

Hypergraphs with independent neighborhoods

Tom Bohman^{1,*} Alan Frieze^{1,†} Dhruv Mubayi^{1,2,‡} Oleg Pikhurko^{1,§}

January 4, 2008

Abstract

For each $k \geq 2$, let $\rho_k \in (0, 1)$ be the largest number such that there exist k -uniform hypergraphs on n vertices with independent neighborhoods and $(\rho_k + o(1))\binom{n}{k}$ edges as $n \rightarrow \infty$. We prove that $\rho_k = 1 - 2 \log k/k + \Theta(\log \log k/k)$ as $k \rightarrow \infty$. This disproves a conjecture of Füredi and the last two authors.

1 Introduction

The *neighborhood* $N(S)$ of a $(k-1)$ -set S in a k -uniform hypergraph (henceforth a k -graph) is the set of vertices v such that $S \cup \{v\}$ is an edge. For $n \geq k \geq 2$, let $f(n, k)$ be the maximum number of edges in a k -graph on n vertices such that all its neighborhoods are *independent* sets (that is, span no edge). Mantel proved in 1907 that $f(n, 2) = \lfloor n^2/4 \rfloor$, and this was the first result in extremal graph theory. Thus the problem of computing $f(n, k)$ is a natural generalization of Mantel's result.

A k -graph is *odd* if it has a vertex partition $X \cup Y$ such that all edges have an odd number of points less than k in Y . It is easy to see that all neighborhoods in an odd k -graph are independent sets. Let $b(n, k)$ be the maximum number of edges in an odd k -graph. Then the previous observation implies that $f(n, k) \geq b(n, k)$. It was conjectured in [8] that there exists some function $n_0(k)$ such that $n > n_0(k)$ implies

$$f(n, k) = b(n, k). \tag{1}$$

There was some evidence for this, as it reduces to Mantel's theorem for $k = 2$, and it was proved for $k = 3$ by Füredi, Pikhurko, and Simonovits [9, 10], thereby settling a conjecture of Mubayi and Rödl [18]. Recently, (1) has also been proved for $k = 4$ [8]. As we will show here, (1) is not that far from the truth for $k = 5$.

¹Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213

²Department of Mathematics, Statistics, and Computer Science, University of Illinois, Chicago, IL 60607

*Research supported in part by NSF grant DMS-0701183

†Research supported in part by NSF grant CCR-0502793

‡Research supported in part by NSF grant DMS 0653946

§Research supported in part by NSF grant DMS-0457512

Since exact results are rare in extremal hypergraph theory, one often studies asymptotics. In this case, we can define $\rho_k = \lim_{n \rightarrow \infty} f(n, k) / \binom{n}{k}$ which is easily shown to exist [12]. Now conjecture (1) implies that $\rho_k = 1/2$ for all even k and $\rho_k \uparrow 1/2$ as $k \rightarrow \infty$ for odd k . Thus a weaker statement than (1) would be that $\rho_k = \lim_{n \rightarrow \infty} b(n, k) / \binom{n}{k}$, and an even weaker statement is that $\rho_k \rightarrow 1/2$ as $k \rightarrow \infty$.

In this paper we show that conjecture (1) is false for all $k \geq 7$, and in fact that $\rho_k \rightarrow 1$. This follows from an old construction of Kim and Roush [16] which gives lower bounds for the Turán problem for complete k -graphs. Thus the small cases shed little light on the behavior of ρ_k .

We are able to obtain rather sharp estimates on the rate at which ρ_k converges to 1:

Theorem 1. *As $k \rightarrow \infty$, we have*

$$1 - \frac{2 \log k}{k} + (1 + o(1)) \frac{\log \log k}{k} \leq \rho_k \leq 1 - \frac{2 \log k}{k} + (5 + o(1)) \frac{\log \log k}{k},$$

where \log denotes the natural logarithm. Furthermore, for $k \geq 7$, we have $\rho_k > 1/2$, hence (1) is false for $k \geq 7$.

This leaves open the cases $k = 5$ and 6 , where we believe that (1) still holds.

Conjecture 1. *$f(n, k) = b(n, k)$ for $k \in \{5, 6\}$ and n sufficiently large.*

We will present the lower bounds in Theorem 1 via constructions in the next section. Sections 3 and 4 are devoted to the proof of the upper bound. We close with some concluding remarks, related open problems, and a sketch of the proof that shows that $40/81 = 0.493... \leq \rho_5 < 0.534$.

We associate a k -graph with its edge set. For a vertex subset S of size $k - 1$, let $d(S) = |N(S)|$. Let $\binom{V}{k} = \{X \subset V : |X| = k\}$. We denote $[n] = \{1, \dots, n\}$. Let $\text{Bin}(k, p)$ denote the binomial distribution with parameters k and p . In Sections 2–4, the asymptotic notation ($O(1)$, $o(1)$, etc) will refer to the case when k is fixed and $n \rightarrow \infty$.

