gamma)} A_{(w,\gamma)}. \tag{18}$$

It is easy to see that  $Y_u - Y'_u \leq X_u$ .

**Claim 8.3**  $\Pr[u \text{ is colored}] = \mathbb{E}[X_u] \geq (\frac{1}{2} - \varepsilon) \theta \geq (a + \varepsilon) \theta$ .

**Proof:** By inclusion-exclusion,

$$\mathbb{E}[Y_u] \geq \sum_{\gamma} \frac{\theta}{2} p_c(u, \gamma) - \sum_{\gamma, \gamma'} \frac{\theta^2}{4} p_c(u, \gamma) p_c(u, \gamma') = \frac{\theta}{2} p_c(\mathcal{P}_u) \left( 1 - \frac{\theta}{2} p_c(\mathcal{P}_u) \right).$$

Using (InvP), (InvE) and (InvH) in Lemma A.5, we know that  $p_c(\mathcal{P}_u) \geq 1 - 6(\sqrt{\theta} + \theta) - 2\varepsilon \geq 1 - 12\sqrt{\theta} - 2\varepsilon$ . Also, by (InvP),  $p_c(\mathcal{P}_u) \leq p(\mathcal{P}_u) \leq 2$ , so by algebra we get  $\mathbb{E}[Y_u] \geq \theta(1/2 - \varepsilon) - O(\theta^{3/2})$ .

Moreover,

$$\mathbb{E}[Y'_u] = \theta^2 \sum_{\gamma} \sum_{w \sim u} p_c(u, \gamma) p_c(w, \gamma) = \theta^2 \xi(u, p_c).$$

Using that  $\xi(u, p_c) \leq \xi(u, p) \leq 2K$  for all  $u$  (from **(InvE)**), we infer  $\mathbb{E}[Y'_u] \leq 2K\theta^2$ . Since  $K\theta^2 = O(\theta^{3/2})$ ,

$$\Pr[u \text{ colored}] = \mathbb{E}[X_u] \geq \mathbb{E}[Y_u - Y'_u] \geq \theta(1/2 - \varepsilon) - O(\theta^{3/2}) \geq (\frac{1}{2} - 2\varepsilon) \theta$$

The final inequality holds for large enough  $\Delta$  **(15)**, and the claim follows using the definition of  $a$ .  $\blacksquare$

**Corollary 8.4**  $\mathbb{E}[d'(v)] \leq (1 - a\theta) \cdot d(v)$

**Proof:** Suppose  $X := \sum_{u \sim v} X_u$  is the expected number of neighbors of  $v$  that get colored; by Claim **8.3**, we get that

$$\mathbb{E}[X] \geq \sum_{u \sim v} \mathbb{E}[X_u] \geq d(v) \cdot a\theta.$$

Finally, observing that  $d'(v) = d(v) - X$  completes the proof.  $\blacksquare$

### 8.2.1 The Concentration Bound for Degrees

We want to show that  $X - \mathbb{E}[X]$  is small **whp** $_{\Delta}$ . Let  $Y := \sum_{u \sim v} Y_u$ , and  $Y' := \sum_{u \sim v} Y'_u$ . For the upper tail, observe that as  $Y - Y' \leq X \leq Y$ , we have

$$X - \mathbb{E}[X] \leq Y - \mathbb{E}[Y - Y'] = (Y - \mathbb{E}[Y]) + \mathbb{E}[Y']. \quad (19)$$

But  $Y$  is the sum of independent  $\{0, 1\}$ -valued r.v.s  $\{Y_u\}_{u \sim v}$ , and by inclusion-exclusion again each  $\mathbb{E}[Y_u] \leq \sum_{\gamma} \frac{\theta}{2} p(u, \gamma)$ . Thus

$$\mathbb{E}[Y] \leq \theta/2 \sum_{u \sim v} \sum_{\gamma} p_c(u, \gamma) \leq O(\theta d(v)) = O(\theta \Delta).$$

By the tail bound Theorem **A.1**, setting  $\lambda_D^{(1)} := O(\sqrt{\theta \Delta \ln \Delta} + \ln \Delta)$  suffices to give  $\Pr[|Y - \mathbb{E}[Y]| \leq \lambda_D^{(1)}] \leq 1/\text{poly}(\Delta)$ . Plugging this into **(19)** and using the bound of  $\mathbb{E}[Y'] = \mathbb{E}[\sum_{u \sim v} Y'_u] \leq 2K\theta^2 \Delta =: \lambda_D^{(2)}$  gives us that the total deviation of  $X$  above its mean is at most  $\lambda_D^{(1)} + \lambda_D^{(2)} = O(K\theta^2 \Delta)$  **whp** $_{\Delta}$ .

For the lower tail, observe that

$$\mathbb{E}[X] - X \leq \mathbb{E}[Y] - (Y - Y') = (\mathbb{E}[Y] - Y) + \mathbb{E}[Y'] + (Y' - \mathbb{E}[Y']). \quad (20)$$

Since  $\mathbb{E}[Y] - Y \leq \lambda_D^{(1)}$  and  $\mathbb{E}[Y'] \leq \lambda_D^{(2)}$  by the preceding argument, so we focus on bounding the upper tail of  $Y'$ . For this we use the concentration inequality for polynomials from Theorem **A.2**. The parameters are:

$$\mu_0 = \mathbb{E}[Y'] \leq d(v) \cdot 2K\theta^2 \leq 2K\theta^2 \Delta \quad (21)$$

$$\mu_1 \leq 2 \max_{(v, \gamma)} \sum_{u \sim v} \mathbb{E}[A_{(u, \gamma)}] = 2 \max_{(v, \gamma)} \sum_{u \sim v} \theta p_c(u, \gamma) \leq 2\theta \hat{p} \Delta = O(\Delta^{-3\varepsilon}) \leq 1 \quad (22)$$

$$\mu_2 = 2. \quad (23)$$

Plugging this into Corollary A.3 of the aforementioned concentration inequality, we get that **whp** $_{\Delta}$  the deviation  $|Y' - \mathbb{E}[Y']|$  is  $O(\sqrt{2K\theta^2\Delta \ln \Delta} + \ln^2 \Delta) =: \lambda_D^{(3)}$ . Substituting into (20), we have that **whp** $_{\Delta}$ ,

$$\mathbb{E}[X] - X \leq \lambda_D^{(1)} + \lambda_D^{(2)} + \lambda_D^{(3)} = O(K\theta^2\Delta) =: \lambda_D \quad (24)$$

This proves Lemma 8.2.

Finally, the LLL part: here the bad event  $\mathcal{B}_v^d = \{\omega : d'(v) > (1 - a\theta)d(v) + \lambda_D\}$ , and this depends on the  $A, \eta, M$  r.v.s for vertices at distance at most 2 from  $v$  (since it depends on whether  $u \in N(v)$  survive, which depend on their neighbors). Hence the dependency is at most  $\Delta^4 \times s$ .

### 8.3 The Entropy

Now to show invariant (InvH), that the entropy of  $\{p'(v, \gamma)\}_{\gamma}$  remains high, where the entropy is defined as

$$h(v, p') := - \sum_{\gamma} p'(v, \gamma) \ln p'(v, \gamma) \quad (25)$$

**Lemma 8.5**  $\mathbb{E}[h(v, p')] \geq h(v, p) - C\theta \xi(v, p_a)$ .

**Proof:** Recall that  $p'(v, \gamma) := p(v, \gamma) \cdot M_{(v, \gamma)}$  from (14), where  $M_{(v, \gamma)}$  is a m.o. r.v..

$$\begin{aligned} h(v, p') &= - \sum_{\gamma} p'(v, \gamma) \ln p'(v, \gamma) \\ &= - \sum_{\gamma} p(v, \gamma) M_{(v, \gamma)} \ln p(v, \gamma) - \sum_{\gamma} p(v, \gamma) M_{(v, \gamma)} \ln M_{(v, \gamma)} \end{aligned}$$

Hence, taking expectations,

$$\begin{aligned} \mathbb{E}[h(v, p')] &= - \sum_{\gamma} \mathbb{E}[M_{(v, \gamma)}] p(v, \gamma) \ln p(v, \gamma) - \sum_{\gamma} p(v, \gamma) \mathbb{E}[M_{(v, \gamma)} \ln M_{(v, \gamma)}] \\ &= h(v, p) - \sum_{\gamma} p(v, \gamma) \mathbb{E}[M_{(v, \gamma)} \ln M_{(v, \gamma)}] \end{aligned}$$

In fact, if the probability  $p(v, \gamma)$  is greater than  $\hat{p}$  for some  $\gamma$  (i.e., if  $p_a(v, \gamma) > 0$ ), then the definition of the stopped product implies that  $M_{(v, \gamma)} \equiv 1$ . Hence, we get the stronger claim that

$$\mathbb{E}[h(v, p')] = h(v, p) - \sum_{\gamma} p_a(v, \gamma) \mathbb{E}[M_{(v, \gamma)} \ln M_{(v, \gamma)}] \quad (26)$$

Now, recall that for an r.v.  $X$ , we defined  $\kappa(X) = \mathbb{E}[X \ln X]$  in §6.2.3.

