

Counting independent sets in triangle-free graphs

Jeff Cooper*, Dhruv Mubayi†

November 15, 2011

Abstract

Ajtai, Komlós, and Szemerédi proved that for sufficiently large t every triangle-free graph with n vertices and average degree t has an independent set of size at least $\frac{n}{100t} \log t$. We extend this by proving that the number of independent sets in such a graph is at least

$$2^{\frac{1}{2400} \frac{n}{t} \log^2 t}.$$

This result is sharp for infinitely many t, n apart from the constant. An easy consequence of our result is that there exists $c' > 0$ such that every n -vertex triangle-free graph has at least

$$2^{c' \sqrt{n} \log n}$$

independent sets. We conjecture that the exponent above can be improved to $\sqrt{n}(\log n)^{3/2}$. This would be sharp by the celebrated result of Kim which shows that the Ramsey number $R(3, k)$ has order of magnitude $k^2/\log k$.

1 Introduction

An independent set in a graph $G = (V, E)$ is a set $I \subset V$ of vertices such that no two vertices in I are adjacent. The independence number of G , denoted $\alpha(G)$, is the size of the largest independent set in G . Determining the independence number of a graph is one of the most pervasive and fundamental problems in graph theory. The independence number naturally arises when studying other fundamental graph parameters like the chromatic number (minimum size of a partition of V into independent sets), clique number (independence number of the complementary graph), minimum vertex cover

*Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, IL 60607; email: jcoope8@uic.edu

†Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, IL 60607; email: mubayi@math.uic.edu; research supported in part by NSF grant DMS 0969092.

(complement of a maximum independent set), matching number (independence number in the line graph) and many others.

Throughout this paper, we suppose that G is a graph with n vertices and average degree t . Turán's [11] basic theorem of extremal graph theory, in complementary form, states that $\alpha(G) \geq \lceil n/(t+1) \rceil$ for any graph G . This bound is tight, as demonstrated by the complement of Turán's graph $G = \overline{T(n, r)}$ which, in the case $n = kr$ is the disjoint union of r cliques, each with k vertices (then $\alpha(G) = r$ and $t = k - 1$). Since G contains large cliques it is natural to ask whether Turán's bound on $\alpha(G)$ can be improved if we prohibit cliques of a prescribed (small) size in G .

In [1], Ajtai, Komlós, and Szemerédi showed that if G contains no K_3 , then this is indeed the case, by improving Turán's bound by a factor that is logarithmic in t . More precisely, they proved that if G is triangle-free, then

$$\alpha(G) \geq \frac{n}{100t} \log t.$$

Shortly after, Shearer [10] improved this to $\alpha(G) \geq (1 - o(1)) \frac{n}{t} \log t$ (assume for convenience throughout this paper that $\log = \log_2$). Random graphs [10] show that for infinitely many t and n with $t = t(n) \rightarrow \infty$ as $n \rightarrow \infty$, there are n -vertex triangle-free graphs with average degree t and independence number $(2 - o(1))((n/t) \log t)$. Consequently, the results of [1, 10] cannot be improved apart from the multiplicative constant.

There is a tight connection between the problem of determining $\alpha(G)$ and questions in Ramsey theory. More precisely, determining the minimum possible $\alpha(G)$ for a triangle-free G is equivalent to determining the Ramsey number $R(3, k)$, which is the minimum n so that every graph on n vertices contains a triangle or an independent set of size k . Moreover, the above lower bounds for $\alpha(G)$ are equivalent to the upper bound $R(3, t) = O(t^2/\log t)$. It was a major open problem, dating back to the 1940's, to determine the order of magnitude of $R(3, t)$, and this was achieved by Kim [7] who showed that for every n sufficiently large, there exists an n -vertex triangle-free graph G with $\alpha(G) < 9\sqrt{n \log n}$. As a consequence, the upper bound $R(3, t) = O(t^2/\log t)$ from [1] is of the correct order of magnitude.

