Counting independent sets in hypergraphs

Jeff Cooper*, Kunal Dutta†, Dhruv Mubayi‡

February 7, 2014

Abstract

Let G be a triangle-free graph with n vertices and average degree t. We show that G contains at least

$$e^{(1-n^{-1/12})\frac{1}{2}\frac{n}{t}\ln t(\frac{1}{2}\ln t-1)}$$

independent sets. This improves a recent result of the first and third authors [8]. In particular, it implies that as $n \to \infty$, every triangle-free graph on n vertices has at least $e^{(c_1-o(1))\sqrt{n}\ln n}$ independent sets, where $c_1 = \sqrt{\ln 2}/4 = 0.208138...$ Further, we show that for all n, there exists a triangle-free graph with n vertices which has at most $e^{(c_2+o(1))\sqrt{n}\ln n}$ independent sets, where $c_2 = 2\sqrt{\ln 2} = 1.665109...$ This disproves a conjecture from [8].

Let H be a $(k+1)$-uniform linear hypergraph with n vertices and average degree t. We also show that there exists a constant c_k such that the number of independent sets in H is at least

$$e^{c_k \frac{n}{t^{1/k}} \ln^{1+1/k} t}.$$

This is tight apart from the constant c_k and generalizes a result of Duke, Lefmann, and Rödl [9], which guarantees the existence of an independent set of size $\Omega(\frac{n}{t^{1/k}} \ln^{1/k} t)$. Both of our lower bounds follow from a more general statement, which applies to hereditary properties of hypergraphs.

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

†Algorithms and Complexity Department, Max Planck Institute for Informatics, Saarbrücken, Germany. (part of this work was done at: Indian Statistical Institute, New Delhi, India); email: kdutta@mpi-inf.mpg.de

‡Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, IL 60607; Research partially supported by NSF grants DMS 0969092 and 1300138; email: mubayi@uic.edu

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.

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

In [3], Ajtai, Komlós, and Szemerédi gave a semi-random algorithm for finding an independent set of size at least $\frac{n}{100t}\ln t$ in any triangle-free graph G with n vertices and average degree t. By analyzing their algorithm, the first and third authors [8] recently showed that for any such graph,

$$i(G) \geq 2^{\frac{1}{2400}\frac{n}{t}\log_2^2 t}. \tag{1}$$

As a consequence, they proved that every triangle-free graph has at least $2^{\Omega(\sqrt{n}\ln n)}$ independent sets and conjectured that this could be improved to $2^{\Omega(\sqrt{n}\ln^{3/2} n)}$, based on the best constructions of Ramsey graphs by Kim [12].

In this paper, we give a simpler proof of (1), which substantially improves the constant in the exponent and avoids any analysis of the algorithm in [3]. Further, we show that our bound is not far from optimal, by disproving the conjecture in [8] and constructing a triangle-free graph with at most $2^{O(\sqrt{n}\ln n)}$ independent sets. The construction is obtained by modifying the graph obtained by the triangle-free process. Our bounds follow from the detailed analysis of this process by Bohman-Keevash [6] and Fiz Pontiveros-Griffiths-Morris [10].

Theorem 1. *Let G be a triangle-free graph with n vertices and average degree t. Then*

$$i(G) \geq \max\{e^{(1-n^{-1/12})\frac{1}{2}\frac{n}{t}\ln t(\frac{1}{2}\ln(t)-1)}, 2^t\}.$$

Consequently, for every triangle-free graph H on n vertices,

$$i(H) \geq e^{(1-o(1))\frac{\sqrt{n\ln 2}\ln n}{4}}.$$

The constant in the exponent above is $\sqrt{\ln 2}/4 \approx 0.2081$. As we show below it is not far from optimal as we have an upper bound with exponent $2\sqrt{\ln 2} \approx 1.665$.