**Claim 8.6**  $\kappa(M_{(v, \gamma)}) = \mathbb{E}[M_{(v, \gamma)} \ln M_{(v, \gamma)}] \leq \theta C \sum_{w \sim v} p_c(w, \gamma)$ .

**Proof:** Observe that  $M_{(v, \gamma)}$  is a stopped product of a bunch of m.o. r.v.s  $M_{(v, \gamma)}^w$  of neighbors  $w \sim v$ , each of which is either 1 (if  $(w, \gamma) \notin \mathcal{T}$ ) or itself a product of two m.o. r.v.s as in (13). Using properties of the  $\kappa(\cdot)$  function from §6.2.3, we get

$$\kappa(M_{(v, \gamma)}) \leq \sum_{w \sim v} (\theta p_c(w, \gamma)) \cdot \max \kappa(M_{v, \gamma}^w) \quad (27)$$

Using the definition of  $C$  gives us the claim. ■

Substituting Claim 8.6 into (26),

$$\mathbb{E}[h(v, p')] \geq h(v, p) - \theta C \sum_{\gamma} \sum_{w \sim v} p_a(v, \gamma) p_c(w, \gamma) \geq h(v, p) - \theta C \xi(v, p_a). \quad (28)$$

This proves the lemma. ■

**Lemma 8.7** *The deviation  $|h(v, p') - \mathbb{E}[h(v, p')]|$  is at most  $\sqrt{p^* \ln^3 \Delta}$  **whp** $_{\Delta}$ .*

**Proof:** By definition, the entropy  $h(v, p')$  is a sum of independent r.v.s  $p'(v, \gamma) \ln \frac{1}{p'(v, \gamma)}$ ; since  $p'(v, \gamma) \in [1/s, p^*]$ , these r.v.s are  $[0, m]$ -bounded where  $m := (p^* \ln s) \leq p^* \ln \Delta$ . Moreover, the mean  $\mu$  satisfies  $\mu \leq \ln \Delta$  because the entropy can be at most  $\ln s \leq \ln \Delta$ . Hence by the tail bound from Theorem A.1, the deviation from the mean is at most  $\lambda_H = O(\sqrt{\mu m \ln \Delta} + m \ln \Delta) = \sqrt{p^* \ln^3 \Delta}$  **whp** $_{\Delta}$ . ■

For the LLL application, since the entropy depends only on the  $p'(v, \gamma)$  values, the bad event for  $v$  is dependent only on the random choices of  $\{v\} \cup N(v)$ , and hence easily handled for the usual reasons.

## 8.4 The Energy

The calculations above show that the decrease in entropy at vertex  $v$  depends on the energy of edges incident to  $v$ , so it remains to show that this energy is small (and in fact, decreases over the course of the algorithm). This is technically the most interesting part of the analysis. We're interested in

$$\xi_G(v, p_a) = \sum_{u \sim_G v} \underbrace{\xi(uv, p_a)}_{\sum_{\gamma} p_a(u, \gamma) p_a(v, \gamma)}, \quad (29)$$

and want to show that this energy drops by a factor of  $\approx (1 - \theta/2)$  each time. Formally, we prove the following:

**Lemma 8.8** *For a vertex  $v$ , **whp** $_{\Delta}$  the energy*

$$\xi_{G'}(v, p'_a) \leq (1 - b\theta) \cdot \xi_G(v, p_a) + O(\theta \Delta^{-2\varepsilon} \ln \Delta).$$

As for the other invariants, we first bound the expectation and then show a large-deviations bound. The expectation calculation itself proceeds via two claims — the first claim quantifies the change in energy due to considering the new probability distribution  $p'_a$  instead of  $p_a$  (keeping the graph  $G$  fixed), and the second captures the change due to considering graph  $G'$  instead of  $G$  (but now keeping the distribution  $p'$  fixed). Remember that the energy of non-edges is zero by definition.

**Claim 8.9** *For any  $v$ ,  $\sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)] \leq \sum_{u \in G} \xi(uv, p_a)$ .*

**Claim 8.10** *For any  $v$ ,  $\mathbb{E}[\sum_{u \in G'} \xi(uv, p'_a)] \leq (1 - a\theta) \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)]$ .*

In the latter claim, observe that  $G'$  is a random variable itself, and hence we cannot just push the expectation inside the sum. Before we prove Claim 8.9, we will use the following observation which allows us to consider the unstopped product instead of the stopped product.

**Fact 8.11** For any  $u, v \in G$ , and  $\gamma \in L$ ,  $p'_a(u, \gamma)p'_a(v, \gamma) \leq p_a(u, \gamma)p_a(v, \gamma) \prod_{w \sim u} M_{(u, \gamma)}^w \prod_{w \sim v} M_{(v, \gamma)}^w$

**Proof:** Indeed, the products are stopped only when the terms multiplied give us value at least  $\hat{p}$ , but then corresponding  $p'_a$  values get zeroed out. ■

**Proof of Claim 8.9:** Using Fact 8.11, we can replace the stopped product by the usual ones:

$$\begin{aligned} \mathbb{E}[p'_a(u, \gamma)p'_a(v, \gamma)] &\leq p_a(u, \gamma)p_a(v, \gamma) \mathbb{E} \left[ \prod_{w \sim u} M_{(u, \gamma)}^w \prod_{x \sim v} M_{(v, \gamma)}^x \right] \\ &\leq p(u, \gamma)p(v, \gamma) \mathbb{E} \left[ \prod_{w \sim u: w \not\sim v} M_{(u, \gamma)}^w \right] \mathbb{E} \left[ \prod_{x \sim v: x \not\sim u} M_{(v, \gamma)}^x \right] \mathbb{E} \left[ \prod_{w: uvw \in \Delta} M_{(u, \gamma)}^w M_{(v, \gamma)}^w \right] \end{aligned}$$

But the first two expectations equal to one, since each of these is a product of independent m.o. r.v.s. For the third one, with probability  $(1 - \theta p_c(w, \gamma))$  the pair  $(w, \gamma) \notin \mathcal{T}$  and we get 1, else with probability  $\theta p_c(w, \gamma)$  the pair  $(w, \gamma) \in \mathcal{T}$  and the random color class modifier gives zero. In all cases the final expectation would be at most one. ■

**Proof of Claim 8.10:** In this claim, it suffices to bound, for each color  $\gamma$ ,  $\mathbb{E}[p'_a(u, \gamma)p'_a(v, \gamma)\mathbf{1}_{(u \in G')}]$ . The crucial observation is that for any  $\gamma$ , the events

$$\{u \in G'\} = \{\mathcal{C}_u = \emptyset\} \tag{30}$$

$$= \{\mathcal{C}_u \subseteq \{(u, \gamma)\}\} - \{\mathcal{C}_u = \{(u, \gamma)\}\} \tag{31}$$

$$= \{\mathcal{C} \cap (\{u\} \times (L \setminus \{\gamma\})) = \emptyset\} - \{\mathcal{C}_u = \{(u, \gamma)\}\} \tag{32}$$

However, when  $\mathcal{C}_u = \{(u, \gamma)\}$ , then  $A_{(u, \gamma)} = 1$  and  $\eta(u, \gamma) = 1$ , which means that  $p'_a(v, \gamma) = 0$  due to the modifier  $M_{(v, \gamma)}^u$ , and hence  $\mathbb{E}[p'_a(u, \gamma)p'_a(v, \gamma)\mathbf{1}_{(\mathcal{C}_u = \{(u, \gamma)\})}] = 0$ . Moreover, the former event says that  $u$  did not get any color from the set  $L \setminus \gamma$ , which is independent of all decisions for the color  $\gamma$ . Hence

$$\mathbb{E}[p'_a(u, \gamma)p'_a(v, \gamma)\mathbf{1}_{(u \in G')}] = \mathbb{E}[p'_a(u, \gamma)p'_a(v, \gamma)] \cdot \Pr[\mathcal{C} \cap (\{u\} \times (L \setminus \{\gamma\})) = \emptyset] \tag{33}$$