In this paper, our goal is to take the result of Ajtai, Komlós, and Szemerédi [1] further by not only finding an independent set of the size guaranteed by their result, but by showing that many of the vertex subsets of approximately that size are independent sets.

Definition 1. Given a graph G , let $i(G)$ denote the number of independent sets in G .

Upper bounds for $i(G)$ have been motivated by combinatorial group theory. In [2], Alon showed that if G is a d -regular graph, then $i(G) \leq 2^{(1/2+o(d))n}$; he also conjectured

that

$$i(G) \leq (2^{d+1} - 1)^{n/2d}.$$

Kahn [6] proved this conjecture for d -regular bipartite graphs. Galvin [5] obtained a similar bound for any d -regular graph G , namely

$$i(G) \leq 2^{n/2(1+1/d+c/d\sqrt{\log d/d})}$$

for some constant c . Finally, Zhao [12] recently resolved Alon's conjecture.

In this paper, we consider lower bounds for $i(G)$. This problem is fundamental in extremal graph theory, indeed, the Erdős-Stone theorem [4] gives a lower bound for $i(G)$ that is the correct order of magnitude *provided n/t is a constant*. More recently, the problem in the range $t = \Theta(n)$ has been investigated by Razborov [9], Nikiforov [8], and Reiher. For example, the results of Razborov and Nikiforov determine $g(\rho, 3)$, the minimum triangle density of an n -vertex graph with edge density $\frac{1}{2} < \rho < 1$. Looking at the complementary graph, this gives tight lower bounds on the number of independent sets of size three in a graph with density $1 - \rho \sim \frac{t}{n}$.

Lower bounds for $i(G)$ appear not to have been studied with the same intensity when t is much smaller than n , in particular, when $t \rightarrow \infty$ and $t/n \rightarrow 0$. Let us make some easy observations that are relevant for our work here. We assume that $\alpha := \alpha(G) \leq n/4$. Since every subset of an independent set is also independent, Turán's theorem implies

$$i(G) \geq 2^\alpha \geq 2^{n/(t+1)}.$$

In Section 2, we will improve this to

$$i(G) \geq 2^{\frac{1}{250} \frac{n}{t} \log t}. \tag{1}$$

Our proof uses the standard probabilistic argument which establishes the order of magnitude given by Turán's bound on $\alpha(G)$. This result is certainly not new, and we present it only to serve as a warm-up for our main result in Section 3. Let us observe below that the result is essentially tight.

As no subset of size more than $\alpha(G)$ is independent, an easy upper bound on $i(G)$ (using $\alpha \leq n/4$) is

$$i(G) \leq \sum_{i=0}^{\alpha} \binom{n}{i} \leq 2^{\binom{n}{\alpha}}. \tag{2}$$

Since $\alpha(\overline{T(kr, r)}) = r = n/(t+1)$ (recall that $n = kr$ and $t = k-1$), this bound implies that as $n \rightarrow \infty$

$$i(\overline{T(kr, r)}) \leq 2^{\binom{kr}{r}} \leq 2^{(ek)^r} = 2^{1+r \log ek} = 2^{(1+o(1)) \frac{n}{t} \log t}.$$

Thus, apart from the constant, the exponent in (1) cannot be improved.

Our main result addresses the case where G contains no triangles. As in the case of the independence number, prohibiting triangles improves the bound in (1).

Theorem 2. (Main Result) *Suppose that G is a triangle-free graph on n vertices with average degree t , where t is sufficiently large. Then*

$$i(G) \geq 2^{\frac{n}{2400t} \log^2 t}. \quad (3)$$

Suitable modifications of Random graphs provide constructions of n -vertex triangle-free graphs G with average degree $t = t(n) \rightarrow \infty$ as $n \rightarrow \infty$, and $\alpha(G) = O((n/t) \log t)$. Plugging this into (2), we see that Theorem 2 is tight (apart from the constant) for infinitely many t . However, it remains open if the theorem is sharp for all t where $t = n^{1/2+o(1)}$. Indeed, the open problem that remains is to obtain a sharp lower bound on $i(G)$ for triangle-free graphs with no restriction on degree. Since all subsets of the neighborhood of a vertex of maximum degree are independent, $i(G) > 2^t$. Combining this with (3) we get

$$i(G) > \max\{2^t, 2^{\frac{n}{2400t} \log^2 t}\} > 2^{cn^{1/2} \log n}$$

for some constant $c > 0$. We conjecture that this can be improved as follows.

