

Counting Independent Sets in Hypergraphs

JEFF COOPER¹, KUNAL DUTTA^{2†} and DHARUV MUBAYI^{1‡}

¹Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, IL 60607, USA

(e-mail: jcoope8@uic.edu, mubayi@uic.edu)

²Algorithms and Complexity Department, Max Planck Institute for Informatics, Saarbrücken, Germany
(e-mail: kdutta@mpi-inf.mpg.de)

Received 11 October 2013; revised 8 February 2014; first published online 8 April 2014

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

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

independent sets. This improves a recent result of the first and third authors [8]. In particular, it implies that as $n \rightarrow \infty$, every triangle-free graph on n vertices has at least $e^{(c_1 - o(1))\sqrt{n} \ln n}$ independent sets, where $c_1 = \sqrt{\ln 2}/4 = 0.208138\dots$. 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 $c_2 = 2\sqrt{\ln 2} = 1.665109\dots$. 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

$$\exp\left(c_k \frac{n}{t^{1/k}} \ln^{1+1/k} t\right).$$

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

$$\Omega\left(\frac{n}{t^{1/k}} \ln^{1/k} t\right).$$

Both of our lower bounds follow from a more general statement, which applies to hereditary properties of hypergraphs.

2010 *Mathematics subject classification*: Primary 05D40
Secondary 05C65, 05D05

[†] Part of this work was done at Indian Statistical Institute, New Delhi, India.

[‡] Research partially supported by NSF grants DMS 0969092 and 1300138.

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 .

Ajtai, Komlós and Szemerédi [3] 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 analysing their algorithm, the first and third authors [8] recently showed that, for any such graph,

$$i(G) \geq 2^{\frac{1}{2400} \frac{n}{t} \log_2^2 t}. \quad (1.1)$$

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 (1.1), 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 and Keevash [6] and Fiz Pontiveros, Griffiths and Morris [10].

Theorem 1.1. *Let G be a triangle-free graph with n vertices and average degree t . Then*

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

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

$$i(H) \leq \exp \left((1 - o(1)) \frac{\sqrt{n \ln 2 \ln n}}{4} \right).$$

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

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

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,

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

so the constant in the exponent of Theorem 1.1 is within a factor of 8 of the best possible constant.

1.1. Linear hypergraphs

Fix $k \geq 1$. 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

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

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

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

This leads to our second theorem.

Theorem 1.3. Fix $k \geq 1$. 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 ,

$$i(H) \geq \exp \left(c''_k \frac{n}{t^{1/k}} \ln^{1+1/k} t \right). \tag{1.2}$$

Ajtai, Komlós, Pintz, Spencer and Szemerédi [2] 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,

$$i(H) \leq \exp \left(b''_k \frac{n}{t^{1/k}} \ln^{1+1/k} t \right),$$

so (1.2) is tight up to the constant in the exponent.

1.2. Hereditary properties

Colbourn, Hoffman, Phelps, Rödl and Winkler [7] counted the number of partial $S(t, t + 1, n)$ Steiner systems by analysing 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 1.1 and 1.3 both follow from a more general result (Theorem 1.4 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 1.4. *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 such 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

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

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. Ajtai, Erdős, Komlós and Szemerédi [1] asked whether 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 1.4 implies that if there exists c_r such that every K_r -free graph G satisfies $\alpha(G) \geq c_r \frac{n}{t} \ln t$, then

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

2. Lower bounds

Theorems 1.1 and 1.3 follow from the linear case of Theorem 1.4. We will prove Theorem 1.4 for linear hypergraphs and afterwards describe the changes needed for nonlinear hypergraphs.

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

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 $0 \leq t \leq np$,

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

Set-up. 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 set-up.

Claim 2.1. *If H is linear and $p = t^{-1/(k+1)}$, 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 2.2. *If $p = t^{(\epsilon-1)/(k(2k+1))}$, 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 1.4 (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 implicitly 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 $p = t^{-1/(k+1)}$. 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 $\epsilon < 4/(k + 1)$, we have

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

By the Chernoff bound,

$$\mathbb{P} \left[|n' - np| > \frac{np}{n^{\epsilon/20}} \right] \leq 2e^{-np/3n^{\epsilon/20}} < n^{-2}. \tag{2.1}$$

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 2.1,

$$\mathbb{P} [m' > (1 + \lambda)ntp^{k+1}] < n^{-2}. \tag{2.2}$$

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

$$\begin{aligned} N &> (1 - 1/n)e^{-g \ln p} = (1 - 1/n)\exp\left(\left(1 - 6\lambda\right)\frac{1}{k + 1}\frac{n}{t^{1/k}}f\left(t^{\frac{1}{k+1}}\right)\ln t\right) \\ &> (1 - 1/n)\exp\left(\left(1 - n^{-\epsilon/21}\right)\frac{1}{k + 1}\frac{n}{t^{1/k}}f\left(t^{\frac{1}{k+1}}\right)\ln t\right). \end{aligned} \tag{2.3}$$

