A hypergraph extension of Turán's theorem

Dhruv Mubayi *

June 6, 2005

Abstract

Fix $l \geq r \geq 2$. Let $H_{l+1}^{(r)}$ be the r-uniform hypergraph obtained from the complete graph K_{l+1} by enlarging each edge with a new set of $r-2$ vertices. Thus $H_{l+1}^{(r)}$ has $(r-2)\binom{l+1}{2}+l+1$ vertices and $\binom{l+1}{2}$ edges. We prove that the maximum number of edges in an n-vertex r-uniform hypergraph containing no copy of $H_{l+1}^{(r)}$ is

$$\frac{(l)_r}{l^r}\binom{n}{r}+o(n^r)$$

as $n \to \infty$. This is the first infinite family of irreducible r-uniform hypergraphs for each odd $r > 2$ whose Turán density is determined.

Along the way we give three proofs of a hypergraph generalization of Turán's theorem. We also prove a stability theorem for hypergraphs, analogous to the Simonovits stability theorem for complete graphs.

1 Introduction

Given a family $\mathcal{F}$ of r-uniform hypergraphs (r-graphs for short), and an r-graph G, we say that G is $\mathcal{F}$-free if G contains no member of $\mathcal{F}$ as a subhypergraph. The extremal number $\mathrm{ex}(n, \mathcal{F})$ is the maximum number of edges in an $\mathcal{F}$-free n-vertex r-graph (in case $\mathcal{F}$ is a single r-graph F, we write $\mathrm{ex}(n, F)$ instead of $\mathrm{ex}(n, \{F\})$). The Turán density of F is defined as

$$\pi(F) := \lim_{n\to\infty} \frac{\mathrm{ex}(n, F)}{\binom{n}{r}}.$$

*Department of Mathematics, Statistics, & Computer Science, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, IL 60607-7045; email: mubayi@math.uic.edu. Research partially supported by National Science Foundation Grant DMS-0400812, and an Alfred P. Sloan Research Fellowship.

2000 Mathematics Subject Classification: 05C35, 05C65

Keywords: *Turán's theorem, extremal hypergraph problems, stability*

When F is an r-graph, $\pi(F) \neq 0$, and $r > 2$, determining $\pi(F)$ is a hard problem, even for very simple r-graphs F. A result of Erdős and Simonovits implies that if H is an r-graph containing two vertices x, y such that $x \cup S \in H$ iff $y \cup S \in H$, and no edge contains both x and y, then $\pi(H) = \pi(H - y)$. Consequently, when studying $\pi(H)$, we may restrict to the case when H contains no two vertices x and y as above. In this case we say that H is *irreducible*. When $r = 3$, the value of $\pi(F)$ is known for very few irreducible r-graphs F. This lack of knowledge of the behavior of π prevents us from understanding general phenomenon of the extremal theory of hypergraphs. It is therefore of interest to increase the list of irreducible hypergraphs with known Turán density.

Until the late 1990's the number of irreducible r-graphs with known Turán density was less than ten (see the survey of Füredi [8]). In the past few years, there has been some progress, beginning with de Caen and Füredi's proof [2] of Sós' conjecture that $\pi(F) = 3/4$, where F is the Fano plane (see also Füredi-Simonovits [12] and Keevash-Sudakov [14] for exact results and further extensions). Extending this method, the author and Rödl [18] determined π for about ten more irreducible 3-graphs, but in each case the value was $3/4$. They also conjectured that $\pi(F) = 4/9$, where $F = \{123, 124, 125, 345\}$, and gave the lower bound. This conjecture was recently proved by Füredi-Pikhurko-Simonovits [9] and exact results and further extensions were obtained by the same authors in [10]. Another recent result, due to Keevash and Sudakov [15], determines $\pi(C_3^{(2r)})$, where $C_3^{(2r)}$ is the $(2r)$-graph obtained by letting P_1, P_2, P_3 be pairwise disjoint sets of size r, and taking as edges the three sets $P_i \cup P_j$ with $i \neq j$. This result settled a conjecture of Frankl [7]. In spite of this recent activity, until the current work, there were only finitely many irreducible 3-graphs whose Turán density was known.