2 Construction

In this section we prove the lower bound in Theorem 1 by means of a construction due to Kim and Roush.

Construction 1 (Kim and Roush [16]). Let $Y_1 \cup \dots \cup Y_l$ be a partition of $[n]$ into sets, each of size $\lfloor n/l \rfloor$ or $\lceil n/l \rceil$. Let the k -graph \mathcal{H} consist of all k -sets that have at least one point in each Y_i . Partition \mathcal{H} into $\mathcal{H}_1 \cup \dots \cup \mathcal{H}_l$, where

$$\mathcal{H}_j = \left\{ S \in \mathcal{H} : \sum_{i=1}^l i |S \cap Y_i| \equiv j \pmod{l} \right\}.$$

By the Pigeonhole Principle, we may assume that there is an $a \in [l]$ with $|\mathcal{H}_a| \leq |\mathcal{H}|/l$. Now let

$$\mathcal{F} = \mathcal{H} \setminus \mathcal{H}_a.$$

□

Proposition 1. *For any $\delta > 0$ there is a $k_0 = k_0(\delta)$ such that for all $k \geq k_0$ and all sufficiently large n (i.e. $n > n_0(k, \delta)$), Construction 1 produces a k -graph \mathcal{F} on n vertices with independent neighborhoods such that*

$$|\mathcal{F}| > \left(1 - \frac{2 \log k}{k} + (1 - \delta) \frac{\log \log k}{k}\right) \binom{n}{k}.$$

Proof. To see that \mathcal{F} has independent neighborhoods, consider a $(k-1)$ -set S . Then $N(S)$ cannot have a point in each Y_i for then $\left\{ \sum_{i=1}^l i |(S \cup \{v\}) \cap Y_i| : v \in N(S) \right\}$ covers all congruence classes modulo l . But then $N(S)$ is an independent set, since every edge of \mathcal{F} has a point in each Y_i .

Let $k > k_0(\delta)$ be fixed and $n \rightarrow \infty$. If l is a fixed function of k then we have

$$\begin{aligned} |\mathcal{F}| &\geq \left(1 - \frac{1}{l}\right) \left(\binom{n}{k} - l \binom{n - \lfloor n/l \rfloor}{k} \right) \\ &\geq \left[\left(1 - \frac{1}{l}\right) \left(1 - l(1 - 1/l)^k\right) + \Theta\left(\frac{1}{n}\right) \right] \binom{n}{k} \\ &= \left(1 - \frac{1}{l} - (l-1)(1 - 1/l)^k + \Theta\left(\frac{1}{n}\right)\right) \binom{n}{k}. \end{aligned}$$

Set $l = \lceil k / ((2 - \epsilon) \log k) \rceil$, where $\epsilon = \log \log k / \log k$. Then using $(1 - 1/l)^k < e^{-k/l}$ and $k^\epsilon = \log k$, we obtain

$$|\mathcal{F}| \geq \left(1 - \frac{(2 - \epsilon) \log k}{k} - \frac{1}{k} + \Theta\left(\frac{1}{n}\right)\right) \binom{n}{k}.$$

This gives the required bound. □

Proposition 2. *For any $k \geq 7$, we have $\rho_k > 1/2$.*

Proof. Let us take $l = 3$ in Construction 1. The Inclusion-Exclusion Principle shows that $|\mathcal{H}| / \binom{n}{k} = 1 - 3 \cdot (2/3)^k + 3 \cdot (1/3)^k + o(1)$. The right-hand side assumes value $\frac{602}{729} > \frac{3}{4}$ for $k = 7$ and, as it is not hard to show, is an increasing function of $k \geq 7$. Since \mathcal{F} contains at least $2/3$ edges of \mathcal{H} , the proposition follows. □

3 Lemmas

This section contains some auxiliary results needed in the proof of the upper bound of Theorem 1.

Lemma 1. For every k there is an n_0 such that for all n , x , and y with $x + y = n \geq n_0$ and $\frac{4n}{k-1} \leq y \leq \frac{n}{2}$, we have

$$\max_{0 \leq i \leq k-1} \frac{\binom{x}{i} \binom{y}{k-i-1}}{\binom{n}{k-1}} \leq 4 \left(\frac{n}{ky} \right)^{1/2}.$$