Theorem 2. *For all n, there exists a triangle-free graph G on n vertices with*

$$i(G) \leq e^{(1+o(1))(2\sqrt{\ln 2})\sqrt{n}\ln n}.$$

Using random graphs, one can show that for $t < n^{1/3}$, there is a triangle-free graph G with independence number at most $(2n/t)\ln t$. Consequently,

$$i(G) \le \sum_{i=1}^{\alpha(G)} \binom{n}{i} \le 2\binom{n}{\alpha(G)} \le 2\left(\frac{te}{2\ln t}\right)^{\frac{2n}{t}\ln t} < 2e^{\ln(te)\frac{2n}{t}\ln t} = e^{(1+o(1))\frac{2n}{t}\ln^2 t},$$

so the constant in the exponent of Theorem 1 is within a factor of 8 of the best possible constant.

1.1 Linear hypergraphs

Fix $k \ge 1$. Using the semi-random method, Ajtai, Komlós, Pintz, Spencer, and Szemerédi [2] showed that there exists c_k such that every $(k+1)$-uniform hypergraph H with n vertices, average degree t, and girth 5 satisfies $\alpha(H) \ge c_k \frac{n}{t^{1/k}} \ln^{1/k} t$. A hypergraph is *linear* (or has girth 3) if any two edges intersect in at most one vertex. Duke, Lefmann, and Rödl [9] (using the result of [2]) showed that there exists c'_k such that every linear $(k+1)$-uniform hypergraph H with n vertices and average degree t satisfies

$$\alpha(H) \ge c'_k \frac{n}{t^{1/k}} \ln^{1/k} t.$$

This leads to our second theorem.

Theorem 3. *Fix $k \ge 1$. There exists $c''_k > 0$ such that the following holds: For every $(k+1)$-uniform, linear hypergraph H on n vertices with average degree t,*

$$i(H) \ge e^{c''_k \frac{n}{t^{1/k}} \ln^{1+1/k} t}. \tag{2}$$

In [2], Ajtai, Komlós, Pintz, Spencer, and Szemerédi observed that, for infinitely many t and n, there exists a $(k+1)$-uniform, linear hypergraph H with n vertices, average degree t, and independence number at most $b'_k \frac{n}{t^{1/k}} \ln^{1/k} t$. For this hypergraph,

$$i(H) \le e^{b''_k \frac{n}{t^{1/k}} \ln^{1+1/k} t},$$

so (2) is tight up to the constant in the exponent.

1.2 Hereditary Properties

In [7], Colbourn, Hoffman, Phelps, Rödl, and Winkler counted the number of partial $S(t, t+1, n)$ Steiner systems by analyzing a semi-random algorithm; Using the same techniques, Grable and Phelps [11] extended their result to partial $S(t, k, n)$ Steiner systems. Asratian and Kuzjurin [5] gave a simpler proof of the bound in [11], which avoids any algorithm analysis. Theorems 1 and 3 both follow from a more general result (Theorem 4 below), which is based on this simpler proof. Since our proof avoids any analysis of how the independent sets are obtained, we are able to extend the bound in [8] from triangle-free graphs to a more general hypergraph setting. Recall that a *hereditary property* $\mathcal{P}$ of hypergraphs is any set of hypergraphs which is closed under vertex-deletion.

Theorem 4. *Fix $k \geq 1$ and $\epsilon \in (0, \frac{4}{k+1})$. Let $\mathcal{P}$ be any hereditary hypergraph property. Suppose there exists a non-decreasing function f so that every $(k+1)$-uniform hypergraph $H \in \mathcal{P}$ with n vertices and average degree at most t satisfies*

$$\alpha(H) \geq \frac{n}{t^{1/k}} f(t).$$

Then there exists $n_0 = n_0(\epsilon)$ such that every $(k+1)$-uniform hypergraph $H \in \mathcal{P}$ with $n \geq n_0$ vertices and average degree at most $t < n^k$ satisfies

$$i(H) \geq e^{\alpha' \frac{n}{t^{1/k}} \ln t},$$

where

$$\alpha' = \begin{cases} (1 - n^{-\epsilon/21}) \frac{1}{k+1} f(t^{\frac{1}{k+1}}), & \text{if } H \text{ is linear} \\ (1 - n^{-\epsilon/21}) \frac{1-\epsilon}{k(2k+1)} f(t^{\frac{2k+\epsilon}{2k+1}}), & \text{otherwise.} \end{cases}$$