Conjecture 3. *There is an absolute positive constant c such that every n -vertex triangle-free graph G satisfies*

$$i(G) > 2^{cn^{1/2}(\log n)^{3/2}}.$$

The conjecture, if true, is sharp (apart from the constant in the exponent) by the graphs (due to Kim [7] and more recently Bohman [3]) which show that $R(3, t) = \Omega(t^2/\log t)$. Indeed, their graphs are triangle-free and have independence number

$$\alpha(G) = \Theta(t) = \Theta\left(\frac{n}{t} \log t\right) = \Theta(\sqrt{n \log n}),$$

so $i(G) \leq 2^{O(\sqrt{n \log^3 n})}$ by (2).

As mentioned before, throughout the paper, all logarithms are base 2. For a graph G , let $n(G)$, $e(G)$ and $t(G)$ denote the number of vertices, edges, and average degree of G .

2 General case

In this section, we give the simple proof of (1). Our purpose in doing this is to familiarize the reader with the general approach to the proof of Theorem 2 in the next section.

Proposition 4. *If G is a graph on n vertices with average degree t , where $2 \leq t \leq \frac{n}{800}$, then $i(G) \geq 2^{\frac{1}{250} \frac{n}{t} \log t}$.*

Proof. Set $k = \lfloor \frac{1}{100} \frac{n}{t} \rfloor$. Pick a k -set uniformly at random from all k -sets in $V(G)$. Let H be the subgraph induced by the k vertices. Then

$$\mathbf{E}[e(H)] = \frac{1}{2} nt \frac{\binom{n-2}{k-2}}{\binom{n}{k}} = \frac{1}{2} nt \frac{k(k-1)}{n(n-1)} < \frac{1}{2} \frac{tk^2}{n}.$$

Recall that Markov's inequality states that if X is a positive random variable and $a > 0$, then $\Pr[X \geq a] \leq \mathbf{E}[X]/a$; hence $\Pr[e(H) \geq 2\frac{1}{2} \frac{tk^2}{n}] \leq 1/2$. So for at least half of the choices for H , $e(H) \leq \frac{tk^2}{n}$. Therefore, the number of choices of H for which $e(H) \leq \frac{tk^2}{n}$ is at least

$$\frac{1}{2} \binom{n}{k} \geq \frac{1}{2} \left(\frac{n}{k}\right)^k > 2^{\frac{k}{2} \log n/k} = 2^{\frac{k}{2}(\log n - \log k)}. \quad (4)$$

Now, if $e(H) < \frac{tk^2}{n} = \frac{1}{100}k$, then at most $\frac{1}{50}k$ of the vertices in H have degree at least one. This in turn implies that H contains an independent set I of size at least $\frac{49}{50}k$. The set I can be obtained from any H which contains it; the number of ways to pick the $\frac{1}{50}k$ vertices of $H - I$ is at most

$$\binom{n}{k/50} \leq \left(\frac{50ne}{k}\right)^{k/50} = 2^{\frac{k}{50} \log 50ne - \frac{k}{50} \log k} \leq 2^{\frac{k}{50}(\log n - \log k) + \frac{k}{50} \log 100t}. \quad (5)$$

Combining this with (4) and using $\frac{20}{23} \frac{1}{100} \frac{n}{t} < k \leq \frac{1}{100} \frac{n}{t}$ for $t \leq \frac{n}{800}$,

$$i(G) \geq 2^{k(\frac{1}{2} - \frac{1}{50})(\log n - \log k) - \frac{k}{50} \log 100t} \geq 2^{k\frac{24}{50} \log 100t - \frac{k}{50} \log 100t} = 2^{k\frac{23}{50} \log 100t} > 2^{\frac{1}{250} \frac{n}{t} \log t}.$$

□

3 Triangle-free graphs

In this section we prove our main result, Theorem 2. We begin with some modifications of a lemma from [1] (see the proof of Lemma 4 in [1]).