Proof. Let $n_0 = n_0(k)$ be sufficiently large. Let $p = x/n$ and $q = y/n = 1-p$. For $0 \leq i \leq k-1$, let $p_i = \binom{x}{i} \binom{y}{k-i-1} \binom{n}{k-1}^{-1}$ and $b_i = \binom{k-1}{i} p^i q^{k-1-i}$. We begin by noting that the hypergeometric distribution (as given by p_i) can be bounded by the binomial distribution (as given by b_i). Consider an experiment in which we choose $k-1$ elements of $[n]$ uniformly at random with replacement. Let $X \subset [n]$ with $|X| = x$, and let \mathcal{D} be the event that the $k-1$ random choices are distinct. Note that b_i is the probability that exactly i of our randomly chosen elements fall in X and p_i is the probability that exactly i of our randomly chosen elements fall in X when we condition on \mathcal{D} . Therefore,

$$p_i \leq \frac{b_i}{Pr(\mathcal{D})} \leq \frac{b_i}{1 - \binom{k-1}{2} \frac{1}{n}}. \quad (2)$$

Note that $b_i < b_{i+1}$ if and only if $i < (k-1)p - (1-p)$. Therefore, if we set $i_0 = \lfloor (k-1)p \rfloor$ and $i_1 = i_0 + 1$ then $\max_i b_i = \max\{b_{i_0}, b_{i_1}\}$. By assumptions on y we have $i_0 \geq 1$, so we can apply an estimate given in Bollobás [2, Inequality (1.5)], giving

$$\begin{aligned} b_{i_0} &\leq \left(\frac{(k-1)}{2\pi i_0 (k-1-i_0)} \right)^{1/2} \left(\frac{(k-1)p}{i_0} \right)^{i_0} \left(\frac{(k-1)q}{k-1-i_0} \right)^{k-1-i_0} \\ &\leq \frac{4}{3} \left(\frac{n}{2\pi ky} \right)^{1/2} \times e^2, \end{aligned}$$

where we used the estimates $\frac{i_0}{k-1} \geq p - \frac{1}{k-1} \geq \frac{3}{4}p$, $\left(\frac{(k-1)p}{i_0} \right)^{i_0} \leq \left(\frac{i_0+1}{i_0} \right)^{i_0} \leq e$, etc. A similar upper bound holds for b_{i_1} . Now the lemma follows from (2) since $\frac{4}{3}e^2/\sqrt{2\pi} < 4$. \square

Lemma 2. For every k there is an n_0 such that for all $n \geq n_0$ the following holds. Suppose that we have two families \mathcal{F} and \mathcal{G} of k -subsets and $(k-1)$ -subsets of $[n]$, respectively, such that $|\mathcal{F}| \geq (1-f) \binom{n}{k}$ and $|\mathcal{G}| \geq g \binom{n}{k-1}$. Let $[n] = X \cup Y$ with $x = |X|$ and $y = |Y|$ satisfying $\frac{4n}{k-1} \leq y \leq \frac{n}{2}$. Suppose that reals $0 < f', g' < 1$ satisfy

$$g'f + f'g > f + f'g' + 4f'\sqrt{n/ky}. \quad (3)$$

Then there is an i , $0 \leq i \leq k-1$, with

$$|\mathcal{F}_i| = |\{K \in \mathcal{F} : |K \cap X| = i\}| \geq (1-f') \binom{x}{i} \binom{y}{k-i} \quad (4)$$

and

$$|\mathcal{G}_i| = |\{L \in \mathcal{G} : |L \cap X| = i\}| \geq g' \binom{x}{i} \binom{y}{k-i-1}. \quad (5)$$

Proof. Suppose on the contrary that no such i exists. Consider

$$s = \frac{(1-g')}{\binom{n}{k}} |\mathcal{F}| + \frac{f'}{\binom{n}{k-1}} |\mathcal{G}| \geq (1-g')(1-f) + f'g. \quad (6)$$

Since either \mathcal{F}_i or \mathcal{G}_i is small (as defined by (4),(5)) for each i , we have $s \leq \sum_{i=0}^k \max(a_i, b_i)$, where

$$\begin{aligned} a_i &= (1-g')(1-f') \frac{\binom{x}{i} \binom{y}{k-i}}{\binom{n}{k}} + f' \frac{\binom{x}{i} \binom{y}{k-i-1}}{\binom{n}{k-1}} \\ b_i &= (1-g') \frac{\binom{x}{i} \binom{y}{k-i}}{\binom{n}{k}} + f'g' \frac{\binom{x}{i} \binom{y}{k-i-1}}{\binom{n}{k-1}}. \end{aligned}$$

Since

$$a_i - b_i = \frac{\binom{x}{i} \binom{y}{k-i}}{\binom{n}{k}} \times (1-g')f' \times \left(-1 + \frac{(n-k+1)(k-i)}{k(y-k+i+1)} \right),$$

there is an i_0 such that $a_i \geq b_i$ for $0 \leq i < i_0$ and $a_i \leq b_i$ for $i_0 \leq i \leq k$. Hence,