Remark 5. In [1], Ajtai, Erdős, Komlós, and Szemerédi asked if every K_r-free graph has independence number at least $\Omega(\frac{n}{t} \ln t)$. They gave a lower bound of $\Omega(\frac{n}{t} \ln \ln t)$, which Shearer [16] later improved to $\Omega(\frac{n}{t} \frac{\ln t}{\ln \ln t})$ for sufficiently large t. Theorem 4 implies that if there exists c_r so that every K_r-free graph G satisfies $\alpha(G) \geq c_r \frac{n}{t} \ln t$, then

$$i(G) \geq \binom{n}{\Omega(\frac{n}{t} \ln t)} = e^{\Omega(\frac{n}{t} \ln^2 t)}.$$

2 Lower Bounds

Theorems 1 and 3 follow from the linear case of Theorem 4. We will prove Theorem 4 for linear hypergraphs and afterward describe the changes needed for non-linear hypergraphs.

We first state a version of the Chernoff bound and two claims, which contain the main differences between the linear and non-linear cases. The proofs of the claims will follow the proof of the theorem.

Chernoff Bound (Chernoff bound [14]). *Suppose X is the sum of n independent variables, each equal to 1 with probability p and 0 otherwise. Then for any $0 \leq t \leq np$,*

$$\mathbb{P}(|X - np| > t) < 2e^{-t^2/3np}.$$

Setup. *Fix $k \geq 1$ and $\epsilon \in (0, \frac{4}{k+1})$. Let H be a $(k+1)$-uniform hypergraph with n vertices, average degree at most $t < n^k$, and maximum degree at most $tn^{\epsilon/8}$. Select each vertex of H independently with probability p. Let m' denote the sum of vertex degrees in the subgraph induced by the selected vertices.*

The next two claims come under the assumption of the setup.

Claim 6. *If H is linear and $p = t^{\frac{-1}{k+1}}$, Then for all $n > n_0(\epsilon)$,*

$$\mathbb{P}\left[m' > ntp^{k+1} + \frac{ntp^{k+1}}{n^{\epsilon/20}}\right] < n^{-2}.$$

Claim 7. *If $p = t^{\frac{\epsilon-1}{k(2k+1)}}$, then for all $n > n_0(\epsilon)$,*

$$\mathbb{P}\left[m' > ntp^{k+1} + \frac{ntp^{k+1}}{n^{\epsilon/20}}\right] < n^{-2}.$$

Proof of Theorem 4 (linear case). Fix $k \geq 1$ and $\epsilon \in (0, \frac{4}{k+1})$. Let $H \in \mathcal{P}$ be a $(k+1)$-uniform, linear hypergraph with n vertices and average degree at most $t < n^k$. We assume $n \geq n_0$, where n_0 is chosen implicity so that several inequalities throughout the proof are satisfied. We consider two cases. In Case 1, we require that the maximum degree of H is at most $tn^{\epsilon/8}$, while Case 2 requires the maximum degree of H to be at least $tn^{\epsilon/8}$.

Case 1: The maximum degree of H is at most $tn^{\epsilon/8}$.

Select each vertex of H independently with probability $p = t^{-\frac{1}{k+1}}$. Let H' denote the subgraph of H induced by the selected vertices. Let n' denote the the number of vertices in H'. Since $t < n^k$ and $\epsilon < \frac{4}{k+1}$,

$$np = nt^{-\frac{1}{k+1}} > n^{1-k/(k+1)} = n^{\frac{1}{k+1}} > n^{\epsilon/4}.$$

By the Chernoff bound,

$$\mathbb{P}[|n' - np| > \frac{np}{n^{\epsilon/20}}] \leq 2e^{-np/3n^{\epsilon/20}} < n^{-2}. \tag{3}$$

Let m' denote the sum of vertex degrees in H'. By linearity of expectation,

$$\mathbb{E}[m'] = ntp^{k+1}.$$

Set $\lambda = n^{-\epsilon/20}$. By Claim 6,

$$\mathbb{P}[m' > (1+\lambda)ntp^{k+1}] < n^{-2}. \tag{4}$$

Therefore, by the union bound, with probability at least $1 - 2n^{-2} > 1 - 1/n$, H' satisfies both

$$m' \leq (1+\lambda)ntp^{k+1}$$

and

$$n' \geq (1-\lambda)np.$$

Let $t' = (1+3\lambda)tp^k$. Then with probability at least $1 - 1/n$, H' has average degree at most

$$m'/n' \leq \frac{(1+\lambda)ntp^{k+1}}{(1-\lambda)np} \leq (1+3\lambda)tp^k = t'.$$