Using (32) again, we know that

$$\Pr[\mathcal{C} \cap (\{u\} \times (L \setminus \{\gamma\})) = \emptyset] = \Pr[u \in G'] + \Pr[\mathcal{C}_u = \{(u, \gamma)\}] \tag{34}$$

$$\leq \Pr[u \in G'] + \Pr[A_{(u, \gamma)} = 1] \tag{35}$$

$$\leq (1 - (a + \varepsilon)\theta) + \theta\hat{p}, \tag{36}$$

where the first expression is from Claim 8.3 and the second one from  $p_c(u, \gamma) \leq \hat{p}$ . But  $\theta\hat{p} \leq \varepsilon\theta$  for large  $\Delta$ . Thus the claim is proved. ■

Combining Claims 8.9 and 8.10, and using  $a = b$ ,

$$\mathbb{E}[\xi_{G'}(v, p'_a)] = \sum_{u \in G'} \mathbb{E}[\xi(uv, p'_a)] \leq (1 - b\theta) \sum_{u \in G} \xi(uv, p_a) = (1 - b\theta) \xi_G(v, p_a). \tag{37}$$

We next show that the r.v.  $\xi_{G'}(v, p'_a)$  is concentrated around its mean.

### 8.4.1 The Concentration Bound for Energy

Fix some  $v \in G$ . We want to show that  $\sum_{u \in G} \xi(uv, p'_a) \mathbf{1}_{(u \in G')}$  is concentrated around its mean. The idea is simple: denoting  $Q_u = \xi(uv, p')$  and  $R_u = \mathbf{1}_{(u \in G')}$ , and letting  $q_u = \mathbb{E}[Q_u]$ ,  $r_u = \mathbb{E}[R_u]$  being their expectations, the triangle inequality gives us that

$$\left| \sum_u Q_u R_u - \mathbb{E}[Q_u R_u] \right| \leq \left| \sum_u (Q_u - q_u) R_u \right| + \left| \sum_u q_u (R_u - r_u) \right| + \left| \sum_u q_u r_u - \mathbb{E}[Q_u R_u] \right|$$

The following three claims now bound the three expressions on the right.

**Claim 8.12**  $\left| \sum_{u \in G} \xi(uv, p'_a) \mathbf{1}_{(u \in G')} - \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)] \mathbf{1}_{(u \in G')} \right| \leq O(\theta \Delta^{-7\epsilon} \ln \Delta)$  **whp $_{\Delta}$** .

**Proof:** The left hand side is at most  $\sum_{u \in N_G(v)} |\xi(uv, p'_a) - \mathbb{E}[\xi(uv, p'_a)]|$ . Each  $\xi(uv, p'_a)$  is a sum of independent  $\widehat{p}^2$ -bounded r.v.s, and hence deviates from its mean by at most  $O(\sqrt{\mathbb{E}[\xi(uv, p'_a)] \widehat{p}^2 \ln \Delta} + \widehat{p}^2 \ln \Delta)$  **whp $_{\Delta}$**  by the tail bound in Theorem A.1. Summing this over all  $u \in N_G(v)$  (and taking a union bound over these  $\Delta$  events), we get the total deviation **whp $_{\Delta}$**  is

$$\begin{aligned} & \sum_{u \in N_G(v)} O(\widehat{p} \sqrt{\mathbb{E}[\xi(uv, p'_a)] \ln \Delta} + \widehat{p}^2 \ln \Delta) \\ & \leq O\left(\widehat{p} \sqrt{\ln \Delta} \cdot \sqrt{\Delta} \sqrt{\mathbb{E}\left[\sum_{u \in N_G(v)} \xi(uv, p')\right]}\right) + O(\widehat{p}^2 \Delta \ln \Delta) \\ & \leq O(\widehat{p} \sqrt{\Delta K \ln \Delta} + \widehat{p}^2 \Delta \ln \Delta). \end{aligned}$$

The first inequality uses Cauchy-Schwarz; the next one uses the expectation bound from Claim 8.9 and invariant (InvE). The dominant term is  $O(\widehat{p} \sqrt{\Delta K \ln \Delta}) = O(\theta \Delta^{-7\epsilon} \ln \Delta)$ , which completes the claim.  $\blacksquare$

**Claim 8.13**  $\left| \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)] \mathbf{1}_{(u \in G')} - \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)] \mathbb{E}[\mathbf{1}_{(u \in G')}] \right| \leq O(\theta \Delta^{-2\epsilon} \ln \Delta)$  **whp $_{\Delta}$** .

**Proof:** By Claim 8.9, each term

$$\mathbb{E}[\xi(uv, p'_a)] \leq \xi(uv, p_a) = \sum_{\gamma} p_a(u, \gamma) p_a(v, \gamma) \leq \widehat{p} \sum_{\gamma} p_a(v, \gamma) \leq 2\widehat{p}. \quad (38)$$

Using  $c_u := \mathbb{E}[\xi(uv, p'_a)]$  and observing that  $c_u = 0$  for all  $u \notin N_G(v)$ , we want to bound the deviation

$$\left| \sum_{u \in N_G(v)} c_u (\mathbf{1}_{(u \in G')} - \Pr[u \in G']) \right|.$$

The argument now follows that of §8.2.1, with the only difference that the variables are in  $[0, 2\widehat{p}]$  rather than  $\{0, 1\}$ , which merely multiplies the deviation from the mean by a factor of  $2\widehat{p}$ . Hence, **whp $_{\Delta}$** , we have

$$\left| \sum_{u \in N_G(v)} c_u \mathbf{1}_{(u \in G')} - \sum_{u \in N_G(v)} c_u \Pr[u \in G'] \right| \leq O(\theta^2 K \widehat{p} \Delta) = O(\theta \Delta^{-2\epsilon} \ln \Delta).$$

This proves the claim.  $\blacksquare$

**Claim 8.14**  $\left| \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)] \mathbb{E}[\mathbf{1}_{(u \in G')}] - \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a) \mathbf{1}_{(u \in G')}] \right| \leq O(\theta \Delta^{-1/2})$  **whp $_{\Delta}$** .



**Proof:** Using (33) and (35), summing over all colors  $\gamma$ , and using the fact that  $\Pr[A_{(u,\gamma)} = 1] \leq \theta\hat{p}$  for all  $\gamma$ ,

$$\mathbb{E}[\xi(uv, p'_a)]\mathbb{E}[\mathbf{1}_{(u \in G')}] \leq \mathbb{E}[\xi(uv, p'_a)\mathbf{1}_{(u \in G')}] \leq \mathbb{E}[\xi(uv, p'_a)] (\mathbb{E}[\mathbf{1}_{(u \in G')}] + \theta\hat{p}).$$

Rearranging, summing over all  $u \in G$ , and using that  $\mathbb{E}[\xi(uv, p'_a)] \leq 2\hat{p}$  by the calculation in (38),

$$0 \leq \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)\mathbf{1}_{(u \in G')}] - \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)]\mathbb{E}[\mathbf{1}_{(u \in G')}] \leq \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)]\theta\hat{p} \quad (39)$$

$$\leq \Delta \cdot 2\theta \cdot \hat{p}^2. \quad (40)$$

This is at most  $O(\theta\Delta^{-1/2})$ , which proves the claim.  $\blacksquare$

Putting Claims 8.12–8.14 together, we get that **whp** $_{\Delta}$

$$\sum_{u \in G} \xi(uv, p'_a)\mathbf{1}_{(u \in G')} \leq \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)\mathbf{1}_{(u \in G')}] + O(\theta\Delta^{-2\varepsilon} \ln \Delta) \quad (41)$$

$$\leq (1 - b\theta) \cdot \xi(uv, p_a) + O(\theta\Delta^{-2\varepsilon} \ln \Delta) \quad (42)$$

where the latter inequality follows from Claims 8.9 and 8.10.

Finally, for the LLL, the bad events in this case are again dependent only on the random choices within distance 2 of  $v$ , which means the dependency is  $O(\Delta^4 s)$ .