$$s \leq \sum_{i=0}^{i_0-1} a_i + \sum_{i=i_0}^k b_i.$$

Let $P = \binom{n}{k}^{-1} \sum_{i=0}^{i_0-1} \binom{x}{i} \binom{y}{k-i}$ and $P' = \binom{n}{k-1}^{-1} \sum_{i=0}^{i_0-1} \binom{x}{i} \binom{y}{k-i-1}$. Let us choose a random $(k-1)$ -subset L of $[n]$ and then let K be obtained from L by adding a random vertex $x \notin L$. Then K is also uniformly distributed. Note that P (resp. P') is the probability that K (resp. L) has less than i_0 vertices in X . Since $L \subset K$, $P \leq P'$. On the other hand $P' - P$ is exactly the probability that $x \in X$ and $|L \cap X| = i_0 - 1$. It follows from Lemma 1 that $\Pr(|L \cap X| = i_0 - 1) \leq 4\sqrt{n/ky}$ and so $P' - P \leq 4\sqrt{n/ky}$. Hence,

$$\begin{aligned} s &\leq P(1-g')(1-f') + P'f' + (1-P)(1-g') + (1-P')f'g' \\ &\leq P(1-g')(1-f') + P'f' + (1-P)(1-g') + (1-P)f'g' + 4f'\sqrt{n/ky} \\ &= 1-g' + f'g' + 4f'\sqrt{n/ky}. \end{aligned}$$

From (6), we obtain that

$$(1-g')(1-f) + f'g \leq s \leq 1-g' + f'g' + 4f'\sqrt{n/ky},$$

and this contradicts (3). □

4 The Upper Bound on ρ_k

Take small $\delta > 0$. Let $k \geq k_0(\delta)$ be sufficiently large. Choose large $n_0 = n_0(k, \delta)$. With foresight, we define

$$c_0 = 4 + \delta \quad c_1 = 5 + 2\delta \quad c_2 = 5 + 3\delta \quad c_3 = 5 + 6\delta.$$

For brevity of notation, let $\epsilon = \log \log k / \log k$. We will show that for all $k > k_0$ we have

$$\rho_k < 1 - \frac{(2 - (5 + 7\delta)\epsilon) \log k}{k} = 1 - \frac{2 \log k}{k} + (5 + 7\delta) \frac{\log \log k}{k}.$$

Suppose that this is false for some $k > k_0$. Then for infinitely many n , in particular for some $n > n_0(k, \delta)$, we can find a k -graph \mathcal{F} with vertex set $[n]$ and independent neighborhoods such that

$$|\mathcal{F}| > \left(1 - \frac{(2 - c_3\epsilon) \log k}{k}\right) \binom{n}{k}.$$

Define

$$l = \left\lceil \frac{k}{(\log k)^{c_0}} \right\rceil.$$

Our goal is to find sets $A_1, \dots, A_l, B_1, \dots, B_l \subset [n]$ such that the following conditions hold.

Condition 1: For every $i \in [l]$, the set A_i is independent (with respect to \mathcal{F}), is disjoint from B_i , and has size

$$a = \left\lceil \left(1 - \frac{(2 - c_1\epsilon) \log k}{k}\right) n \right\rceil \quad (7)$$

Condition 2: The sets B_1, \dots, B_l are pairwise disjoint, each of size

$$b = \left\lceil \frac{(2 - c_2\epsilon) \log k}{k} n \right\rceil. \quad (8)$$

Indeed, if we have such sets then, for any $1 \leq i < j \leq l$, the set $A_i \cap A_j$ has at most $n - 2b$ elements because its complement contains $B_i \cup B_j$ as a subset. Since every k -set in $\cup_{i=1}^l \binom{A_i}{k}$ is missing from \mathcal{F} , we have by a simple version of the Inclusion-Exclusion Principle that

$$l \binom{a}{k} - \binom{l}{2} \binom{n - 2b}{k} \leq \binom{n}{k} - |\mathcal{F}| < \frac{2 \log k}{k} \binom{n}{k}.$$

Dividing by $\binom{n}{k}$ and using $k > k_0(\delta)$ and $n > n_0(k, \delta)$, we get

$$(1 - \delta) \left(\frac{l}{k^{2 - c_1\epsilon}} - \frac{l^2}{2k^{4 - 2c_2\epsilon}} \right) \leq \frac{2 \log k}{k},$$

which is a contradiction (for $\delta < 1$ and $k \geq k_0(\delta)$).

Before proceeding with an argument that gives the sets $A_1, \dots, A_l, B_1, \dots, B_l$, we need two observations regarding $(k - 1)$ -sets of large degree. First, observe that for every $(k - 1)$ -set S , we have

$$d(S) < \left(1 - \frac{\log k - 2 \log \log k}{k}\right) n, \quad (9)$$

for otherwise $\binom{n}{k} - |\mathcal{F}| \geq \binom{d(S)}{k} > \frac{1}{2} \frac{\log^2 k}{k} \binom{n}{k}$ which is a contradiction. Let \mathcal{G} be the collection of $(k - 1)$ -sets $S \in \binom{[n]}{k-1}$ such that $d(S) \geq n - b$.