Since $\mathcal{P}$ is hereditary, $H' \in \mathcal{P}$. Thus, with probability at least $1 - 1/n$, H' has an independent set of size at least

$$\begin{aligned}
\frac{n'}{t'^{1/k}} f(t') \geq \frac{(1-\lambda)np}{((1+3\lambda)tp^k)^{1/k}} f((1+3\lambda)tp^k) &= \frac{(1-\lambda)n}{(1+3\lambda)^{1/k} t^{1/k}} f((1+3\lambda)tp^k) \\
&\geq \frac{(1-\lambda)n}{(1+3\lambda)t^{1/k}} f((1+3\lambda)tp^k) \\
&> (1-6\lambda)\frac{n}{t^{1/k}} f((1+3\lambda)tp^k) \\
&\geq (1-6\lambda)\frac{n}{t^{1/k}} f(tp^k),
\end{aligned}$$

where we used that f is non-decreasing in the last inequality.

Let $g = (1-6\lambda)\frac{n}{t^{1/k}} f(tp^k)$. Suppose I is an independent set in H with at least g vertices. Then

$$\mathbb{P}[I \subset V(H')] = p^{|I|} \leq p^g.$$

Let N denote the number of independent sets in H with at least g vertices, and let the random variable N' denote the number of independent sets in H' with at least g vertices. By Markov's inequality,

$$1 - 1/n < \mathbb{P}[N' \geq 1] \leq \mathbb{E}[N'] \leq Np^g = Ne^{-g \ln p}$$

Thus

$$\begin{aligned}
N > (1-1/n)e^{-g\ln p} &= (1-1/n)e^{(1-6\lambda)\frac{1}{k+1}\frac{n}{t^{1/k}} f(t^{\frac{1}{k+1}}) \ln t} \\
&> (1-1/n)e^{(1-n^{-\epsilon/21})\frac{1}{k+1}\frac{n}{t^{1/k}} f(t^{\frac{1}{k+1}}) \ln t}.
\end{aligned} \tag{5}$$

Case 2: The maximum degree of H is more than $tn^{\epsilon/8}$.

Let

$$K = \{u \in V(H) : \deg(u) > tn^{\epsilon/8}/2\}.$$

Let H' denote the subgraph of H induced by $V(H) - K$, and let $n' = |V(H')|$. Since

$$\begin{aligned}\frac{1}{n'}\sum_{v\in V(H')} \deg_{H'}(v) - \frac{1}{n}\sum_{v\in V(H')} \deg_{H}(v) &\le (\frac{1}{n'} - \frac{1}{n})\sum_{v\in V(H')} \deg_{H}(v)\\ &\le (\frac{1}{n'} - \frac{1}{n})n'tn^{\epsilon/8}/2\\ &= (n-n')\frac{tn^{\epsilon/8}}{2n}\\ &\le \frac{1}{n}\sum_{v\in K} \deg_{H}(v),\end{aligned}$$

the average degree of H' is at most

$$\frac{1}{n}\sum_{v\in K} \deg_{H}(v) + \frac{1}{n}\sum_{v\in V(H')} \deg_{H}(v) = \frac{1}{n}\sum_{v\in V(H)} \deg_{H}(v) \le t.$$

Also, because

$$tn \ge \sum_{u\in V(H)} \deg_H(u) \ge \sum_{u\in K} \deg_H(u) > |K|tn^{\epsilon/8}/2,$$