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

$$\begin{aligned} \frac{1}{n'} \sum_{v \in V(H')} \deg_{H'}(v) - \frac{1}{n} \sum_{v \in V(H')} \deg_H(v) &\leq \left(\frac{1}{n'} - \frac{1}{n}\right) \sum_{v \in V(H')} \deg_H(v) \\ &\leq \left(\frac{1}{n'} - \frac{1}{n}\right) n' t n^{\epsilon/8} / 2 \\ &= (n - n') \frac{t n^{\epsilon/8}}{2n} \\ &\leq \frac{1}{n} \sum_{v \in K} \deg_H(v), \end{aligned}$$

the average degree of H' is at most

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

Also, because

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

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

$$\begin{aligned} i(H') &\geq (1 - 1/n') \exp\left((1 - 6\lambda) \frac{1}{k + 1} \frac{n'}{t^{1/k}} f(t^{\frac{1}{k+1}}) \ln t\right) \\ &> (1 - 2/n) \exp\left((1 - 6\lambda)(1 - n^{-\epsilon/8}) \frac{1}{k + 1} \frac{n}{t^{1/k}} f(t^{\frac{1}{k+1}}) \ln t\right), \end{aligned}$$

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

$$i(H) \geq i(H') \geq \exp\left((1 - n^{-\epsilon/21}) \frac{1}{k + 1} \frac{n}{t^{1/k}} f(t^{\frac{1}{k+1}}) \ln t\right). \tag{2.4}$$

□

The proof of Theorem 1.4 when H is nonlinear is similar. We set $p = t^{(\epsilon-1)/(k(2k+1))}$. Since we still have $np > n^{\epsilon/4}$, (2.1) still holds. We then use Claim 2.2 instead of Claim 2.1 to prove (2.2). The proof then proceeds in the same way until we get to (2.3), where, using the different value of p , we instead obtain

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

Finally, (2.4) becomes

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

We now prove Theorems 1.1 and 1.3.

Proof of Theorem 1.1. 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 1.4 (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

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

Suppose G is a triangle-free graph with $n < n_0$ vertices and average degree t . Choose an integer r such 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,

$$i(G) = i(G')^{1/r} \geq \left(\exp\left((1 - (rn)^{-1/12}) \frac{1}{2} \frac{rn}{t} \ln t \left(\frac{1}{2} \ln(t) - 1 \right) \right) \right)^{1/r} \geq \exp\left((1 - n^{-1/12}) \frac{1}{2} \frac{n}{t} \ln t \left(\frac{1}{2} \ln(t) - 1 \right) \right).$$

This completes the proof of the first bound in Theorem 1.1. 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 neighbourhood of u is an independent set, which contains 2^t independent sets. Therefore, every triangle-free graph has at least

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

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

$$2^{(1-o(1)) \frac{\sqrt{n \ln n}}{4\sqrt{\ln 2}}} = \exp\left((1 - o(1)) \frac{\sqrt{n \ln 2} \ln n}{4} \right)$$

independent sets. □

Proof of Theorem 1.3. 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 1.4 (with $f(t) = c'_k \ln^{1/k} t$) to conclude that for

$$\epsilon = \frac{3}{k + 1} \in \left(0, \frac{4}{k + 1} \right)$$

there exists n_0 such that every $(k + 1)$ -uniform linear hypergraph H with $n \geq n_0$ vertices satisfies

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

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 1.1. □

It only remains to prove the claims stated at the beginning of this section. We first prove Claim 2.1. We will use the following theorem of Kim and Vu [13].

Theorem 2.3. *Suppose F is a hypergraph such that $W = V(F)$ and $|f| \leq s$ 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 $|A| \leq s$, 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 $M_j = \max_{A: |A| \geq j} M_A$ for $j \geq 0$. Then there exist positive constants $a = a(s)$ and $b = b(s)$ such that, for any $\lambda > 0$,

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

Proof of Claim 2.1. Apply Theorem 2.3 with $F = H$ and $\mathbb{P}[z_i = 1] = p = t^{-1/(k+1)}$. Note first that

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

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

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

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

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

Since $t \leq n^k$ and $\epsilon < 4/(k + 1)$, we have $n \geq n^{\epsilon/8} t^{1/(k+1)}$. Further, $n^{\epsilon/8} t^{1/(k+1)} \geq 1$. Therefore $M_0 \leq n$ and $M_1 \leq n^{\epsilon/8} t^{1/(k+1)}$. Theorem 2.3 therefore implies that there exist constants $a = a(k)$ and $b = b(k)$ such that

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

Since $t \leq n^k$ and $\epsilon < 4/(k + 1)$,

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

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

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

□

To prove Claim 2.2, we will apply the following theorem of Alon, Kim and Spencer [4].