Our purpose here is to present an infinite family of irreducible r-graphs whose Turán density is exactly determined. For each odd $r \geq 3$, this is the first such family. Moreover, the values of the Turán densities range all the way from $r!/r^r$ (the smallest possible that is not zero) tending to 1. In the definition below, we use $\dot{\bigcup}$ for disjoint union.

Definition. *Fix $l, r \geq 2$. Let $\mathcal{K}_l^{(r)}$ be the family of r-graphs with at most $\binom{l}{2}$ edges, that contain a set S, called the core, of l vertices, with an edge containing every pair of vertices in S. Let $H_l^{(r)} \in \mathcal{K}_l^{(r)}$ be the r-graph with vertex set $A \dot{\bigcup} (\dot{\cup}_{S \in \binom{A}{2}} B_S)$, where $|A| = l, |B_S| = r - 2$ for every S, and edge set $\{S \cup B_S : S \in \binom{A}{2}\}$.*

Remark. *When $r = 2$, the family $\mathcal{K}_l^{(r)}$ reduces to K_l, the usual complete graph, however when $r > 2$, it contains more than one r-graph. Nevertheless, for each fixed r and l, the family $\mathcal{K}_l^{(r)}$ is finite, since every member of it has at most $\binom{l}{2}$ edges.*

We generalize the definition of the Turán graph to hypergraphs. An r-graph is l-partite if its vertex set can be partitioned into l classes, such that every edge has at most one vertex from each class. Thus in particular, there are no edges if $l < r$. A complete l-partite r-graph is one where all of the allowable edges (given a vertex l-partition) are present. For $n, l, r \geq 1$, let $T_r(n,l)$ be the complete l-partite r-graph on n vertices with no two part sizes differing by more than one. Thus the part sizes are $n_i = \lfloor (n+i-1)/l \rfloor$ for $i \in [l]$. Among all l-partite r-graphs on n vertices, $T_r(n,l)$ has the most edges. The number of edges in $T_r(n,l)$ is

$$t_r(n,l) = \sum_{S \in \binom{[l]}{r}} \prod_{i \in S} n_i.$$

Our main theorem is a generalization of Turán's graph theorem, which is the case $r = 2$ below.

Theorem 1. *(Section 2) Let $n, l, r \geq 2$. Then*

$$\mathrm{ex}(n, \mathcal{K}_{l+1}^{(r)}) = t_r(n,l),$$

and the unique r-graph on n vertices containing no copy of a member of $\mathcal{K}_{l+1}^{(r)}$ for which equality holds is $T_r(n,l)$.

As a consequence of Theorem 1, we obtain the Turán density of the infinite family of irreducible r-graphs $H_l^{(r)}$.

Theorem 2. *(Section 3) Let $l \geq r \geq 2$. Then*

$$\pi(H_{l+1}^{(r)}) = \frac{(l)_r}{l^r},$$

where $(l)_r = l(l-1)\cdots(l-r+1)$.

Along with obtaining exact extremal results, one can ask about the structure of nearly extremal structures. The seminal result in this direction is the Simonovits stability theorem for graphs, proved independently by Erdős and Simonovits (see [19]). It states that if an n-vertex K_{l+1}-free graph (n large) has almost as many edges as $T_2(n,l)$, then its structure is very similar to that of $T_2(n,l)$. Similar theorems for hypergraphs have been proven only recently. Papers [12, 10, 13, 14, 15] each prove stability theorems for hypergraphs, [12, 14] for the Fano plane, [10] for $\{123, 124, 125, 345\}$, [15] for $C_3^{(2r)}$, and [13] for cancellative 3-graphs, and also for $\{123, 124, 345\}$. Stability theorems have also proved useful to obtain exact results. This approach, developed and applied in [14, 15], first proves approximate results, then a stability statement, and finally uses the stability result to guarantee an exact extremal result. Our final contribution is a hypergraph analogue of the Simonovits stability theorem for complete graphs.