$|K| < 2n^{1-\epsilon/8}$, and so $n' > n(1 - 2n^{-\epsilon/8}) > n/2^{8/\epsilon}$. Thus H' has maximum degree at most $tn^{\epsilon/8}/2 < tn'^{\epsilon/8}$. Further, since H has maximum degree at least $tn^{\epsilon/8}$ and at most n^k, $t < n^{k-\epsilon/8}$. Hence $t < n^{k-\epsilon/8} < n'^k$. Thus Case 1 implies that

$$i(H') \ge (1-1/n')e^{(1-6\lambda)\frac{1}{k+1}\frac{n'}{t^{1/k}}f(t^{\frac{1}{k+1}})\ln t} > (1-2/n)e^{(1-6\lambda)(1-n^{-\epsilon/8})\frac{1}{k+1}\frac{n}{t^{1/k}}f(t^{\frac{1}{k+1}})\ln t},$$

where $\lambda = n'^{-\epsilon/20}$. We conclude that

$$i(H) \ge i(H') \ge e^{(1-n^{-\epsilon/21})\frac{1}{k+1}\frac{n}{t^{1/k}}f(t^{\frac{1}{k+1}})\ln t}. \tag{6}$$

□

The proof of Theorem 4 when H is non-linear is similar. We set $p = t^{\frac{\epsilon-1}{k(2k+1)}}$. Since we still have $np > n^{\epsilon/4}$, (3) still holds. We then use Claim 7 instead of Claim 6 to prove (4). The proof then proceeds in the same way until we get to (5), where, using the different value of p, we instead obtain

$$N > (1-1/n)e^{(1-6\lambda)\frac{1-\epsilon}{k(2k+1)}\frac{n}{t^{1/k}}f(t^{\frac{2k+\epsilon}{2k+1}})\ln t}.$$

Finally, (6) becomes

$$e^{(1-n^{-\epsilon/21})\frac{1-\epsilon}{k(2k+1)}\frac{n}{t^{1/k}}f(t^{\frac{2k+\epsilon}{2k+1}})\ln t}.$$

We now prove Theorem 1 and Theorem 3.

Proof of Theorem 1. Shearer [15] showed that every triangle-free graph with n vertices and average degree t has independence number at least $\frac{n}{t}(\ln(t)-1)$. Since being triangle-free is hereditary and graphs are 2-uniform, linear hypergraphs, we may apply Theorem 4 (with $f(t)=\ln(t)-1$) to conclude that for $\epsilon = 21/12 \in (0,2)$, there exists n_0 such that every triangle-free graph G with $n \geq n_0$ vertices and average degree at most t satisfies

$$i(G) \geq e^{(1-n^{-\epsilon/21})\frac{1}{2}\frac{n}{t}\ln t(\frac{1}{2}\ln(t)-1)} > e^{(1-n^{-1/12})\frac{1}{2}\frac{n}{t}\ln t(\frac{1}{2}\ln(t)-1)}.$$

Suppose G is a triangle-free graph with $n < n_0$ vertices and average degree t. Choose an integer r so that $rn \geq n_0$. Let G' be the disjoint union of r copies of G. Then $i(G') = i(G)^r$, so by the previous paragraph,

$$\begin{aligned} i(G) = i(G')^{1/r} &\geq (e^{(1-(rn)^{-1/12})\frac{1}{2}\frac{rn}{t}\ln t(\frac{1}{2}\ln(t)-1)})^{1/r} \\ &\geq e^{(1-n^{-1/12})\frac{1}{2}\frac{n}{t}\ln t(\frac{1}{2}\ln(t)-1)}. \end{aligned}$$

This completes the proof of the first bound in Theorem 1. For the second part, consider a triangle-free graph G having average degree t. G contains a vertex u with degree at least t. The neighborhood of u is an independent set, which contains 2^t independent sets. Therefore, every triangle-free graph has at least

$$\max\{2^t, e^{(1-n^{-1/12})\frac{1}{2}\frac{n}{t}\ln t(\frac{1}{2}\ln(t)-1)}\}$$

independent sets. This is minimized when $t = (\frac{1}{4}+o(1))\sqrt{n/\ln 2}\ln n$, so every triangle-free graph on n vertices has at least

$$2^{(1-o(1))\cdot\frac{\sqrt{n}\ln n}{4\sqrt{\ln 2}}} = e^{(1-o(1))\cdot\frac{\sqrt{n\ln 2}\ln n}{4}}$$