Lemma 5. (Ajtai-Komlós-Szemerédi [1]) *Suppose that G is a triangle-free graph on n vertices with average degree t , and let $k \leq n/100t$. Let H be the subgraph consisting of k vertices chosen uniformly at random from all the k -sets contained in $\{v \in V(G) : \deg(v) \leq 10t\}$. Let M be the subgraph of G consisting of vertices adjacent to no vertex in H . Let n' and t' denote the number of vertices and average degree of M . Then the*

random variables H and M satisfy:

$$\mathbf{E}[n(M)] > n\left(1 - \frac{k}{n-t}\right)^{t+1} > \frac{9n}{10} \quad (6)$$

$$\mathbf{E}[e(M)] > \frac{nt}{2}\left(1 - \frac{k}{n-20t}\right)^{20t+1} > \frac{nt}{10} \quad (7)$$

$$\mathbf{E}[e(H)] = \frac{1}{2}nt\frac{k(k-1)}{n(n-1)} \leq \frac{tk^2}{n} \quad (8)$$

$$\mathbf{Var}[n(M)] < \frac{2nk(t+1)(10t+1)}{n-k-20t-2} < nt \quad (9)$$

$$\mathbf{Var}[e(M)] < 2400kt^4 < 40nt^3 \quad (10)$$

$$\mathbf{Var}[e(H)] \leq \frac{tk^2(10k+n)}{n^2} \quad (11)$$

Further, if $e(M) < (1 + \delta) \mathbf{E}[e(M)]$ and $n(M) > (1 - \delta) \mathbf{E}[n(M)]$, then $n'/t' > \nu n/t$, where $\delta = 800\sqrt{t/n}$ and $\nu = 1 - 1/t - c_{10}\sqrt{t/n}$ for some positive constant c_{10} .

Remark. Ajtai-Komlós-Szemerédi state their lemma for $k = n/100t$ and prove each of the first inequalities in (6), (7), (9), and (10) for all k . They prove each of the second inequalities for $k = n/100t$, but it is easily observed that they continue to hold for $k < n/100t$.

The next lemma is implied by the computation in the proof of Lemma 4 from [1]. However, the last statement of Lemma 6 is crucial to our proof of Theorem 2, so we make the computations in [1] explicit.

Lemma 6. *Suppose G is a triangle-free graph on $n > 2^{50}$ vertices with average degree $t \leq 2\sqrt{n} \log n$ and $k \leq n/100t$. Then G contains a subgraph H with $n(H) = k$ $e(H) \leq k/50$. Moreover, if M is the subgraph of G consisting of vertices adjacent to no vertex in H , then*

1. $n(M) > n(G)/2$ and
2. $n(M)/t(M) > \nu n/t$, where $\nu = 1 - 1/t - c_{10}\sqrt{t/n}$.

Further, if the vertices in H are chosen uniformly at random from all the k -sets contained in $\{v \in V(G) : \deg(v) \leq 10t\}$, then at least half of the choices for H satisfy $e(H) \leq k/50$, along with conditions 1 and 2.

Proof. Recall that for a random variable X and $a > 0$, Chebyshev's inequality states that $\Pr[|X - \mathbf{E}[X]| \geq a] \leq \mathbf{Var}[X]/a^2$. Thus, with $a = k/50 - \mathbf{E}[e(H)]$, Lemma 5 implies

$$\begin{aligned}
\Pr[e(H) \geq k/50] &\leq \frac{tk^2(10k+n)}{n^2(k/50 - \frac{k^2t}{2n})^2} \\
&= \frac{t(10k+n)}{n^2(1/50 - \frac{kt}{2n})^2} \\
&\leq \frac{t(n/10t+n)}{n^2(1/50 - 1/200)^2} \\
&= \frac{1/10+t}{n(3/200)^2} \\
&< 5000 \frac{t}{n} \\
&\leq 5000 \frac{2 \log n}{\sqrt{n}} \\
&\leq 1/1000.
\end{aligned}$$

So with probability at most $1/1000$, the condition $e(H) \leq k/50$ fails.