Theorem 3. *(Section 4) Fix $l \geq r \geq 2$, and $\delta > 0$. Then there exists an $\varepsilon > 0$ and an n_0 such that the following holds for all $n > n_0$: If G is an n-vertex $\mathcal{K}_{l+1}^{(r)}$-free r-graph with at least $t_r(n,l) - \varepsilon n^r$ edges, then G can be transformed to $T_r(n,l)$ by adding and deleting at most δn^r edges.*

We associate every r-graph G with its edge set, and write $V(G)$ for its vertex set. Given a vertex $x \in V(G)$, the *link* of x is $L_G(x) = \{S - \{x\} : x \in S \in G\}$, and the degree is $\deg_G(x) = |L_G(x)|$. The *codegree* of x and y, written $\text{codeg}_G(x,y)$, is the number of edges in G containing both x and y, and the *neighborhood* of x is $N_G(x) = \{y : \textit{codeg}(x,y) > 0\}$. In all cases above, we omit the subscript G if it is obvious from context. For $S \subset V(G)$, we write $G[S]$ for the hypergraph induced by G on S.

2 Three proofs

In this section, we give three proofs of the bound in Theorem 1. Our first proof gives the characterization of the extremal family as well. We begin by noting that for each $k \in [n]$,

$$t_r(n-k, l-1) + k \cdot t_{r-1}(n-k, l-1) \leq t_r(n,l), \tag{1}$$

and if equality holds in (1) then $k = \lfloor n/l \rfloor$ or $\lceil n/l \rceil$. Indeed, for each k, we can consider the LHS as counting the edges in a copy of $T_r(n-k, l-1)$ together with k additional vertices each of whose links is a copy of $T_{r-1}(n-k, l-1)$. Since the vertex partitions of $T_r(n-k, l-1)$ and $T_{r-1}(n-k, l-1)$ are the same (each has $l-1$ parts, no two differing in size by more than one), we may interpret the LHS as the number of edges in a complete l-partite r-graph, where every two of the first $l-1$ part sizes differ by at most one, and the last part has size k. Since $T_r(n,l)$ maximizes the number of edges among all l-partite r-graphs, we conclude that (1) holds, with equality only if $k = \lfloor n/l \rfloor$ or $\lceil n/l \rceil$.

Proofs of Theorem 1

Proof 1 (loosely based on Erdős' 1970 proof [4] of Turán's theorem): We proceed by induction on l, with $l < r$ being trivial. When $r = 2$, the result is Turan's theorem. We therefore assume that $l \geq r > 2$. Let G be an n-vertex $\mathcal{K}_{l+1}^{(r)}$-free r-graph. If $n \leq l$, the result is again trivial, so from now on we assume that $n \geq l+1 \geq r+1 > 3$.

Pick a vertex $x \in V(G)$ of maximum degree Δ. Let $N = N(x)$ be the set of vertices y for which $\text{codeg}_G(x,y) > 0$. Consider the r-graph $G[N]$ induced by N, and suppose that it contains a copy H of a member of $\mathcal{K}_l^{(r)}$. Let $S \subset V(H)$ be the core of H. Form H' from H by adding the vertex x and one edge containing each pair x, v with $v \in S$. These edges exist by the definition of N. Altogether we have added at most l edges, giving $|H'| \leq |H| + l \leq \binom{l}{2} + l = \binom{l+1}{2}$. Therefore $H' \in \mathcal{K}_{l+1}^{(r)}$ which is a contradiction. Consequently, $G[N]$ is $\mathcal{K}_l^{(r)}$-free.

Next consider the $(r-1)$-graph $L(x)$. If $L(x)$ contains a copy H of a member of $\mathcal{K}_l^{(r-1)}$ then by enlarging every edge of H to contain x, we obtain a copy of an $H' \in \mathcal{K}_{l+1}^{(r)}$, since $|H'| = |H| < \binom{l+1}{2}$. Therefore $L(x)$ is $\mathcal{K}_l^{(r-1)}$-free.

Set $k = n - |N|$. By the induction hypothesis, $|G[N]| \leq t_r(n-k, l-1)$ and $\Delta = |L(x)| \leq t_{r-1}(n-k, l-1)$. Since all vertices outside N have degree at most Δ, we conclude that

$$|G| \leq |G[N]| + k \cdot \Delta \leq t_r(n-k, l-1) + k \cdot t_{r-1}(n-k, l-1) \leq t_r(n, l),$$

where the last inequality follows by (1). If equality holds above, then no edge of G contains two vertices in $V(G) - N$, since this would result in over-counting edges in the first inequality. Also, by the discussion after (1), we may assume that $k = \lfloor n/l \rfloor$ or $\lceil n/l \rceil$. Further, by induction we conclude that $G[N]$ is a copy of $T_r(n-k, l-1)$ and the link of each vertex outside N is a copy of $T_{r-1}(n-k, l-1)$. Let us first assume that $l > r$, and fix $z \notin N$. We have already argued that $L(z)$ (which is isomorphic to $T_{r-1}(n-k, l-1)$) has vertex set N. Next we argue that its vertex partition $V_1 \cup \ldots \cup V_{l-1}$ respects that of $G[N]$.