## 8.5 Behavior after $T$ rounds, and Maintaining Invariants

In the previous sections, we showed that if the invariants **(InvP)**–**(InvH)** held at the beginning of a specific round, then using the LLL we can ensure that they hold at the end of the lemma, with some additional loss. Using this, we now show that the bounds we have derived suffice for the invariants hold at each round in  $[1..T]$ . For all these bounds we use that  $\Delta$  is large enough (15).

- **Probability.** After each round, the probability value may increase by  $\lambda_P$ . This means after  $T$  rounds,

$$p^t(\mathcal{P}_v) \leq 1 \pm T\lambda_P = 1 \pm \frac{2\varepsilon \ln(\Delta)}{a\theta} \ln \Delta \cdot \theta\Delta^{-1/8-4\varepsilon} \sqrt{\ln \Delta} = 1 \pm \sqrt{\theta}.$$

- **Degree.** In each round, the degree falls by a multiplicative factor of  $(1 - a\theta)$ , but potentially increases by an additive term of  $\lambda_D$ . This means that after  $t$  rounds, the degree can be (wastefully) bounded by

$$d(v, G^t) \leq (1 - a\theta)^t \cdot \Delta + t\lambda_D \leq e^{-a\theta t} \Delta + O(K\theta^2 \Delta t)$$

Now for  $t = T$ , we get that

$$d(v, G^t) \leq \Delta e^{-2\varepsilon \ln \Delta} + O(K\theta\Delta) \leq \Delta^{1-\varepsilon}. \quad (43)$$

- **Energy.** Again the energy falls by a  $(1 - b\theta)$  factor but potentially increases by an additive term of  $\lambda_E$ . This time we use a slightly better bound<sup>6</sup> of

$$\xi(v, p_a^t) \leq (1 - b\theta)^t \xi(v, p_a^0) + \frac{\lambda_E}{b\theta} \quad (44)$$

$$\leq e^{-b\theta t} K + O(\Delta^{-2\varepsilon} \ln \Delta) \leq 2K. \quad (45)$$

---

<sup>6</sup>Given a system  $x_{t+1} \leq \alpha x_t + \beta$  for  $\alpha \leq 1$ , we know that  $x_t \leq \alpha^t x_0 + \frac{\beta}{1-\alpha}$ .

- **Entropy.** From Lemmas 8.5 and 8.7 we get that

$$\begin{aligned}
h(v, p^t) &\geq h(v, p^0) - \theta C \sum_{t' \leq t} \xi(v, p_a^{t'}) - t\lambda_H \\
&\geq (\ln \Delta - \ln K) - \theta C \sum_{t' \leq t} \left( (1 - b\theta)^{t'} \xi(v, p_a^0) + \frac{\lambda_E}{b\theta} \right) - t\sqrt{p^* \ln^3 \Delta} \\
&\geq (\ln \Delta - \ln K) - \theta C \left( \frac{\xi(v, p_a^0)}{b\theta} + \frac{t\lambda_E}{b\theta} \right) - t\sqrt{p^* \ln^3 \Delta} \\
&\geq (\ln \Delta - \ln K) - \frac{KC}{b} - \frac{tC\theta \cdot O(\Delta^{-2\varepsilon} \ln \Delta)}{b} - t\sqrt{p^* \ln^3 \Delta}
\end{aligned}$$

For  $t \leq T = \frac{\varepsilon}{a\theta} \ln \Delta$ , we get

$$\begin{aligned}
h(v, p^t) &\geq (\ln \Delta - \ln K) - \frac{KC}{b} - \frac{2\varepsilon \ln(\Delta)}{a \cdot b} O(\Delta^{-2\varepsilon} \ln^2 \Delta) - O(\Delta^{-1/8} \ln^{2.5} \Delta) \\
&\geq (1 - \varepsilon) \ln \Delta.
\end{aligned}$$

By (43), the degree of all surviving vertices falls below  $\Delta^{1-\varepsilon}$  after  $T$  rounds, whence we can color them using  $\Delta^{1-\varepsilon}$  more colors. Hence the total number of colors used is  $s + \Delta^{1-\varepsilon} = O(\frac{\Delta}{K}) = O(\frac{\Delta \cdot C}{\ln \Delta})$ . It remains to bound the parameter  $C$ : since the m.o. r.v.s  $M_{(v,\gamma)}^w$  either take on value 0 or  $2r$ , we know that  $\kappa(M_{(v,\gamma)}^w) = \ln(2r) = \ln 2 + \ln r$ . This means the number of colors used is

$$O\left(\frac{\Delta}{\ln \Delta} \ln r\right)$$

This completes the proof of Theorem 6.1.

## 9 $K_r$ -free Graphs

The above analysis was tailored for graphs where each neighborhood is  $r$ -colorable. To color  $K_r$ -free graphs, we use different modifiers which give a weaker guarantee of  $O(\frac{\Delta}{\ln \Delta} (r^2 + r \ln \ln \Delta))$  colors as claimed in Theorem 6.2. Since a coloring algorithm with  $s$  colors gives an independent set of size  $n/s$ , this result matches Shearer's bound for independent sets for values of  $r \in O(\ln \ln \Delta)$ .

In this section, we will assume that  $r \leq c' \sqrt{\log \Delta}$  for some suitably small constant  $c'$ ; for values of  $r \geq c' \sqrt{\log \Delta}$ , the quantity  $O(r^2 + r \ln \ln \Delta)$  is  $\Omega(\ln \Delta)$ , and the trivial  $\Delta$ -coloring satisfies Theorem 6.2. Again, we assume that  $\Delta$  is a suitably large constant.

### 9.1 The Algorithm

The algorithm in this case is very similar in structure to that in §6.3. The only difference is in the modifiers: we replace Step (5) of that algorithm by the following:

5'. For each pair  $(w, \gamma)$ , generate modifiers  $M_{(v,\gamma)}^w$  for all  $v \sim w$  as follows:

- If  $(w, \gamma) \notin \mathcal{T}$  (i.e.,  $A_{(w,\gamma)} = 0$ ) then  $M_{(v,\gamma)}^w = 1$  for all  $v \in N(w)$ .

- Else, if  $(w, \gamma) \in \mathcal{T}$ , then let  $H = G[N(w)]$  be the graph induced on the neighbors of  $w$ . Use Theorem 9.2 on this graph  $H$ , with  $c = \frac{1}{4}$ , and with values  $\{p(v, \gamma)\}_{v \in V(H)}$ . This generates modifiers  $\widetilde{M}_{(v, \gamma)}^w$  for each  $v \in V(H) = N(w)$ , and then define

$$M_{(v, \gamma)}^w = \underbrace{2(1 - \eta(w, \gamma))}_{m.o.r.v.} \cdot \underbrace{\widetilde{M}_{(v, \gamma)}^w}_{m.o.r.v.}$$

We emphasize that we invoke the procedure in Theorem 9.2 once for each  $(w, \gamma) \in \mathcal{T}$ . A few comments on the new modifiers:

- By Theorem 9.2(P3) and the fact that  $c = \frac{1}{4}$ , each modifier has

$$\kappa(M_{(v, \gamma)}^w) \leq O(r^2 + r \ln(\sum_{v \sim w} p(v, \gamma)))$$

as long as  $(w, \gamma)$  is a tentatively chosen pair;  $\kappa(M_{(v, \gamma)}^w) = 0$  if  $(w, \gamma)$  is not tentatively chosen. But the pair  $(w, \gamma)$  is tentatively chosen with probability  $\theta p_c(w, \gamma)$ , and  $p_c(w, \gamma) \neq 0$  implies that the probability is at most  $100 \ln \Delta$ . This implies that  $\kappa(M_{(v, \gamma)}^w) \leq O(r^2 + r \ln \ln \Delta)$ .

- The maximum value of the modifier is given by Theorem 9.2(P2), and again using the bound on total probability of any neighborhood, we get that

$$\widehat{\kappa}(M_{(v, \gamma)}^w) \leq O(\ln \Delta)^r \cdot O(1)^{r^2}. \quad (46)$$

If we define  $\widehat{\kappa} := \max_{w, v, g} \widehat{\kappa}(M_{(v, \gamma)}^w)$ , we can use this to again define  $p^* := 2\widehat{\kappa}\widehat{p}$ , just the value of  $\widehat{\kappa}$  has changed. In the analysis for the  $r$ -local-colorability case, we used that  $p^*$  was only greater than  $\widehat{p}$  by a  $\widehat{\kappa} = 2r \leq \Delta^\varepsilon$  factor. This time  $\widehat{\kappa}$  is given by (46), but for values  $r \leq c'' \sqrt{\ln \Delta}$  for some suitably small constant  $c''$ ,  $\widehat{\kappa}$  is again  $\Delta^\varepsilon$  and we can use  $p^* \leq \widehat{p}\Delta^\varepsilon$  as before.