Set $\delta = 800\sqrt{t/n}$. Again by Lemma 5 and Chebyshev, with $a = \delta \mathbf{E}[n(M)]$,

$$\Pr[n(M) \leq n/2] \leq \Pr[n(M) \leq (1-\delta) \mathbf{E}[n(M)]] \leq \frac{nt}{(\frac{9n}{10})^2 800^2 \frac{t}{n}} \leq 1/1000.$$

Thus the probability that condition (1) fails is at most $1/1000$.

With $a = \delta \mathbf{E}[e(M)]$,

$$\Pr[e(M) \geq (1+\delta) \mathbf{E}[e(M)]] \leq \frac{40nt^3}{\frac{nt}{10} 800^2 \frac{t}{n}} = 1/160.$$

Since $\Pr[e(M) \geq (1+\delta) \mathbf{E}[e(M)] \text{ or } n(M) \leq n/2] < 1/160 + 1/1000$, the last assertion of Lemma 5 implies that the probability of condition (2) failing is at most $1/160 + 1/1000$. Therefore, the probability that condition $e(H) \leq k/50$ fails or condition (1) fails or condition (2) fails is at most $1/1000 + 1/1000 + 1/160 + 1/1000 < 1/2$. \square

Our proof of Theorem 2 is achieved by analyzing Algorithm 1 below. The algorithm is a slight modification of the algorithm from [1] that yields an independent set of size $\frac{1}{100} \frac{n}{t} \log t$. Recall that c_{10} is the constant that appear in Lemma 5.

Theorem 7. *Suppose Algorithm 1 is run on a triangle-free graph G with n vertices and average degree t , where $2^{100} < t < \sqrt{n} \log n$ and $n > (3c_{10})^{12}$. If Algorithm 1 terminates at line 12, then $|I| > \frac{1}{2} \frac{n}{t} \log^2 t$. Otherwise, for each iteration $i = 0, \dots, R-1$, Algorithm*

Algorithm 1: Independent set algorithm

Input: Triangle-free graph G with n vertices, average degree t
Output: Independent set I

```

1  $M_o = G$ ;
2  $R = \lfloor (\log t)/2 \rfloor$ ;
3 for  $i \leftarrow 0$  to  $R$  do
4    $n_i =$  number of vertices in  $M_i$ ;
5    $t_i =$  average degree in  $M_i$ ;
6    $\nu_i = 1 - 1/t_{i-1} - c_{10}\sqrt{t_{i-1}/n_{i-1}}$ ;
7   if  $i = 0$  or  $\nu_i > 1 - 1/\log t$  then
8     Apply Lemma 6 with  $G = M_i$  and  $k = \lfloor \frac{1}{200} \frac{n}{t} \rfloor$ ;
9      $M_{i+1}, H_{i+1} = M, H$  from Lemma 6;
10  else
11     $I =$  Independent set in  $M_{i-1}$  of size  $\lceil n_{i-1}/(t_{i-1} + 1) \rceil$ ;
12    return  $I$ ;
13  end
14 end
15  $H = H_1 \cup \dots \cup H_R$ ;
16  $I =$  Independent set in  $H$  of size  $\lceil \frac{48}{50} kR \rceil$ ;
17 return  $I$ ;
```

1 successfully applies Lemma 6 to the graph M_i to obtain a graph H_{i+1} with $k = \lfloor \frac{1}{200} \frac{n}{t} \rfloor$ vertices. Moreover, the graph H in line 15 is the disjoint union of the H_i , and the independent set I in line 16 consists of $\frac{48}{50}kR$ vertices from H .

Proof. We break our proof into two cases, depending on whether Algorithm 1 terminates at line 17 or 12.