Suppose to the contrary that $G[N]$ has $(l-1)$-partition $W_1 \cup \ldots \cup W_{l-1}$, and $\{v_1, v_2\} \in W_1$, where $v_i \in V_i$. Note that since v_1 and v_2 lie in different parts of $L(z)$, there is an edge of G containing them both. Now pick a vertex $w_j \in W_j$ for each $j > 1$, and consider $S = \{w_2, \ldots, w_{l-1}, v_1, v_2\}$. In order for $G[N]$ to contain at least one edge, we need $n - k \geq l - 1 \geq r$. This follows since $n - k \geq n - \lceil n/l \rceil \geq (l+1) - 2 = l - 1 \geq r$. Therefore every two vertices in different parts of $G[N]$ lie in an edge of $G[N]$. Consequently, for $j \neq j'$, we have $\mathrm{codeg}_{G[N]}(w_j, w_{j'}) > 0$, and $\mathrm{codeg}_{G[N]}(w_j, v_i) > 0$ for $i = 1, 2$. Since v_1 and v_2 also lie in an edge of G (that also contains z), this produces a copy of a member of $\mathcal{K}_l^{(r)}$ with core S. Together with z, we obtain a copy of a member of $\mathcal{K}_{l+1}^{(r)}$, with core $S \cup z$, a contradiction. Therefore each $L(z)$ respects the $(l-1)$-partition of $G[N]$, and G is $T_r(n, l)$ as required.

If $l = r$, then $G[N]$ has no edges, so we cannot use the argument above. In this case we must show that for any two $z, z' \notin N$, the $(r-1)$-partitions of $L(z)$ and $L(z')$ are the same. This follows from an almost identical argument as in the previous paragraph, and we omit the details. □

Proof 2 (based on Turán's original proof of Turán's theorem): For this proof, we need the recurrence

$$t_r(n-1, l) + t_{r-1}(n - \lceil n/l \rceil, l-1) = t_r(n, l).$$

This follows by removing one vertex from $T_r(n,l)$ and counting edges among the remaining $n-1$ vertices, together with edges containing the removed vertex.

Again we proceed by induction on l. Let G be an n-vertex $\mathcal{K}_{l+1}^{(r)}$-free r-graph with $|G| \geq t_r(n,l)$. As in the first proof, we may assume that $n \geq l+1 \geq r+1 > 3$. We know that $t_r(n,l) > t_r(n,l-1)$, so by induction we may assume that $H \subset G$ for some $H \in \mathcal{K}_l^{(r)}$. Let $S = \{w_1, \ldots, w_l\}$ be the core of H. For each $v \in V(G)$, let $s(v)$ be the number of i for which $\operatorname{codeg}(v, w_i) > 0$. If $s(v) = l$ for some v, then $S \cup v$ is the core of a copy of some member of $\mathcal{K}_{l+1}^{(r)}$. We may therefore assume that $s(v) < l$ for each v. Recall that for a vertex x, $|N(x)|$ is the number of y for which $\operatorname{codeg}(x,y) > 0$. By double counting,

$$\sum_{i=1}^{l} |N(w_i)| = \sum_{v \in V(G)} s(v) \leq n(l-1).$$

Consequently, there is an i, for which $|N(w_i)| \leq \lfloor n(l-1)/l \rfloor = n - \lceil n/l \rceil$. As in Proof 1, we know that $L(w_i)$ is $\mathcal{K}_l^{(r-1)}$-free. Therefore by induction

$$|G| \leq |L(w_i)| + |G[V(G) - w_i]| \leq t_{r-1}(n - \lceil n/l \rceil, l-1) + t_r(n-1, l) = t_r(n,l).$$