The runtime: Constructing the modifier requires  $\text{poly}(\Delta) \cdot r$  time and we construct a modifier for each vertex at most  $T$  times. Hence, this algorithm runs in  $O(n \cdot r \cdot \text{poly}(\Delta))$

## 9.2 The Altered Parameters

The analysis remains very similar, we just indicate the changes (in blue). The parameters change slightly from §7.1: we now set  $c := \frac{1}{4}$  (as mentioned in the algorithm description), and set  $b := a - c - \varepsilon$ . This affects the value of  $K$ . Moreover, since the modifiers  $M_{(v, \gamma)}^w$  change, the value of  $C = \max_{w, v, g} \kappa(M^w(v, \gamma))$  becomes  $O(r^2 + r \ln \ln \Delta)$  by the discussion above. To summarize, here are the new parameters; those in blue differ from their counterparts in §7.1.

$$\begin{array}{lll} \varepsilon := 1/100 & \theta := \Delta^{-1/4+2\varepsilon} & s := |L| := \frac{\Delta}{K} \\ \widehat{\kappa} := 2^{O(r^2)} \cdot \ln(\Delta)^{O(r)} & \widehat{p} := \Delta^{-3/4-5\varepsilon} & p^* := \widehat{\kappa}\widehat{p} \\ a := 1/2 - 3\varepsilon & \widehat{c} := 1/4 & b := a - \widehat{c} - \varepsilon \\ C := O(r^2 + r \ln \ln \Delta) & T := \frac{2\varepsilon}{a \cdot \theta} \ln \Delta & K := \frac{(b/4)\varepsilon \ln \Delta}{C} \end{array}$$

### 9.3 The Analysis

Upon closer inspection, the proofs of invariants **(InvP)**, **(InvD)** and **(InvH)** go through verbatim, since they do not use any properties of the modifiers being used. Only the analysis of the invariant **(InvE)** bounding the energy needs to be changed. In particular, Claim 8.9 no longer holds, and we must use the slightly weaker claim below (we will defer the proof to the end of the section).

**Claim 9.1** For any  $v$ ,  $\mathbb{E}[\xi_G(v, p'_a)] = \sum_{u \in G} \mathbb{E}[\xi(uv, p'_a)] \leq (1 + (c + \varepsilon)\theta) \xi_G(v, p_a)$ .

Combining Claims 9.1 and 8.10,

$$\sum_{u \in G'} \mathbb{E}[\xi(uv, p'_a)] \leq (1 - (a - c - \varepsilon)\theta) \sum_{u \in G} \xi(uv, p_a) = (1 - b\theta) \sum_{u \in G} \xi(uv, p_a),$$

since we redefined  $b$  to be  $a - c - \varepsilon$ . This is the analog of (37); we now need to show the concentration. And indeed, the argument in §8.4.1 is almost independent of the modifiers, except for the use of Claim 8.9 in the proof of Claim 8.12; we can now use Claim 9.1 instead and get identical results up to changes in the constants (which are absorbed in the **whp** $_{\Delta}$  claims). This proves that invariant **(InvE)** also holds. The rest of the analysis is unchanged for the new modifiers. Finally, the number of colors used is  $O(\frac{\Delta}{K}) = O(\frac{\Delta \cdot C}{\ln \Delta})$  again; plugging in the value of  $C$  calculated above gives Theorem 6.2.

It only remains to prove Claim 9.1, which we do next, and to give the construction of the new modifiers, which appears in the next section §9.4.

**Proof of Claim 9.1:** We begin as in the proof of Claim 8.9. Using Fact 8.11, we can replace the stopped product with unstopped products, i.e. for any nodes  $u, v \in G$ , and color  $\gamma \in L$

$$\mathbb{E}[p'_a(u, \gamma)p'_a(v, \gamma)] \leq p_a(u, \gamma)p_a(v, \gamma)\mathbb{E}\left[\prod_{w \sim u} M_{(u, \gamma)}^w \prod_{x \sim v} M_{(v, \gamma)}^x\right].$$

Using independence of the random variables,

$$\mathbb{E}\left[\left(\prod_{w \sim u} M_{(u, \gamma)}^w\right)\left(\prod_{w \sim v} M_{(v, \gamma)}^w\right)\right] = \prod_{w \sim v: w \not\sim u} \mathbb{E}\left[M_{(u, \gamma)}^w\right] \prod_{w \sim v: w \not\sim u} \mathbb{E}\left[M_{(v, \gamma)}^w\right] \prod_{w: uvw \in \Delta} \mathbb{E}\left[M_{(u, \gamma)}^w M_{(v, \gamma)}^w\right]. \quad (47)$$

The expectations in the first two products are 1, so focus on the expectations in the last product:

$$\begin{aligned} & \mathbb{E}\left[M_{(u, \gamma)}^w M_{(v, \gamma)}^w\right] \\ &= \mathbb{E}\left[M_{(u, \gamma)}^w M_{(v, \gamma)}^w \mid w \notin \mathcal{T}_v\right] \cdot \Pr[w \notin \mathcal{T}_v] + \mathbb{E}\left[M_{(u, \gamma)}^w M_{(v, \gamma)}^w \mid w \in \mathcal{T}_v\right] \cdot \Pr[w \in \mathcal{T}_v] \end{aligned}$$

When  $w \notin \mathcal{T}_v$ , then all modifiers  $\{M_{(v, \gamma)}^w\}_v$  have value 1. On the other hand, if  $w \in \mathcal{T}_v$ , then either  $\eta_{(w, \gamma)} = 1$  (w.p. half) and  $M_{(v, \gamma)}^w M_{(u, \gamma)}^w = 0$ , or else  $\eta_{(w, \gamma)} = 0$  (also w.p. half) and  $M_{(v, \gamma)}^w M_{(u, \gamma)}^w = 4\widetilde{M}_{(v, \gamma)}^w \widetilde{M}_{(u, \gamma)}^w$ . Moreover,  $\Pr[w \in \mathcal{T}_v] = \theta p_c(w, \gamma) \leq \theta p_a(w, \gamma)$ . Plugging this into the expression above,

$$\begin{aligned} \mathbb{E}\left[M_{(u, \gamma)}^w M_{(v, \gamma)}^w\right] &= 1 \cdot (1 - \theta p_c(w, \gamma)) + \frac{1}{2} \cdot \mathbb{E}\left[4\widetilde{M}_{(v, \gamma)}^w \widetilde{M}_{(u, \gamma)}^w\right] \cdot \theta p_c(w, \gamma) \\ &\leq 1 + 2 \mathbb{E}\left[\widetilde{M}_{(v, \gamma)}^w \widetilde{M}_{(u, \gamma)}^w\right] \cdot \theta p_a(w, \gamma) \end{aligned}$$

Taking the product over all  $w$  that are common neighbors of  $u, v$ ,

$$\begin{aligned} & \prod_{w:uvw \in \Delta} \left( 1 + 2 \mathbb{E} \left[ \widetilde{M}_{(v,\gamma)}^w \widetilde{M}_{(u,\gamma)}^w \right] \cdot \theta p_a(w, \gamma) \right) \\ &= 1 + \sum_{w:uvw \in \Delta} 2\theta p_a(w, \gamma) \cdot \mathbb{E}[\widetilde{M}_{(u,\gamma)}^w \widetilde{M}_{(v,\gamma)}^w] + o(\Delta^{-2-\varepsilon}) \\ &\leq (1 + \varepsilon\theta) + \sum_{w:uvw \in \Delta} 2\theta p_a(w, \gamma) \cdot \mathbb{E}[\widetilde{M}_{(u,\gamma)}^w \widetilde{M}_{(v,\gamma)}^w] \end{aligned}$$

In the second expression, the second-order terms in the product expansion are bounded by  $(\theta \cdot \widehat{p} \cdot \Delta^\varepsilon)^2 \in o(\Delta^{-2-\varepsilon})$ , since  $p_a(v, \gamma) \leq \widehat{p}$  and  $\widetilde{M}_{(u,\gamma)}^w \leq \widehat{\kappa} \leq \Delta^\varepsilon$  (by the comments at the end of §9.1). The subsequent inequality uses that  $o(\Delta^{-2-\varepsilon}) \leq \varepsilon\theta$  for large enough  $\Delta$ .