Claim 1. $|\mathcal{G}| \geq 2\delta\epsilon \binom{n}{k-1}$.

Proof of Claim. Let $|\mathcal{G}| = g \binom{n}{k-1}$. We have

$$k \left(1 - \frac{(2 - c_3 \epsilon) \log k}{k}\right) \binom{n}{k} \leq k |\mathcal{F}| = \sum_{S \in \binom{[n]}{k-1}} d(S) \leq \binom{n}{k-1} (1 - g) \left(1 - \frac{(2 - c_2 \epsilon) \log k}{k}\right) n + g \binom{n}{k-1} \left(1 - \frac{\log k - 2 \log \log k}{k}\right) n,$$

where the last expression comes from (9). Solving for g yields

$$g \geq \frac{(c_3 - c_2) \epsilon - 2k^2/n}{1 - c_2 \epsilon + 2 \log \log k / \log k} > 2\delta \epsilon.$$

(We used the facts that $c_3 - c_2 = 3\delta$ and $c_2 > 2$ in the last inequality.) This completes the proof of Claim 1. \square

Now we describe how to inductively construct the sets A_i and B_i . Suppose that we have constructed $A_1, \dots, A_p, B_1, \dots, B_p$ with $0 \leq p < l$ satisfying Conditions 1 and 2. Let

$$y = \left\lfloor \frac{2n}{(\log k)^{c_0-1}} \right\rfloor. \quad (10)$$

and $x = n - y$. Take an arbitrary partition $[n] = X \cup Y$ with $Y \supset \cup_{j=1}^p ([n] \setminus A_j)$ and $|Y| = y$, which is possible since each set $[n] \setminus A_i$ has $n - a \leq 2n \log k / k$ elements and $p < l$. Our task now is to construct A_{p+1} and B_{p+1} .

For an integer i , define

$$\mathcal{F}_i = \{S \in \mathcal{F} : |S \cap X| = i\} \quad \text{and} \quad \mathcal{G}_i = \{S \in \mathcal{G} : |S \cap X| = i\}.$$

Also, let

$$f = 2 \log k / k, \quad g = 2\delta \epsilon, \quad f' = \log^{2+\delta} k / k, \quad g' = \delta \epsilon.$$

A short calculation shows by (10) that (3) holds:

$$f'(g - g') + (g'f - f - 4f'\sqrt{n/ky}) > \frac{\delta \epsilon \log^{2+\delta} k}{k} - C \frac{\log k}{k} > 0,$$

for some absolute constant C . So Lemma 2 implies that there is an i such that $|\mathcal{F}_i| \geq (1 - f') \binom{x}{i} \binom{y}{k-i}$ and $|\mathcal{G}_i| \geq \delta \epsilon \binom{x}{i} \binom{y}{k-i-1}$.

Let $\lambda = f' / (\delta \epsilon)$. Let us show that there is a $(k-1)$ -set $T_0 \in \mathcal{G}_i$ such that

$$|Y \setminus N(T_0)| \leq \lambda(y - k + i + 1) \quad (11)$$

Suppose on the contrary that no such T_0 exists. Then counting the number of pairs (K, z) with $K \in \mathcal{F}_i$ and $z \in K \cap Y$ in two different ways, we obtain

$$(k-i)(1-f') \binom{x}{i} \binom{y}{k-i} \leq (k-i) |\mathcal{F}_i| < |\mathcal{G}_i| (1-\lambda)(y-k+i+1) + \left(\binom{x}{i} \binom{y}{k-1-i} - |\mathcal{G}_i| \right) (y-k+i+1).$$

It follows that

$$\lambda|\mathcal{G}|(y - k + i + 1) < f' \binom{x}{i} \binom{y}{k-i} (k-i).$$

Since $|\mathcal{G}| \geq \delta \epsilon \binom{x}{i} \binom{y}{k-i+1}$, this contradicts the choice of λ .

Choose an arbitrary set $B_{p+1} \subset X$ that contains all of $X \setminus N(T_0)$ and such that $|B_{p+1}| = b$. (This is possible because $|X| \geq n - lb \geq b$ and $T_0 \in \mathcal{G}$, so $|X \setminus N(T_0)| \leq n - d(T_0) \leq b$.) For every $j \in [p]$, the set $B_j \subset Y$ is disjoint from $B_{p+1} \subset X$, so Condition 2 holds. Let $Z = Y \setminus N(T_0)$ and $A' = [n] \setminus (B_{p+1} \cup Z)$. Note that A' , as a subset of $N(T_0)$, is an independent set. Moreover, by the definition of T_0 (i.e. by (11)), we have

$$|A'| \geq n - b - \lambda y \geq n - \left\lceil \frac{(2 - c_2 \epsilon) \log k}{k} n \right\rceil - \frac{\log^{2+\delta} k}{\delta \epsilon k} \times \frac{2n}{(\log k)^{c_0-1}} \geq a.$$