Although this proof can be extended to give the case of equality, the arguments are not as clean as in Proof 1, and we omit the details. □

Proof 3 (extension of Motzkin and Straus' proof [17] of Turán's theorem): This proof only gives the bound on the number of edges when $l|n$, however for this purpose it is ideally suited. Given an n-vertex r-graph G, define the polynomial

$$f(G, x_1, \ldots, x_n) = \sum_{E \in G} \prod_{i \in E} x_i.$$

The *Lagrange function* of G is

$$\lambda(G) = \max \left\{ f(G, x_1, \ldots, x_n) : x_i \geq 0 \text{ and } \sum_{i=1}^{n} x_i = 1 \right\}.$$

Now let G be an n-vertex $\mathcal{K}_{l+1}^{(r)}$-free r-graph, and let x_i, $i \in [n]$ be chosen for which $f(G, x_1, \ldots, x_n) = \lambda(G)$. Define the support of G by $supp(G) = \{i : x_i > 0\}$. It follows from a lemma of Frankl and Rödl [11] (proved earlier for $r = 2$ by Motzkin and Straus [17]) that if $\{i,j\} \subset supp(G)$, then $\operatorname{codeg}_G(i,j) > 0$. Since G is $\mathcal{K}_{l+1}^{(r)}$-free, we conclude that $|supp(G)| \leq l$. An easy optimization now implies that $\lambda(G) \leq \binom{l}{r}(1/l)^r$. On the other hand, setting each $x_i = 1/n$ gives the lower bound $\lambda(G) \geq |G|/n^r$. Putting this together yields $|G| \leq \binom{l}{r}(n/l)^r$ as needed. □

3 Infinitely many densities

In this section we prove Theorem 2. Denote by $H(k)$ the r-graph obtained from H by replacing each vertex of H by k copies of itself. Call the k copies of vertex v *clones* of v. The supersaturation result of Erdős and Simonovits implies that if $k > 0$ is any fixed integer, then $\pi(H(k)) = \pi(H)$. We need a slightly stronger statement that follows immediately from their argument. For completeness, we sketch the proof.

Lemma 4. *Fix $k, t \geq 1, r \geq 2$, and let $\mathcal{F} = \{H_1, \ldots, H_t\}$ be a (finite) family of r-graphs. Suppose that H is an r-graph satisfying $H \subset H_i(k)$ for every $i \in [t]$. Then $\pi(H) \leq \pi(\mathcal{F})$.*

Proof. (Sketch) In what follows, we write $a \ll b$ to denote that b is much larger than a; for the sake of clarity, we prefer this notation to giving the explicit relationship. Choose $\varepsilon > 0$. Then there exists an $m \gg 1/\varepsilon$ such that every r-graph on m vertices with more than $(\pi(\mathcal{F}) + \varepsilon/2)\binom{m}{r}$ edges contains a copy of some $H_i \in \mathcal{F}$. Choose $n \gg m$.

Suppose that G is an r-graph on n vertices with $|G| > (\pi(\mathcal{F}) + \varepsilon)\binom{n}{r}$. Then an averaging argument (see Erdős-Simonovits [5]) implies that at least $\gamma\binom{n}{m}$ of the m-sets of vertices in G induce an r-graph with more than $(\pi(\mathcal{F}) + \varepsilon/2)\binom{m}{r}$ edges, where $0 < \gamma = \gamma(\varepsilon)$. Each of these m-sets contains a copy of some member of $\mathcal{F}$. Therefore there is an i for which at least $(\gamma/t)\binom{n}{m}$ of the m-sets contain H_i. Consequently the number of copies of H_i in G is at least

$$\frac{(\gamma/t)\binom{n}{m}}{\binom{n-h_i}{m-h_i}} = \frac{\gamma}{t}\frac{(n)_{h_i}}{(m)_{h_i}},$$

where $h_i = |V(H_i)|$. Now, since $n \gg m$, a result of Erdős [3] implies that G contains a copy of $H_i(k)$. Consequently, $H \subset H_i(k) \subset G$, and therefore $\pi(H) \leq \pi(\mathcal{F})$. $\square$

Proof of Theorem 2: We first show that $H_{l+1}^{(r)} \subset H(\binom{l+1}{2} + 1)$ for every $H \in \mathcal{K}_{l+1}^{(r)}$. Pick $H \in \mathcal{K}_{l+1}^{(r)}$, and let $H' = H(\binom{l+1}{2} + 1)$. For each vertex $v \in V(H)$, suppose that the clones of v are $v = v^1, v^2, \ldots, v^{\binom{l+1}{2}+1}$. In particular, identify the first clone of v with v.