independent sets. $\square$

Proof of Theorem 3. Duke, Lefmann, and Rödl [9] showed that every $(k+1)$-uniform linear hypergraph with n vertices and average degree at most t has independence number at least $c'_k\frac{n}{t^{1/k}}\ln^{1/k} t$. Since linearity is a hereditary property, we may apply Theorem 4 (with $f(t) = c'_k \ln^{1/k} t$) to conclude that for $\epsilon = \frac{3}{k+1} \in (0, \frac{4}{k+1})$, there exists n_0 such that every $(k+1)$-uniform linear hypergraph H with $n \geq n_0$ vertices satisfies

$$i(H) \geq e^{(1-n^{-1/(7(k+1))})\frac{c'_k}{k+1}\frac{1}{(k+1)^{1/k}}\frac{n}{t^{1/k}}\ln^{1+1/k} t} > e^{c''_k\frac{n}{t^{1/k}}\ln^{1+1/k} t}.$$

If H is a $(k+1)$-uniform linear hypergraph with $n < n_0$ vertices, then we proceed in the same way as in the proof of Theorem 1. $\square$

It only remains to prove the claims stated at the beginning of this section. We first prove Claim 6. We will use the following theorem of Kim and Vu [13]:

Theorem 8. *Suppose F is a hypergraph such that $W = V(F)$ and $|f| \le s$ for all $f \in F$. Let*

$$Z = \sum_{f \in F} \prod_{i \in f} z_i,$$

where the z_i, $i \in W$ are independent random variables taking values in $[0, 1]$. For $A \subset W$ with $|A| \le s$, let

$$Z_A = \sum_{f \in F : f \supset A} \prod_{i \in f - A} z_i.$$

Let $M_A = \mathbb{E}[Z_A]$ and $M_j = \max_{A:|A| \ge j} M_A$ for $j \ge 0$. Then there exists positive constants $a = a(s)$ and $b = b(s)$ such that for any $\lambda > 0$,

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \ge a\lambda^s \sqrt{M_0 M_1}] \le b|W|^{s-1} e^{-\lambda}.$$

Proof of Claim 6. Apply Theorem 8 with $F = H$ and $\mathbb{P}[z_i = 1] = p = t^{-\frac{1}{k+1}}$. Note first that

$$\mathbb{E}[Z_\emptyset] \le ntp^{k+1} = nt^{1-1} = n.$$

Since the maximum degree of H is at most $tn^{\epsilon/8}$,

$$\mathbb{E}[Z_{\{u\}}] \le tn^{\epsilon/8} p^k = n^{\epsilon/8} t^{\frac{1}{k+1}}$$

for any $u \in V(G)$. By linearity, for any $A \subset V(G)$ with $|A| \ge 2$,

$$\mathbb{E}[Z_A] \le p^{k+1-|A|} \le 1.$$

Since $t \le n^k$ and $\epsilon < \frac{4}{k+1}$, $n \ge n^{\epsilon/8} t^{\frac{1}{k+1}}$. Further, $n^{\epsilon/8} t^{\frac{1}{k+1}} \ge 1$. Therefore $M_0 \le n$ and $M_1 \le n^{\epsilon/8} t^{1/(k+1)}$. Theorem 8 therefore implies that there exist constants $a = a(k)$ and $b = b(k)$ such that

$$\mathbb{P}[|m' - \mathbb{E}[m']| > a((k+3)\ln n)^{k+1} \sqrt{ntp^{k+1} tn^{\epsilon/8} p^k}] \le bn^k e^{-(k+3)\ln n}.$$

Since $t \le n^k$ and $\epsilon < \frac{4}{k+1}$,

$$\sqrt{ntp^{k+1} tn^{\epsilon/8} p^k} = \frac{ntp^{k+1}}{n^{1/2-\epsilon/16} p^{1/2}} \le \frac{ntp^{k+1}}{n^{\epsilon/16}}.$$

Thus, since $\mathbb{E}[m'] \le ntp^{k+1}$,

$$\begin{aligned}
\mathbb{P}[m' > ntp^{k+1} + \frac{ntp^{k+1}}{n^{\epsilon/20}}] &< \mathbb{P}[m' > \mathbb{E}[m'] + a((k+3)\ln n)^{k+1} \frac{ntp^{k+1}}{n^{\epsilon/16}}] \\
&\le bn^k e^{-(k+3)\ln n} \\
&< n^{-2}.
\end{aligned}$$