Let us take for A_{p+1} an arbitrary a -subset of A' . Condition 1 clearly holds, finishing the proof. \square

5 Concluding Remarks and Open Problems

• By suitably modifying the proof that $\rho_4 = 1/2$ from [8], we can obtain fairly good bounds for ρ_5 . Here is a sketch of the proof. Suppose that \mathcal{G} is a 5-graph with independent neighborhoods and $\pi \binom{n}{5}$ edges. By Zykov symmetrization, we may assume that all vertex degrees are $(\pi + o(1)) \binom{n}{4}$. Now let A be a neighborhood of maximum size, say $|A| = \alpha n$, and $B = [n] \setminus A$. Let h_i be the number of edges of \mathcal{G} with exactly i points in B ; note that $h_0 = 0$ by our hypothesis. Let σ_i be the sum, over all 4-sets S with i points in B and $4 - i$ points in A , of $d(S)$. Then one obtains

$$\binom{\alpha n}{3} (1 - \alpha)n \times \alpha n \geq \sigma_1 = 4h_1 + 2h_2 \tag{12}$$

$$\binom{\alpha n}{2} \binom{(1 - \alpha)n}{2} \times \alpha n \geq \sigma_2 = 3h_2 + 3h_3 \tag{13}$$

$$\alpha n \binom{(1 - \alpha)n}{3} \times \alpha n \geq \sigma_3 = 2h_3 + 4h_4. \tag{14}$$

On the other hand, using the fact that all degrees are almost equal we get

$$(1 - \alpha)n \times (\pi + o(1)) \binom{n}{4} = \sum_{x \in B} d(x) = h_1 + 2h_2 + 3h_3 + 4h_4 + 5h_5. \tag{15}$$

Now consider $3/4 \times (12) + 1/6 \times (13) + 1/4 \times (14) + (15)$, observe that $\sum_{i=1}^5 h_i = |\mathcal{G}|$, and divide by n^5 . This gives that, as $n \rightarrow \infty$,

$$\pi \leq \frac{\alpha}{5\alpha - 1} (15(1 - \alpha)\alpha^3 + 5(1 - \alpha)^2\alpha^2 + 5(1 - \alpha)^3\alpha) + o(1).$$

Maximizing this function over all $\alpha \in (0.5, 1)$ yields $\pi < 0.534$ and hence $\rho_5 < 0.534$. On the other hand, we have $b(n, 5) = (\frac{40}{81} + o(1))\binom{n}{5}$ (take $|Y| = (\frac{1}{3} + o(1))n$ for the lower bound), which shows that $\rho_5 \geq \frac{40}{81} > 0.493$.

- Our results are similar in flavor to the following problem about the Turán numbers of complete hypergraphs. Let t_k denote the maximum proportion of edges in a k -graph on n vertices, as $n \rightarrow \infty$, that contains no copy of the complete k -graph on $k + 1$ vertices. Thus $t_2 = 1/2$ by Mantel's theorem. The most famous conjecture in this area, due to Turán [23], is that $t_3 = 5/9$, which is achieved by (among others) the 3-graph with vertex partition Y_1, Y_2, Y_3 into almost equal parts and all edges with two points in Y_i and one point in Y_{i+1} (indices modulo 3) or one point in each Y_i . Perhaps just as interesting is to determine the growth rate of t_k as $k \rightarrow \infty$. Frankl and Rödl [7] proved that $1 - t_k = O(\log k/k)$ via a construction that has similarities to Construction 1 in this paper. On the other hand, the known upper bound is $t_k = 1 - \Omega(1/k)$, where the best results are due to Chung and Lu [3]. It would be very interesting to obtain sharper estimates for t_k . Perhaps the methods of this paper can be used to show that $1 - t_k = \omega(1/k)$, an open question for whose solution de Caen [4, Page 190] offered 500 Canadian dollars.

- For $2 \leq m \leq k$ let the *book* $B_{k,m}$ be the k -graph with the following $m + 1$ edges: $[k - 1] \cup \{k + i - 1\}$ for $i \in [m]$, and $\{k, k + 1, \dots, 2k - 1\}$. The problem of computing the Turán function $\text{ex}(n, B_{k,m})$ has been actively studied [1, 5, 6, 8, 9, 10, 14, 18, 20, 21]. Clearly, the property not containing $B_{k,k}$ as a subgraph is equivalent to having empty neighborhoods, so $f(n, k) = \text{ex}(n, B_{k,k})$. Our results can be modified to show, for example, that for any function $m = m(k) < c_1 \log k$, where c_1 is a constant, we have