Now summing over all colors  $\gamma$  and over all  $u \sim v$ , we get that for vertex  $v$ ,

$$\begin{aligned} \mathbb{E} \left[ \sum_{u \sim v} \xi(uv, p'_a) \right] &\leq \sum_{u \sim v} \sum_{\gamma} p_a(u, \gamma) p_a(v, \gamma) \left[ (1 + \varepsilon\theta) + \sum_{w:uvw \in \Delta} 2\theta p_a(w, \gamma) \mathbb{E} \left[ \widetilde{M}_{(v,\gamma)}^w \widetilde{M}_{(u,\gamma)}^w \right] \right] \\ &\leq (1 + \varepsilon\theta) \xi_G(v, p_a) + \sum_{\gamma} \sum_{u \sim v} \sum_{w:uvw \in \Delta} 2\theta p_a(u, \gamma) p_a(v, \gamma) p_a(w, \gamma) \mathbb{E}[\widetilde{M}_{(u,\gamma)}^w \widetilde{M}_{(v,\gamma)}^w] \end{aligned}$$

And now summing over all  $w \sim v$  instead of only  $w : uvw \in \Delta$ , and interchanging the summations

$$\leq (1 + \varepsilon\theta) \xi_G(v, p_a) + \sum_{\gamma} \sum_{w \sim v} \theta p_a(w, \gamma) \left( \sum_{u \sim v} p_a(v, \gamma) p_a(u, \gamma) \cdot 2 \mathbb{E}[\widetilde{M}_{(u,\gamma)}^w \widetilde{M}_{(v,\gamma)}^w] \right)$$

Applying Theorem 9.2(P1) on the inner summands,

$$\begin{aligned} &\leq (1 + \varepsilon\theta) \xi_G(v, p_a) + \sum_{\gamma} \sum_{w \sim v} \theta p_a(w, \gamma) \cdot c p_a(v, \gamma) \\ &\leq (1 + (c + \varepsilon)\theta) \xi_G(v, p_a) \end{aligned}$$

This completes the proof of Claim 9.1. ■

## 9.4 Constructing a $K_r$ -free Graph Modifier

**Theorem 9.2** *Given an integer  $t \leq r$ , a  $K_t$ -free graph  $H$ , a constant  $c < 1$  called the “contraction” parameter, values  $p : V(H) \rightarrow [0, 1]$ , we can construct modifiers  $\{M(v)\}_{v \in V}$  which are m.o. r.v.s that satisfy the following properties:*

- (P1) For every vertex  $v \in V(H)$ ,  $\sum_{u \sim v} 2p(u)p(v) \mathbb{E}[M(u)M(v)] \leq c \cdot p(v)$ .
- (P2) The maximum value of any  $M(v) \leq \left( \frac{8p(V)}{c} \right)^{t-2} \cdot 16^{\binom{t-2}{2}}$ .
- (P3)  $\mathbb{E}[M(v) \log M(v)] = O(t^2 + t \log(p(V)/c))$ .

Moreover, this construction works in time  $\text{poly}(|V(H)|) \cdot r$ .

Remark: The graph  $H$  should be viewed as the neighborhood of some vertex  $w$ .

**Proof:** The proof is via induction. The base case is when  $t = 2$ ; the graph  $H$  is  $K_2$ -free (i.e., it has no edges), then we return  $M_2(v) = 1$  for all  $v$ . This “trivial modifier” satisfies the properties above. Else  $t \geq 3$ ; in this case we recursively build the modifier. Let  $V = V(H)$  be the vertex set of the graph.

1. Define  $q(v) = p(v)/c$ . Sample a set  $X \subseteq V$  of vertices using a dependent sampling technique of Gandhi et al. [GKPS06] satisfying the following properties:

- (i)  $|X| \in \{\lfloor q(V) \rfloor, \lceil q(V) \rceil\}$ ,
- (ii)  $\Pr[v \in X] = q(v)$ , and
- (iii)  $\Pr[N(v) \cap X] \leq \prod_{u \sim v} (1 - q(u))$ .

Assumption 9.3 implies that  $q(v) \leq 1$  for all  $v$ , so the step is well-defined.

2. Let  $m := |X|$  and let  $X = x_1, x_2, \dots, x_m$ . Define  $V_0 := V - N(X)$  and let  $V_i := N(x_i) - (\cup_{j=0}^{i-1} V_j)$ . This partitions the vertex set  $V$  into  $m + 1$  sets. For  $i \in \{1, \dots, m\}$  the induced graphs  $H[V_i]$  are  $K_{t-1}$ -free; however,  $H[V_0]$  might still contain a  $K_{t-1}$ .
3. Let  $A := \{i \mid p(V_i) > c\}$  and  $B := [1 \dots m] \setminus A$ . Let  $\alpha$  be an r.v. with  $\Pr[\alpha = i] = w_i$ , where

$$w_i = \begin{cases} \frac{3}{4} & \text{if } i = 0 \\ \frac{p(V_i)}{8 \sum_{i \in A} p(V_i)} & \text{if } i \in A \\ \frac{1}{8|B|} & \text{if } i \in B \end{cases}$$

4. For each  $i \in \{1, \dots, m\}$ , recursively construct a modifier  $M'_i$  on the induced  $K_{t-1}$ -free graph  $H_i := H[V_i]$  using contraction parameter  $c' = \frac{1}{2}c$  and values  $p'(v) = p(v)/w_i$  for all  $v \in V_i$ . (Assumption 9.3(ii) implies that the new values  $p'$  satisfy the requirements of the Theorem.) Define  $M'_0$  to be the trivial modifier assigning value 1 to all  $v \in V_0$ . Define

$$M(v) = \sum_{i=0}^m \mathbf{1}_{(\alpha=i)} \cdot \mathbf{1}_{(v \in V_i)} \cdot \frac{1}{w_i} M'_i(v).$$

In other words, the modifier picks  $\alpha$ , defines  $M(v) = 0$  for all  $v \in V \setminus V_\alpha$ , and returns a scaled-up version of the recursively constructed modifier for  $V_\alpha$ .

In order to ensure the algorithm is well-defined, we need some assumptions

**Assumption 9.3** *The construction above satisfies:*

- (i)  $q(v) = p(v)/c \leq 1$  for all  $v \in V$ .
- (ii)  $p'(v) = p(v)/w_i \leq 1$  for all  $v \in V$ .

We now prove that  $M$  satisfies properties (P1)–(P3).

### 9.4.1 Satisfying property (P1)

We need to show that for  $v \in V$ , we have

$$\sum_{u \sim v} 2p(u)p(v)\mathbb{E}[M(u)M(v)] \leq c \cdot p(v).$$

Note that the expectation is over the random choice of  $X$ , the choice of  $\alpha$ , and the internal randomness of the modifiers (denoted as IR).

$$\begin{aligned} & \sum_{u \sim v} 2p(u)p(v)\mathbb{E}_{X,\alpha,IR}[M(u)M(v)] & (48) \\ &= \sum_{u \sim v} 2p(u)p(v)\mathbb{E}_{X,\alpha,IR}[M(u)M(v) \mid v \in N(X)] \Pr_X[v \in N(X)] \\ & \quad + \sum_{u \sim v} 2p(u)p(v)\mathbb{E}_{X,\alpha,IR}[M(u)M(v) \mid v \notin N(X)] \Pr_X[v \notin N(X)] \\ &\leq \sum_{u \sim v} 2p(u)p(v)\mathbb{E}_{X,\alpha,IR}[M(u)M(v) \mid v \in N(X)] \\ & \quad + \sum_{u \sim v} 2p(u)p(v)\mathbb{E}_{X,\alpha,IR}[M(u)M(v) \mid v \notin N(X)] \Pr_X[v \notin N(X)] & (49) \end{aligned}$$

Let us concentrate on the first summand in (49), and condition on some  $X$  such that  $v \in N(X)$ ;

$$\begin{aligned} & \sum_{u \sim v} 2p(u)p(v)\mathbb{E}_{\alpha,IR}[M(u)M(v) \mid v \in N(X), X] \\ &= \sum_{u \sim v} 2p(u)p(v)\mathbb{E}_{\alpha,IR} \left[ \sum_{i=1}^m \mathbf{1}_{(\alpha=i)} \cdot \mathbf{1}_{(u \in V_i)} \frac{1}{w_i} \cdot M'_i(u) \cdot \mathbf{1}_{(\alpha=i)} \cdot \mathbf{1}_{(v \in V_i)} \frac{1}{w_i} \cdot M'_i(v) \mid v \in N(X), X \right] \end{aligned}$$