Let $S = \{w_1, \ldots, w_{l+1}\} \subset V(H)$ be the core of H. For every $1 \leq i < j \leq l+1$, let $E_{ij} \in H$ with $E_{ij} \supset \{w_i, w_j\}$. Replace each vertex z of $E_{ij} - \{w_i, w_j\}$ by z^q where $q > 1$, to obtain an edge $E'_{ij} \in H'$. Continue this procedure for every i, j, making sure that whenever we encounter a new edge it intersects the previously encountered edges only in H. Since the number of clones is $\binom{l+1}{2} + 1$, this procedure can be carried out successfully and results in a copy of $H_{l+1}^{(r)}$ with core S. Therefore $H_{l+1}^{(r)} \subset H' = H(\binom{l+1}{2} + 1)$. Consequently, Lemma 4 implies that $\pi(H_{l+1}^{(r)}) \leq \pi(\mathcal{K}_{l+1}^{(r)})$.

As $H_{l+1}^{(r)}$ contains a core of size $l+1$, we conclude that $H_{l+1}^{(r)} \not\subset T_r(n,l)$. Therefore

$$\lim_{n\to\infty} \frac{t_r(n,l)}{\binom{n}{r}} \le \pi(H_{l+1}^{(r)}) \le \pi(\mathcal{K}_{l+1}^{(r)}) \le \lim_{n\to\infty} \frac{t_r(n,l)}{\binom{n}{r}},$$

where the last inequality follows from Theorem 1. Since $t_r(n,l) = [(l)_r/(l^r)]\binom{n}{r} + o(n^r)$, the result follows. □

4 Stability

In this section we prove Theorem 3. It is more convenient to prove the following result, which is easily seen to be equivalent to Theorem 3. For a set X of vertices in a hypergraph G, let $e_G(X)$ be the number of edges that contain at least two vertices from X. If it is obvious from context, we will omit the subscript G. We write $a = b \pm c$ to mean that $b - c \le a \le b + c$.

Theorem 5. *Fix $l+1 \ge r \ge 2$. For every δ, there exist ε and M such that if $n > M$ and G is an n-vertex $\mathcal{K}_{l+1}^{(r)}$-free r-graph with $|G| > t_r(n,l) - \varepsilon n^r$, then G has a vertex partition $W_1 \cup \ldots \cup W_l$ satisfying $\sum_i e(W_i) < \delta n^r$.*

Proof. Our proof uses induction on l, with the case $l = r-1$ trivial. The case $r = 2$ is the content of the Simonovits stability theorem, so we further assume that $r > 2$. So assume that $l \ge r > 2$. Choose $\delta = \delta_l > 0$. Our goal is to obtain $\varepsilon = \varepsilon_l$ and $M = M_l$ satisfying the theorem. In what follows, the notation $a \ll b$ means that b is much larger than a, and unless specifically mentioned, we can let $b^{10} > (10lr)^{10}a$ (note that both $a, b < 1$). Choose $\delta_{l-1} \ll \delta_l$. If the theorem holds for ε and M, then it also holds for $\varepsilon' < \varepsilon$ and $M' > M$. Hence by induction there exist $1/M_{l-1} \ll \varepsilon_{l-1} \ll \delta_{l-1}$ for which the theorem holds for $l-1$. Next we describe our choices of ε_l and ε'. For $0 \le x \le 1$, define

$$f_{l,r}(x) = x\binom{l-1}{r-1}\left(\frac{1-x}{l-1}\right)^{r-1} + \binom{l-1}{r}\left(\frac{1-x}{l-1}\right)^{r}.$$