$$\pi(B_{k,k-m}) = 1 - \Theta\left(\frac{\log k}{k}\right) \tag{16}$$

as $k \rightarrow \infty$, where $\pi(F) = \lim_{n \rightarrow \infty} \text{ex}(n, F)/\binom{n}{k}$ denotes the *Turán density* of a k -graph F . Indeed, the upper bound on $\pi(B_{k,k-m})$ follows from Theorem 1 and the trivial observation that $\text{ex}(n, B_{k,k}) \geq \text{ex}(n, B_{k,k-m})$. The lower bound (16) can be obtained by taking the k -graph \mathcal{F} of Construction 1 with $l = k/c_2 \log k$ where $c_2 \gg \max(c_1, 1)$ and removing those edges of \mathcal{F} that intersect some part Y_i in at most m vertices. As $n \rightarrow \infty$, the proportion of edges that we delete is approximately at most

$$l \times \Pr(\text{Bin}(k, 1/l) < m) \leq l e^{-c_2 \log k/4} < \frac{1}{k^2}.$$

(We apply the Chernoff bound here, see e.g. [11, Corollary 2.3].) Therefore, the size of the family \mathcal{F} is at least $(1 - 1/l)(1 - 1/k^2)\binom{n}{k}$, and (16) follows.

On the other hand, it is easy to show that $\pi(B_{k,m}) = o(1)$ if $m = o(k)$. Determining the behavior of $\pi(B_{k,m})$ for the intermediate values of m is an interesting open problem.

- A related problem which has been studied a fair amount recently (see, e.g., [13, 15, 17, 22]) is the maximum possible minimum degree (of $(k - 1)$ -sets) that a k -graph can have without containing some fixed configuration. Let $g(n, k)$ denote the maximum minimum degree of a k -graph on n vertices with independent neighborhoods. Then it was shown in [19] that the limit $\gamma_k = \lim_{n \rightarrow \infty} g(n, k)/n$ exists. It is trivial to see that $\gamma_k \leq 1/2$ for all k , and odd k -graphs show that if k is even, we have equality. It would be interesting to determine the behavior of γ_k for k odd. As with t_k , the small cases seem difficult. For $k = 3$, the construction for t_3 above minus the edges with one point in each Y_i shows that $\gamma_3 \geq 1/3$. In fact, we make the following conjecture.

Conjecture 2. *For every $\epsilon > 0$, there exists n_0 such that if $n > n_0$ and \mathcal{G} is an n -vertex 3-graph with every pair lying in at least $(1/3 + \epsilon)n$ edges, then \mathcal{G} contains a neighborhood that is not an independent set. In particular, $\gamma_3 = 1/3$.*

- Construction 1 has the following generalization. We begin with some definitions that establish the general setting. Let $a, l \geq 2$ be fixed parameters. Consider the digraph D with vertex set \mathbb{Z}_a^l and an arc from $x = (x_1, \dots, x_l)$ to $y = (y_1, \dots, y_l)$ if and only if there exists a coordinate k such that

$$y_i = \begin{cases} x_i & \text{if } i \neq k \\ x_i - 1 & \text{if } i = k \end{cases}.$$

Note that the out-degree of each vertex is l . We say that a subset X of \mathbb{Z}_a^l is a *perfect cover* of D if the out-neighborhoods of the elements of X form a partition of \mathbb{Z}_a^l . In other words, the set X is a perfect cover if for every $y \in \mathbb{Z}_a^l$ there exists a unique $x \in X$ such that the arc (x, y) (i.e. the arc directed from x to y) is in D . Note that a perfect cover contains a^l/l vertices.

Suppose X is a perfect cover of D . Let n be large and fix a partition Y_1, \dots, Y_l of $[n]$. For each k -set S let $y_S \in \mathbb{Z}_a^l$ be the vector $y_S = (y_1, \dots, y_l)$ where $y_i \equiv |S \cap Y_i| \pmod{a}$ for $i = 1, \dots, l$. Now we are ready to define our family with independent neighborhoods. Let \mathcal{F} be the collection of k -sets S such that $S \cap Y_i \neq \emptyset$ for $i = 1, \dots, l$ and $y_S \notin X$. We claim that the collection \mathcal{F} has independent neighborhoods. To see this, consider a $(k - 1)$ -set T . Since X is a perfect cover, there exists $x \in X$ such that (x, y_T) is an arc in D . It follows that there exists an index k such that $T \cup \{z\} \notin \mathcal{F}$ for all $z \in Y_k$. In other words, the neighborhood of T (in the hypergraph \mathcal{F}) does not intersect Y_k . Since every edge in \mathcal{F} intersects Y_k , it follows that \mathcal{F} has independent neighborhoods.

In order to ensure a lower bound on the cardinality of the collection \mathcal{F} , we consider situations where there is a partition of \mathbb{Z}_a^l into perfect covers X_1, \dots, X_l . Each X_i corresponds to a collection \mathcal{F}_i . Furthermore, each set S that intersects Y_1, \dots, Y_l is excluded from exactly one of the collections \mathcal{F}_i . Therefore, there is an index i such that $|\mathcal{F}_i|$ is at least $(1 - 1/l)$ times the number of k -sets S that intersect Y_1, \dots, Y_l .