Since the internal randomness for the modifiers at the next level and  $\alpha$  are independent, we get

$$= \sum_{u \sim v} 2p(u)p(v) \sum_{i=1}^m \mathbb{E}_{\alpha} [\mathbf{1}_{(\alpha=i)} \mid v \in N(X), X] \mathbb{E}_{IR} \left[ \frac{1}{w_i} \cdot M'_i(u) \cdot \frac{1}{w_i} \cdot M'_i(v) \mid v \in N(X), X \right]$$

Rearranging the sum, and noting that  $\mathbb{E}_{\alpha}[\mathbf{1}_{(\alpha=i)} \mid X] = \Pr[\alpha = i \mid X] = w_i$ , we get

$$\begin{aligned} &= \sum_{i=1}^m w_i \sum_{u \sim v, u \in V_i} 2 \frac{p(u)}{w_i} \frac{p(v)}{w_i} \mathbb{E}_{IR} [M'_i(u)M'_i(v)] \\ &= \sum_{i=1}^m w_i \sum_{u \sim v, u \in V_i} 2p'(u)p'(v) \mathbb{E}_{IR} [M'_i(u)M'_i(v)] \end{aligned}$$

Applying the induction hypothesis on  $H[V_i]$  with values  $p'$ ,

$$\leq \sum_{i=1}^m w_i c' p'(v) \leq \frac{c}{2} p(v)$$



We now turn to the second summand in (49). In particular, consider the expectation

$$\begin{aligned} & \mathbb{E}_{X,\alpha,IR} [M(u)M(v) \mid v \notin N(X)] \\ &= \Pr[\alpha = 0] \cdot \mathbb{E}_{X,IR} [M(u)M(v) \mid v \notin N(X), \alpha = 0] \\ & \quad + \Pr[\alpha \neq 0] \cdot \mathbb{E}_{X,\alpha,IR} [M(u)M(v) \mid v \notin N(X), \alpha \neq 0] \end{aligned}$$

If we ensure that the value  $w_0$  is chosen independently of  $X$ , we get that  $\{\alpha = 0\}$  is independent of  $X, IR$ . Moreover, since  $v \notin N(X)$ , it lies in  $V_0$ . By construction,  $M(u)M(v)$  will be non-zero only if  $u$  also lies in  $V_0$ , which causes the second summand above to disappear, and give

$$\begin{aligned} &= \Pr[\alpha = 0] \cdot \mathbb{E}_{X,IR} [M(u)M(v) \mid v \notin N(X)] \\ &= \Pr[\alpha = 0] \mathbb{E}_X \left[ \frac{1}{w_0} \mathbf{1}(u \notin N(X)) \frac{1}{w_0} \mathbf{1}(v \notin N(X)) \mid v \notin N(X) \right] \\ &= \Pr[\alpha = 0] \frac{1}{w_0^2} \Pr_X [u \notin N(X) \mid v \notin N(X)] \\ &\leq \Pr[\alpha = 0] \frac{1}{w_0^2} = \frac{1}{w_0} \end{aligned}$$

This shows that the second summand of (49) is upper bounded by

$$\begin{aligned} \sum_{u \sim v} p(u)p(v) \frac{1}{w_0} \Pr[v \notin N(X)] &\leq \frac{p(v)}{w_0} \left( \sum_{u \sim v} p(u) \right) \left( \prod_{u \sim v} (1 - q(u)) \right) \\ &\leq \frac{p(v)}{w_0} \cdot c \cdot \left( \sum_{u \sim v} q(u) \right) \exp \left( - \sum_{u \sim v} q(u) \right) \\ &\leq \frac{p(v)}{w_0} \cdot c \cdot \frac{1}{e} \leq \frac{c}{2} p(v). \end{aligned}$$

In the first inequality we used the negative correlation property of the dependent sampling scheme of Gandhi et al., in the second we used the definition of  $q(u) = p(u)/c$ , in the third inequality we used that  $x \exp(-x) \leq \frac{1}{e}$  for all  $x$ . The final inequality uses  $w_0 = 3/4 \geq 2/e$ . Hence the two summands sum up to at most  $cp(v)$ , proving Property (P1).

#### 9.4.2 Satisfying Properties (P2)

**Claim 9.4** *The new values satisfy the following:*

- (i)  $w_i \geq \frac{c}{8p(V)}$  for all  $i \in \{1, \dots, m\}$ .
- (ii)  $p'(V_i) \leq 8p(V)$ .

**Proof:** For (i), for  $i \in A$ ,  $w_i = \frac{p(V_i)}{8 \sum_{i \in A} p(V_i)} \geq \frac{c}{8p(V)}$  by the definition of  $A$ . Moreover, for  $i \in B$ ,  $w_i \geq \frac{1}{8|X|} \geq \frac{c}{8p(V)}$ . (Here we ignore the issues caused by  $|X|$  being an integer adjacent to  $p(V)$  rather than being equal to it.) Note that (i) does not make any claims about  $w_0$ .

For (ii),  $p'(V_0) = \frac{4}{3}p(V_0) \leq 8p(V)$ . For  $i \in A$ ,

$$p'(V_i) = \frac{p(V_i)}{w_i} = 8 \sum_{i \in A} p(V_i) \leq 8p(V).$$

For  $i \in B$ ,  $p'(V_i) = \frac{p(V_i)}{w_i} \leq \frac{c}{w_i} \leq 8p(V)$  by part (i). ■

To prove property (P2), by the IH the maximum value of the recursively constructed modifier  $M_i$  is

$$\left(\frac{8 \cdot p'(V_i)}{c'}\right)^{t-3} \cdot 16^{\binom{t-3}{2}} \leq \left(\frac{8 \cdot p(V)}{c}\right)^{t-3} \cdot 16^{t-3} \cdot 16^{\binom{t-3}{2}} = \left(\frac{8 \cdot p(V)}{c}\right)^{t-3} \cdot 16^{\binom{t-2}{2}},$$

using the definition of  $c'$  and Claim 9.4(ii). If we consider  $i \in \{1, \dots, m\}$ , then scaling up by  $1/w_i$  causes the maximum value to be at most

$$\frac{1}{w_i} \times \left(\frac{8 \cdot p(V)}{c}\right)^{t-3} \cdot 16^{\binom{t-2}{2}} \leq \left(\frac{8 \cdot p(V)}{c}\right)^{t-2} \cdot 16^{\binom{t-2}{2}}$$

by Claim 9.4(i). If we consider  $i = 0$ , then using that  $M_0 \equiv 1$ , the maximum value is  $4/3$ , which is only smaller (since  $t \geq 3$ ).

For property (P3), observe that  $\mathbb{E}[M \log M] \leq \mathbb{E}[M] \cdot \log M_{\max}$ , where  $M_{\max}$  is the maximum value  $M$  takes. But if  $M(v)$  is a m.o. r.v. then  $\mathbb{E}[M(v)] = 1$ , so  $\mathbb{E}[M(v) \log M(v)]$  is bounded by the logarithm of the expression in property (P2). ■

### 9.4.3 Satisfying the Assumptions

We still have to address the issue of the validity of the assumptions in Assumption 9.3. We assume we start off with a  $K_r$ -free graph with  $p(V) \leq O(\log \Delta)$ , and a contraction parameter  $c = \frac{1}{4}$  (say). Let  $p^t(v)$  be the probability values at some stage where the current vertex set is  $V^t$  (which is  $K_t$ -free), then by Claim 9.4 and algebra.

- (a)  $p^t(V^t) \leq 8^{r-t} p(V)$ , and
- (b) for all  $v \in V^t$ ,  $p^t(v) \leq p(v) \cdot \frac{p(V)^{r-t} 16^{\binom{r-t}{2}}}{c^{r-t}}$ .

Consequently,  $p^t(v) \leq p(v) \cdot (p(V)^r \cdot 16^{r^2})$ , and if we start off with  $p(v) \leq p^*$  and  $r \ll \sqrt{\log \Delta}$ , we ensure assumption (ii) that  $p^t(v) \leq 1$  for all stages  $t$ . Assumption (i) demands  $q^t(v) = \frac{p^t(v)}{c^t} = O(p^t(v) \cdot 2^t) \leq 1$  which is satisfied again by the same conditions.