$\square$

To prove Claim 7, we will apply the following theorem of Alon, Kim, and Spencer [4]:

Theorem 9. *Let $X_1, \dots, X_n$ be independent random variables with*

$$\mathbb{P}[X_i = 0] = 1 - p_i \text{ and } \mathbb{P}[X_i = 1] = p_i.$$

For $Y = Y(X_1, \dots, X_n)$, suppose that

$$|Y(X_1, \dots, X_{i-1}, 1, X_{i+1}, \dots, X_n) - Y(X_1, \dots, X_{i-1}, 0, X_{i+1}, \dots, X_n)| \le c_i$$

for all $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$, $i = 1, \dots, n$. Then for

$$\sigma^2 = \sum_{i=1}^{n} p_i(1 - p_i)c_i^2$$

and a positive constant α with $\alpha \max_i c_i < 2\sigma^2$,

$$\mathbb{P}[|Y - \mathbb{E}[Y]| > \alpha) \le 2e^{-\frac{\alpha^2}{4\sigma^2}}.$$

Proof of Claim 7. Recall that $p = t^{\frac{\epsilon-1}{k(2k+1)}}$. The random variable m' is determined by the n independent, indicator random variables $\mathbf{I}[v \in V(H')]$. Each of these affects m' by at most $\deg(v) \le tn^{\epsilon/8}$. Set $\alpha = \frac{ntp^{k+1}}{n^{\epsilon/16}}$ and $\sigma^2 = n^{1+\epsilon/4}p(1-p)t^2$. Note that $\alpha tn^{\epsilon/8} \le 2\sigma^2$. Also, because $t \le n^k$,

$$\frac{\alpha^2}{4\sigma^2} = \frac{np^{2k+1}}{16n^{\epsilon/4+\epsilon/8}(1-p)} \ge \frac{np^{2k+1}}{16n^{\epsilon/4+\epsilon/8}} = \frac{nt^{\frac{\epsilon-1}{k}}}{16n^{3\epsilon/8}} \ge \frac{n^{\epsilon}}{16n^{3\epsilon/8}} = n^{5\epsilon/8}/16.$$

Since $\mathbb{E}[m'] \le ntp^{k+1}$, Theorem 9 implies

$$\mathbb{P}[m' > ntp^{k+1} + \frac{ntp^{k+1}}{n^{\epsilon/20}}] < \mathbb{P}[m' > \mathbb{E}[m'] + \frac{ntp^{k+1}}{n^{\epsilon/16}}] \le 2e^{-n^{5\epsilon/8}/16} < n^{-2}.$$

□

3 Upper Bound for Triangle-free Graphs

In this section we prove Theorem 2. We use the results of Bohman-Keevash [6] and Fiz Pontiveros-Griffiths-Morris [10] on the triangle-free graph process: Let G be the maximal graph in which the triangle-free process terminates.

Theorem 10 (Bohman-Keevash, Fiz Pontiveros-Griffiths-Morris)**.** *With high probability, every vertex of G has degree $d \le (1 + o(1))\sqrt{\frac{1}{2}n \ln n}$, and independence number $\alpha \le (1 + o(1))\sqrt{2n \ln n}$.*

Let $r > 0$ be a real parameter to be optimized later. Construct the graph G' from G as follows:

Construction of G': We take the strong graph product of G and $\bar{K}_r$, the empty graph on r vertices: Replace each vertex v of G by a copy C_v of $\bar{K}_r$. Introduce a complete bipartite graph between all the vertices of C_v and C_u if and only if $\{u, v\} \in E(G)$. We obtain the graph G'. Notice that $|V(G')| = N = nr$.

Define the function $f : V(G') \to V(G)$, such that given any $i \in C_u \subset V(G')$, $f(i) = u$. For a set $S \subset V(G')$, define $f(S) = \bigcup_{i \in S}\{f(i)\}$.