Theorem 2.4. *Let X_1, \dots, X_n be independent random variables with*

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

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

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

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 $\alpha \max_i c_i < 2\sigma^2$,

$$\mathbb{P}[|Y - \mathbb{E}[Y]| > \alpha] \leq 2 \exp\left(-\frac{\alpha^2}{4\sigma^2}\right).$$

Proof of Claim 2.2. Recall that $p = t^{(\epsilon-1)/(k(2k+1))}$. 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 $\text{deg}(v) \leq tn^{\epsilon/8}$. Set

$$\alpha = \frac{ntp^{k+1}}{n^{\epsilon/16}} \quad \text{and} \quad \sigma^2 = n^{1+\epsilon/4}p(1 - p)t^2.$$

Note that $\alpha tn^{\epsilon/8} \leq 2\sigma^2$. Also, because $t \leq n^k$,

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

Since $\mathbb{E}[m'] \leq ntp^{k+1}$, Theorem 2.4 implies

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

□

3. Upper bound for triangle-free graphs

In this section we prove Theorem 1.2. We use the results of Bohman and Keevash [6] and Fiz Pontiveros, Griffiths and Morris [10] on the triangle-free graph process. Let G be the maximal graph in which the triangle-free process terminates.

Theorem 3.1 ([6, 10]). *With high probability, every vertex of G has degree*

$$d \leq (1 + o(1))\sqrt{\frac{1}{2}n \ln n},$$

and independence number $\alpha \leq (1 + o(1))\sqrt{2n \ln n}$.

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 $f : V(G') \rightarrow V(G)$, 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 3.2. *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 $i, j \in I$. 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 1.2. We shall show that G' is the required graph. By Claim 3.2,

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

To finish the proof, note that

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

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 $r = \frac{1}{2} \log_2 n$,

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

□

Acknowledgements

We thank the referee for helpful comments.

References

- [1] Ajtai, M., Erdős, P., Komlós, J. and Szemerédi, E. (1981), On Turán's theorem for sparse graphs. *Combinatorica* **1** 313–317.
- [2] Ajtai, M., Komlós, J., Pintz, J., Spencer, J. and Szemerédi, E. (1982) Extremal uncrowded hypergraphs. *J. Combin. Theory Ser. A* **32** 321–335.
- [3] Ajtai, M., Komlós, J. and Szemerédi, E. (1981) A dense infinite Sidon sequence. *Europ. J. Combin.* **2** 1–11.
- [4] Alon, N., Kim, J.-H. and Spencer, J. (1997) Nearly perfect matchings in regular simple hypergraphs. *Israel J. Math.* **100** 171–187.
- [5] Asratian, A. S. and Kuzjurin, N. N. (2000) On the number of partial Steiner systems. *J. Combin. Des.* **8** 347–352.
- [6] Bohman, T. and Keevash, P. (2013) Dynamic concentration of the triangle-free process. [arXiv.1302.5963](https://arxiv.org/abs/1302.5963)
- [7] Colbourn, C. J., Hoffman, D. G., Phelps, K. T., Rödl, V. and Winkler, P. M. (1991) The number of t -wise balanced designs. *Combinatorica* **11** 207–218.
- [8] Cooper, J. and Mubayi, D. Counting independent sets in triangle-free graphs. *Proc. Amer. Math. Soc.*, to appear.
- [9] Duke, R. A., Lefmann, H. and Rödl, V. (1995) On uncrowded hypergraphs. *Random Struct. Alg.* **6** 209–212.
- [10] Fiz Pontiveros, G., Griffiths, S. and Morris, R. (2013) The triangle-free process and $r(3, k)$. [arXiv.1302.6279](https://arxiv.org/abs/1302.6279)
- [11] Grable, D. A. and Phelps, K. T. (1996) Random methods in design theory: A survey. *J. Combin. Des.* **4** 255–273.
- [12] Kim, J. H. (1995) The Ramsey number $R(3, t)$ has order of magnitude $t^2/\log t$. *Random Struct. Alg.* **7** 173–207.
- [13] Kim, J. H. and Vu, V. H. (2000) Concentration of multivariate polynomials and its applications. *Combinatorica* **20** 417–434.
- [14] Molloy, M. and Reed, B. (2002) *Graph Colouring and the Probabilistic Method*, Vol. 23 of *Algorithms and Combinatorics*, Springer.
- [15] Shearer, J. B. (1983) A note on the independence number of triangle-free graphs. *Discrete Math.* **46** 83–87.
- [16] Shearer, J. B. (1995) On the independence number of sparse graphs. *Random Struct. Alg.* **7** 269–271.