Line 17: We need to show that Lemma 6 can be applied at every iteration and that the graph H in line 15 contains an independent set of size at least $\frac{48}{50}kR \geq \frac{1}{500} \frac{n}{t} \log t$.

If $i = 0$, then $k \leq n/100t$, $t \leq \sqrt{n} \log n$, and $n > t > 2^{50}$, so we may apply Lemma 6 to obtain graphs M_1 and H_1 , where $|V(H_1)| = k$. Suppose that $i > 0$ and that Lemma 6 was successfully applied at all previous iterations. Using 1 of Lemma 6, $i < R$, and $R = \lfloor (\log t)/2 \rfloor < (\log n)/2$,

$$n_i \geq n/2^i > n/2^R > \sqrt{n} > \sqrt{t} > 2^{50}. \quad (12)$$

By the condition in line 7, $\nu_i > 1 - 1/\log t$ for each iteration i . So by 2 of Lemma 6,

$\frac{n_i}{t_i} \geq \frac{n}{t} \nu_1 \nu_2 \dots \nu_i > \frac{n}{t} (1 - 1/\log t)^R$. Thus

$$\frac{n_i}{t_i} > \frac{n}{t} (1 - 1/\log t)^R > \frac{n}{t} \left(1 - \frac{R}{\log t}\right) \geq \frac{1}{2} \frac{n}{t}.$$

In particular,

$$k \leq \frac{1}{200} \frac{n}{t} \leq \frac{1}{100} \frac{n_i}{t_i}, \quad (13)$$

and also, $t < \sqrt{n} \log n$ and $n_i < n$ yield

$$t_i < 2n_i \frac{t}{n} \leq 2n_i \frac{\log n}{\sqrt{n}} \leq 2n_i \frac{\log n_i}{\sqrt{n_i}} = 2\sqrt{n_i} \log n_i. \quad (14)$$

The inequalities (12), (13), and (14) ensure that we may again apply Lemma 6 with M_i, M_{i+1}, H_{i+1} , and k playing the roles of G, M, H , and k , respectively. Applying Lemma 6 R times yields a collection of sparse graphs H_1, H_2, \dots, H_R , each with k vertices. Each H_i contains at most $\frac{2}{50}k$ vertices of degree at least one, so each H_i contains an independent set of size at least $\frac{48}{50}k$. By definition of M_i , these independent sets may be combined into one independent set of size at least $\frac{48}{50}kR$.

Line 12: Suppose that the algorithm terminates at line 12 during iteration $i + 1$. Then n_i, t_i (and n_{i+1}, t_{i+1}) have been defined and $\nu_{i+1} = 1 - 1/t_i - c_{10} \sqrt{t_i/n_i}$. If $t_i \leq (\frac{3}{2})^{2/3}$, then $1/t_i > 1/\log^3 t$. Assume $t_i > (\frac{3}{2})^{2/3}$. Then

$$\frac{2}{3} t_i^{-1/3} > 1/t_i. \quad (15)$$

By (12), $n_i > \sqrt{n}$, so for $n > (3c_{10})^{12}$,

$$n_i^3 > n^{3/2} > (3c_{10})^6 n > (3c_{10})^6 t_i.$$

This implies

$$\frac{1}{3} t_i^{1/3} > c_{10} \sqrt{t_i/n_i}. \quad (16)$$

Combining (15) and (16) yields

$$1 - t_i^{-1/3} < 1 - 1/t_i - c_{10} \sqrt{\frac{t_i}{n_i}} = \nu_{i+1} \leq 1 - 1/\log t.$$

Thus $1/t_i > 1/\log^3 t$. Since $t \geq 2^{100}$ (which implies that $t > (\log^5 t + \log^2 t)^2$) and $i < R \leq \frac{\log t}{2}$,

$$\frac{n_i}{t_i + 1} > \frac{n}{2^i (\log^3 t + 1)} \geq \frac{n}{\sqrt{t} (\log^3 t + 1)} = \frac{n}{t} \frac{\sqrt{t}}{(\log^3 t + 1)} > \frac{n}{t} \log^2 t.$$

Turan's theorem now implies that M_i contains an independent set of size at least $\frac{n_i}{t_i + 1} > \frac{1}{2} \frac{n}{t} \log^2 t$. \square

We now complete the proof of Theorem 2 by obtaining a lower bound on the number of outcomes given by line 17 of Algorithm 1.