Claim 11. *For every $S \subset V(G')$, S is independent only if $f(S)$ is independent in G. Further $|S| \leq r|f(S)|$.*

Proof. Given an independent set $I \subset G'$, consider $i, j \in I$. Clearly, if $f(i) \neq f(j)$, then $f(i), f(j)$ are not adjacent in G, by the construction. Further, if $f(i) = f(j)$, then i, j must belong to some copy of $\bar{K}_r$ in G'. □

Proof of Theorem 2. We shall show that G' is the required graph. By Claim 11,

$$\begin{aligned} i(G') &\leq \sum_{I \subset G : I \text{ ind. set}} 2^{r|I|} \\ &\leq \alpha \binom{n}{\alpha} 2^{r\alpha} \\ &\leq e^{\ln \alpha + \alpha \ln(ne/\alpha) + r\alpha \ln 2}. \end{aligned}$$

To finish the proof, note that

$$\begin{aligned} \ln \alpha + \alpha \ln(ne/\alpha) + r\alpha \ln 2 &= (\frac{\ln n}{2} + r \ln 2 + o(1))\alpha \\ &\leq (\frac{\ln(N/r)}{2} + r \ln 2 + o(1))\sqrt{2(N/r)\ln(N/r)} \end{aligned}$$

where the last line was obtained by substituting the value of α in terms of N and r. Now maximizing the above expression with respect to r, we get that when $r = \frac{1}{2}\log_2 n$,

$$i(G') \leq e^{(1+o(1))2\sqrt{N \ln 2}\ln(N)}.$$

□

4 Acknowledgments

We thank the referee for helpful comments.

References

[1] M. Ajtai, P. Erdős, J. Komlós, and E. Szemerédi, *On Turán's theorem for sparse graphs*, Combinatorica **1** (1981), no. 4, 313–317. MR 647980 (83d:05052)

[2] M. Ajtai, J. Komlós, J. Pintz, J. Spencer, and E. Szemerédi, *Extremal uncrowded hypergraphs*, J. Combin. Theory Ser. A **32** (1982), no. 3, 321–335. MR 657047 (83i:05056)

[3] 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)

[4] Noga Alon, Jeong-Han Kim, and Joel Spencer, *Nearly perfect matchings in regular simple hypergraphs*, Israel J. Math. **100** (1997), 171–187. MR 1469109 (98k:05112)

[5] A. S. Asratian and N. N. Kuzjurin, *On the number of partial Steiner systems*, J. Combin. Des. **8** (2000), no. 5, 347–352. MR 1775787 (2001d:05011)

[6] Tom Bohman and Peter Keevash, *Dynamic concentration of the triangle-free process*, http://arxiv.org/abs/1302.5963 (2013).

[7] Charles J. Colbourn, Dean G. Hoffman, Kevin T. Phelps, Vojtěch Rödl, and Peter M. Winkler, *The number of t-wise balanced designs*, Combinatorica **11** (1991), no. 3, 207–218. MR 1122007 (93b:05014)

[8] Jeff Cooper and Dhruv Mubayi, *Counting independent sets in triangle-free graphs*, Proc. Amer. Math. Soc. (Accepted).

[9] Richard A. Duke, Hanno Lefmann, and Vojtěch Rödl, *On uncrowded hypergraphs*, Random Structures Algorithms **6** (1995), no. 2-3, 209–212. MR 1370956 (96h:05146)

[10] Gonzalo Fiz Pontiveros, Simon Griffiths, and Robert Morris, *The triangle-free process and r(3,k)*, http://arxiv.org/abs/1302.6279 (2013).

[11] David A. Grable and Kevin T. Phelps, *Random methods in design theory: a survey*, J. Combin. Des. **4** (1996), no. 4, 255–273. MR 1391809 (97d:05031)

[12] 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)

[13] Jeong Han Kim and Van H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), no. 3, 417–434.

[14] Michael Molloy and Bruce Reed, *Graph colouring and the probabilistic method*, Algorithms and Combinatorics, vol. 23, Springer-Verlag, Berlin, 2002. MR 1869439 (2003c:05001)

[15] 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)

[16] ______, *On the independence number of sparse graphs*, Random Structures Algorithms **7** (1995), no. 3, 269–271. MR 1369066 (96k:05101)