Note that Construction 1 is given by this general setting by taking $a = l$ and letting

$$X_j = \left\{ x \in \mathbb{Z}_l^l : \sum_{i=1}^l ix_i = j \right\}.$$

For a second example, set $a = 2$ and suppose $l = 2^b$ for some integer $b \geq 2$. Fix a Hamming code $H \subseteq \{0, 1\}^{l-1}$; that is, fix a set of strings $H \subseteq \{0, 1\}^{l-1}$ with the property that every string in $\{0, 1\}^{l-1}$ is either in H or adjacent (in the $(l-1)$ -cube) to exactly one element of H . Note that

$$X = \{(x_1, \dots, x_l) \in \mathbb{Z}_2^l : (x_1, \dots, x_{l-1}) \in H\}$$

is a perfect cover of \mathbb{Z}_2^l . Furthermore the collection $X, X + e_1, X + e_2, \dots, X + e_{l-1}$ is a partition of \mathbb{Z}_2^l into perfect covers. Thus, the Hamming code gives another construction that achieves the bound given by Construction 1.

References

- [1] B. Bollobás, Three-graphs without two triples whose symmetric difference is contained in a third. *Discrete Math.*, 8 (1974) 21–24.
- [2] B. Bollobás, *Random Graphs*, 2nd Edition, Cambridge University Press, 2001.
- [3] F. Chung and L. Lu, An upper bound for the Turán number $t_3(n, 4)$, *J. Combin. Theory Ser. A*, 87 (1999), no. 2, 381–389.
- [4] D. de Caen, *The current status of Turán’s problem on hypergraphs*, Extremal problems for finite sets (Visegrád, 1991), Bolyai Soc. Math. Stud., vol. 3, János Bolyai Math. Soc., Budapest, 1994, pp. 187–197.
- [5] P. Frankl and Z. Füredi. A new generalization of the Erdős-Ko-Rado theorem. *Combinatorica*, 3 (1983) 341–349.
- [6] P. Frankl and Z. Füredi. Extremal problems whose solutions are the blowups of the small Witt-designs. *J. Combin. Th. A*, 52 (1989) 129–147.
- [7] P. Frankl and V. Rödl, Lower bounds for Turán’s problem, *Graphs Combin.* 1 (1985), no. 3, 213–216.
- [8] Z. Füredi, D. Mubayi and O. Pikhurko, Quadruple systems with independent neighborhoods, submitted
- [9] Z. Füredi, O. Pikhurko and M. Simonovits, On triple systems with independent neighborhoods, *Combin. Probab. Comput.* 14 (2005), 795–813.
- [10] Z. Füredi, O. Pikhurko and M. Simonovits, The Turán density of the hypergraph $\{abc, ade, bde, cde\}$. *Electron. J. Combin.* 10 (2003), Research Paper 18, 7 pp. (electronic).

- [11] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*. Wiley, 2000.
- [12] G. Katona, T. Nemetz and M. Simonovits, On a problem of Turán in the theory of graphs, *Mat. Lapok* 15 (1964) 228–238.
- [13] P. Keevash, A hypergraph regularity method for generalised Turán problems, manuscript, 2007.
- [14] P. Keevash and D. Mubayi. Stability results for cancellative hypergraphs. *J. Combin. Th. B*, 92 (2004) 163–175.
- [15] P. Keevash and Y. Zhao, Codegree problems for projective geometries, *J. Combin. Theory Ser. B*, 97 (2007), 919–928.
- [16] K. Kim and F. Roush, On a problem of Turán, *Studies in pure mathematics*, 423-425, Birkhäuser, Basel-Boston, Mass. 1983.
- [17] D. Mubayi, The co-degree density of the Fano plane, *J. Combin. Theory Ser. B*, 95 (2005), no. 2, 333–337.
- [18] D. Mubayi and V. Rödl, On the Turán number of triple systems, *J. Combin. Theory Ser. A*, 100 (2002), no. 1, 136–152.
- [19] D. Mubayi and Y. Zhao, Codegree density of hypergraphs, *J. Combin. Theory Ser. A*, to appear.
- [20] O. Pikhurko. An exact Turán result for the generalized triangle. *Combinatorica*, to appear.
- [21] A. F. Sidorenko. The maximal number of edges in a homogeneous hypergraph containing no prohibited subgraphs. *Math Notes*, 41 (1987) 247–259. Translated from *Mat. Zametki*.
- [22] V. Rödl, A. Ruciński and E. Szemerédi, A Dirac-type theorem for 3-uniform hypergraphs, *Combinatorics, Probability and Computing*, to appear
- [23] P. Turán. On an extremal problem in graph theory (in Hungarian). *Mat. Fiz. Lapok*, 48 (1941) 436–452.