Proof of Theorem 2. Recall that we are to show that if G is a triangle-free graph on n vertices with average degree t sufficiently large, then $i(G) \geq 2^{\frac{n}{2400t} \log^2 t}$. Assume $t > \max\{(3c_{10})^{12}, 2^{100}\}$. Then $n > t > (3c_{10})^{12}$. Also, if $t > \sqrt{n} \log n$, then G has a vertex whose neighborhood contains at least $t > \frac{n}{\sqrt{n}} \log n > \frac{n}{t} \log^2 n$ vertices. Since G is triangle-free, this neighborhood forms an independent set, which contains at least $2^{\frac{n}{t} \log^2 n} > 2^{\frac{n}{2400t} \log^2 t}$ subsets, which are also independent. Thus we may assume that $t \leq \sqrt{n} \log n$; in particular, G satisfies the hypotheses of Theorem 7.

If the algorithm terminates at line 12, then G contains an independent set of size at least $\frac{n}{2t} \log^2 t$; so G contains at least $2^{\frac{n}{2t} \log^2 t} > 2^{\frac{n}{2400t} \log^2 t}$ independent sets, and we are done. Thus we may assume that Algorithm 1 terminates at line 17. Consequently, at each iteration i , the algorithm applies Lemma 6 to pick a sparse graph with $k = \lfloor \frac{1}{2} \frac{n}{100t} \rfloor$ vertices. The vertices in this graph are chosen from

$$L_i = \{v \in V(M_i) : \deg(v) \leq 10t_i\}.$$

Note that

$$n_i t_i = \sum_{v \in L_i} \deg(v) + \sum_{v \in V(M_i) - L_i} \deg(v) \geq \sum_{v \in V(M_i) - L_i} \deg(v) \geq (n_i - |L_i|) 10t_i.$$

This, together with 1 in Lemma 6, implies $|L_i| \geq \frac{9}{10} n_i > \frac{9}{10} n_{i-1} / 2 \geq \frac{9}{10} n / 2^i$. At least half of the k -sets in L_i satisfy the conditions of Lemma 6, so the number of choices for H_i is at least

$$\frac{1}{2} \binom{|L_i|}{k} \geq \frac{1}{2} \binom{.9n/2^i}{k}.$$

Therefore, the number of choices for the sequence H_1, \dots, H_R is at least

$$\begin{aligned} \prod_{i=0}^{R-1} \frac{1}{2} \binom{|L_i|}{k} &\geq \frac{1}{2^R} \prod_{i=0}^{R-1} \binom{.9n/2^i}{k} \geq \frac{1}{2^R} \left(\frac{.9n}{k}\right)^{kR} 2^{-kR^2/2} \\ &= 2^{kR \log .9n - kR \log k - kR^2/2 - R} \\ &= 2^{kR(\log n - \log k) - kR^2/2 + kR \log .9 - R} \\ &> 2^{kR(\log n - \log k) - \frac{kR}{2} \log t - kR - R} \\ &> 2^{kR(\log n - \log k) - \frac{kR}{4} \log t - 2kR}. \end{aligned} \tag{17}$$

Recall that Algorithm 1 obtains an independent set I of size $\lceil \frac{48}{50} kR \rceil$ from the graph $H = H_1 \cup \dots \cup H_R$. For a fixed I , the number of graphs H that yield I is at most the

number of possibilities for $H - I$. This is at most $\binom{n}{|V(H-I)|}$, which is at most

$$\begin{aligned} \binom{n}{\frac{2}{50}kR} &\leq \left(\frac{50ne}{2kR}\right)^{\frac{2}{50}kR} = 2^{\frac{2}{50}kR \log n + \frac{2}{50}kR \log 50e - \frac{2}{50}kR \log 2kR} \\ &< 2^{\frac{2}{50}kR(\log n - \log k) + \frac{2}{50}kR \log 50e} \\ &< 2^{\frac{2}{50}kR(\log n - \log k) + kR}. \end{aligned} \quad (18)$$