## A Probabilistic Tools and Useful Lemmas

### A.1 Concentration Bounds

The following large-deviation bound is standard, see, e.g., [AS92].

**Theorem A.1 (A Large Deviation Bound)** *For independent  $[0, m]$ -bounded random variables  $X_1, X_2, \dots$ , with  $X := \sum_i X_i$  having mean  $\mathbb{E}[X] \leq \mu$ , given any  $\lambda \geq 0$ ,*

$$\Pr[|X - \mathbb{E}[X]| \geq \lambda] \leq 2 \exp \left\{ -\frac{\lambda^2}{m(2\mu + \lambda)} \right\}.$$

*In particular, this probability is at most  $1/\text{poly}(\Delta)$  when  $\lambda = O(\sqrt{\mu m \ln \Delta} + m \ln \Delta)$ .*

For a multilinear polynomial  $f(x) = f(x_1, x_2, \dots, x_n)$  with nonnegative coefficients and degree  $\leq q$ , and  $n$  independent random variables  $Y = (Y_1, Y_2, \dots, Y_n)$ , we define  $\mu_r$  (for every  $r \leq q$ ) as follows

$$\mu_r = \max_{\substack{S \subseteq [n], |S|=r \\ S=\{s_1, \dots, s_r\}}} \left( \frac{\partial^r f}{\partial x_{s_1} \partial x_{s_2} \dots \partial x_{s_r}} \Big|_{\mathbb{E}[Y_1], \dots, \mathbb{E}[Y_n]} \right)$$

Building on a long line of work, the following bound is presented by Schudy and Sviridenko [SS12].

**Theorem A.2 (Large Deviation for Polynomials)** *Consider independent  $[0, m]$ -bounded r.v.s  $X_1, X_2, \dots$  and let  $X := (X_1, X_2, \dots, X_n)$ . Let  $f(x) = f(x_1, x_2, \dots, x_n)$  be a multilinear polynomial of degree  $q$  with non-negative coefficients, and let  $f(X)$  have moment parameters  $\mu_0, \mu_1, \dots, \mu_q$ . There exists a universal constant  $C = C(q)$  such that*

$$\Pr [|f(X) - \mathbb{E}[f(X)]| \geq \lambda] \leq e^2 \max \left\{ \max_{r=1, \dots, q} \exp \left\{ -\frac{\lambda^2}{C \cdot m^r \cdot \mu_0 \mu_r} \right\}, \max_{r=1, \dots, q} \exp \left\{ -\left( \frac{\lambda}{C \cdot m^r \cdot \mu_r} \right)^{1/r} \right\} \right\}.$$

**Proof:** Use Theorem 1.2 of the Schudy-Sviridenko paper [SS12] and the observation that any  $[0, m]$ -bounded r.v. is moment bounded by parameter  $L = m$ .  $\blacksquare$

**Corollary A.3** *Consider independent  $[0, 1]$ -bounded r.v.s  $X_1, X_2, \dots$  and let  $X := (X_1, X_2, \dots, X_n)$ . Let  $f(x) = f(x_1, x_2, \dots, x_n)$  be a multilinear polynomial of degree 2 with non-negative coefficients, and let  $f(X)$  have mean  $\mathbb{E}[f(X)] \leq \mu$  and moment parameters  $\mu_1, \mu_2 \leq O(1)$ . Then*

$$\Pr [|f(X) - \mathbb{E}[f(X)]| \geq \lambda] \leq e^2 \max \left\{ \exp \left\{ -\frac{\lambda^2}{O(\mu)} \right\}, \exp \left\{ -O(\lambda)^{1/2} \right\} \right\}.$$

In particular, this probability is at most  $1/\text{poly}(\Delta)$  when  $\lambda = O(\sqrt{\mu \ln \Delta} + \ln^2 \Delta)$ .

## A.2 The Lovász Local Lemma

The following theorem essentially follows from Moser and Tardos [MT10].

**Theorem A.4** *Consider a set of  $n$  independent random variables  $\mathcal{F} = \{X_i\}_{i=1}^n$ , and assume that sampling each r.v. from the underlying distribution can be done in constant time. Given a collection of  $m$  subsets  $\{S_j \subseteq \mathcal{F}\}_{j=1}^m$  such that the “bad” event  $\mathcal{B}_j$  is completely determined by the r.v.s in subset  $S_j$ , define the degree  $d_j = |\{j' \in [m] \mid S_j \cap S_{j'} \neq \emptyset\}|$ . Define  $p_j := \Pr[\mathcal{B}_j]$ . Suppose*

$$(\max_j p_j) \cdot (\max_j d_j) \leq 1/4$$

*then there is an algorithm running in time  $\text{poly}(m, n)$  to find a setting of the random variables  $X_i$  such that none of the bad events occur.*

## A.3 Auxiliary Lemmas

**Lemma A.5** *Suppose  $h(v, p) \geq (1 - \delta) \ln \Delta$  and  $p(\mathcal{P}_v) \in (1 \pm \nu)$ , then  $p_a(\mathcal{P}_v) \geq 1 - 6(\delta + \nu)$ . If  $\xi_G(v, p) \leq 2K$  also holds, then  $p_c(\mathcal{P}_v) \geq 1 - 6(\delta + \nu) - 2\varepsilon$ .*

**Proof:** First, we prove the bound on  $p_a(\mathcal{P}_v)$ . Recall that  $p_a(v, \gamma) = p(v, \gamma) \cdot \mathbf{1}_{(p(v, \gamma) \leq \hat{p})}$ . Since any non-zero probability is at least  $1/s \geq 1/\Delta$ , the entropy

$$h(v, p) = - \sum_{\gamma} p(v, \gamma) \ln p(v, \gamma) \leq \left( \sum_{\gamma: p(v, \gamma) \in (1/s, \hat{p}]} p(v, \gamma) \right) \ln \Delta + \left( \sum_{\gamma: p(v, \gamma) > \hat{p}} p(v, \gamma) \right) \ln 1/\hat{p} \quad (50)$$

Let  $B := \sum_{\gamma: p(v, \gamma) > \hat{p}} p(v, \gamma)$ ; since  $p(\mathcal{P}_v) \in (1 \pm \nu)$ , we have that

$$p_a(\mathcal{P}_v) = \sum_{\gamma: p(v, \gamma) \in [0, \hat{p}]} p(v, \gamma) = \sum_{\gamma: p(v, \gamma) \in (1/s, \hat{p}]} p(v, \gamma) \in (1 - B \pm \nu).$$

Moreover,  $\ln 1/\hat{p} = (\frac{3}{4} + 5\varepsilon) \ln \Delta$ . Finally, by assumption,  $h(v, p) \geq (1 - \delta) \ln \Delta$ . Substituting into (50) and dividing throughout by  $\ln \Delta$ , and using  $\varepsilon = 1/100$ , we get

$$(1 - B + \nu) + \left( \frac{3}{4} + 5\varepsilon \right) B \geq 1 - \delta \quad \implies \quad B \leq 5(\delta + \nu). \quad (51)$$

Hence  $p_a(\mathcal{P}_v) \in [1 - 6(\delta + \nu), 1 + \nu]$ , which proves the first part of the claim.

Next, the bound on  $p_c(\mathcal{P}_v)$ . Recall that  $p_c(v, \gamma) = p_a(v, \gamma) \cdot \mathbf{1}_{(\sum_{u \sim v} p(u, \gamma) \leq 100 \ln \Delta)}$ . Since  $\varepsilon = 1/100$ , the threshold for zeroing out is  $\frac{\ln \Delta}{\varepsilon} \geq \frac{K}{\varepsilon}$ . Consequently, if  $S_v := \{\gamma \mid \sum_{u \sim v} p(u, \gamma) \geq \frac{K}{\varepsilon}\}$ , then  $p_c(\mathcal{P}_v) \geq p_a(\mathcal{P}_v) - \sum_{\gamma \in S_v} p(v, \gamma)$ . To bound the latter sum, observe that

$$\sum_{\gamma \in S_v} p(v, \gamma) \leq \frac{\varepsilon}{K} \sum_{\gamma \in S_v} p(v, \gamma) \sum_{u \sim v} p(u, \gamma) \leq \frac{\varepsilon}{K} \sum_{u \sim v} \xi_G(uv, p) = \frac{\varepsilon}{K} \xi_G(v, p) \leq 2\varepsilon.$$

Hence  $p_c(\mathcal{P}_v) \geq 1 - 6(\delta + \nu) - 2\varepsilon$  ■