For a fixed H , the number of partitions $H_1 \cup \dots \cup H_R = H$ is at most the number of partitions of kR elements into R sets of size k , which is less than

$$\binom{kR}{k}^R \leq (Re)^{kR} = 2^{kR \log Re} < 2^{kR \log R + 2kR} < 2^{kR \frac{1}{4}R + 2kR} \leq 2^{kR \frac{1}{8} \log t + 2kR}. \quad (19)$$

Since each H yields an independent set, the total number of independent sets that can be returned at line 17 of the algorithm is at least

$$\frac{\# \text{ of ways to obtain } H}{(\# \text{ of } H \text{ that yield a fixed } I)(\# \text{ of partitions that yield } H)}.$$

Since $\frac{n}{268t} \leq k \leq \frac{n}{200t}$ and $R > (\log t)/3$, (17), (18), and (19) imply that this is at least

$$\begin{aligned} 2^{\frac{48}{50}kR(\log n - \log k) - \frac{5}{8}kR \log t - 5kR} &\geq 2^{\frac{48}{50}kR \log 200t - \frac{5}{8}kR \log t - 5kR} \\ &> 2^{\frac{134}{400}kR \log t} \\ &> 2^{\frac{134}{1200}k \log^2 t} \\ &> 2^{\frac{1}{2400} \frac{n}{t} \log^2 t}. \end{aligned}$$

□

References

- [1] Miklós Ajtai, János Komlós, and Endre Szemerédi, *A dense infinite Sidon sequence*, European J. Combin. **2** (1981), no. 1, 1–11. MR 611925 (83f:10056)
- [2] Noga Alon, *Independent sets in regular graphs and sum-free subsets of finite groups*, Israel J. Math. **73** (1991), no. 2, 247–256. MR 1135215 (92k:11024)
- [3] Tom Bohman, *The triangle-free process*, Adv. Math. **221** (2009), no. 5, 1653–1677. MR 2522430 (2010h:05271)
- [4] Paul Erdős and Arthur H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc. **52** (1946), 1087–1091. MR 0018807 (8,333b)

- [5] David Galvin, *An upper bound for the number of independent sets in regular graphs*, Discrete Math. **309** (2009), no. 23-24, 6635–6640. MR 2558628 (2010m:05217)
- [6] Jeff Kahn, *An entropy approach to the hard-core model on bipartite graphs*, Combin. Probab. Comput. **10** (2001), no. 3, 219–237. MR 1841642 (2003a:05111)
- [7] Jeong Han Kim, *The Ramsey number $R(3, t)$ has order of magnitude $t^2/\log t$* , Random Structures Algorithms **7** (1995), no. 3, 173–207. MR 1369063 (96m:05140)
- [8] Vladimir Nikiforov, *The number of cliques in graphs of given order and size*, Trans. Amer. Math. Soc. **363** (2011), no. 3, 1599–1618. MR 2737279
- [9] Alexander A. Razborov, *On the minimal density of triangles in graphs*, Combin. Probab. Comput. **17** (2008), no. 4, 603–618. MR 2433944 (2009i:05118)
- [10] James B. Shearer, *A note on the independence number of triangle-free graphs*, Discrete Math. **46** (1983), no. 1, 83–87. MR 708165 (85b:05158)
- [11] Paul Turán, *On an extremal problem in graph theory (in hungarian)*, Math. Fiz. Lapok **48** (1941), 436–452.
- [12] Yufei Zhao, *The number of independent sets in a regular graph*, Combin. Probab. Comput. **19** (2010), no. 2, 315–320. MR 2593625 (2011e:05200)