It is easy to see that $f_{l,r}(x)$ has a unique maximum at $x = 1/l$, where its value is $\binom{l}{r}/l^r$. Since $f''(1/l) < 0$, there exist ε_l and ε' such that if $f_{l,r}(x) > \binom{l}{r}/l^r - 2\varepsilon_l$, then $x = 1/l \pm \varepsilon'$. Since for fixed ε', we can always make ε_l smaller with the condition still satisfied, we may assume that $\varepsilon_l \ll \varepsilon' \ll \varepsilon_{l-1}$. Finally, choose $M_l \gg M_{l-1}$. Putting this all together, the hierarchy of constants is

$$\frac{1}{M_l} \ll \frac{1}{M_{l-1}} \ll \varepsilon_l \ll \varepsilon' \ll \varepsilon_{l-1} \ll \delta_{l-1} \ll \delta_l.$$

Now suppose that $n > M_l$, and G satisfies the conditions of the theorem. We will argue as in our first proof of Theorem 1, refining the steps as needed. Let $x, \Delta, N(x), k, L(x)$ be as in that proof, and let $X = V(G) - N(x)$. As before, we can argue that $G[N(x)]$ is $\mathcal{K}_l^{(r)}$-free and $L(x)$ is $\mathcal{K}_l^{(r-1)}$-free. Therefore $|G[N(x)]| \leq t_r(n-k, l-1)$ and $\Delta = |L(x)| \leq t_{r-1}(n-k, l-1)$. This gives

$$\begin{aligned} |G| &\leq |G[N(x)]| + k \cdot \Delta - e_G(X) && (2) \\ &\leq t_r(n-k, l-1) + k \cdot t_{r-1}(n-k, l-1) - e_G(X) && (3) \\ &= t_r(n, l) - e_G(X). && (4) \end{aligned}$$

Claim 1:

$$k = \left(\frac{1}{l} \pm \varepsilon'\right) n.$$

Proof: First observe that $1/M_l \ll \varepsilon_l$ implies that

$$|G| > t_r(n,l) - \varepsilon_l n^r > \left(\binom{l}{r}\frac{1}{l^r} - \varepsilon_l\right) n^r - \varepsilon_l n^r = \left(\binom{l}{r}\frac{1}{l^r} - 2\varepsilon_l\right) n^r. \quad (5)$$

On the other hand, setting $\kappa = k/n$,

$$f_{l,r}(\kappa) \cdot n^r \quad \geq \quad t_r(n-k, l-1) + k \cdot t_{r-1}(n-k, l-1). \quad (6)$$

Now (2), (3), (5), and (6) yield

$$f_{l,r}(\kappa) > \binom{l}{r}\frac{1}{l^r} - 2\varepsilon_l.$$

By the choice of ε_l and ε', we conclude that $\kappa = (1/l \pm \varepsilon')$. □

Claim 2:

$$\Delta = |L(x)| > t_{r-1}(n-k, l-1) - \varepsilon_{l-1}(n-k)^{r-1}.$$

Proof: Otherwise, (3) implies that

$$t_r(n,l) - \varepsilon_l n^r < |G| < t_r(n,l) - k\varepsilon_{l-1}(n-k)^{r-1}.$$

This yields $\varepsilon_l n^r > \varepsilon_{l-1} k (n-k)^{r-1}$. By Claim 1, this implies that $\varepsilon_l > \varepsilon_{l-1}(1/l - \varepsilon')(1 - 1/l - \varepsilon')$. Since $\varepsilon' < 1/(2l)$, and $l \geq 3$, this yields $\varepsilon_l > \varepsilon_{l-1}/(4l)$, which contradicts $\varepsilon_l \ll \varepsilon_{l-1}$. □

Now consider $L(x)$. This $(r-1)$-graph has vertex set $N(x)$ of size $n-k$ and by Claim 2, $|L(x)| > t_{r-1}(n-k, l-1) - \varepsilon_{l-1}(n-k)^{r-1}$. Since $M_l \gg M_{l-1}$, Claim 1 implies that $n - k \gg M_{l-1}$. Moreover, we have already argued that $L(x)$ is $\mathcal{K}_l^{(r-1)}$-free. Since $r \geq 3$, and the Simonovits stability theorem is the case $r = 2$, we may apply the induction hypothesis to $L(x)$. We conclude that $N(x)$ has a vertex partition $W_1 \cup \ldots \cup W_{l-1}$, with $\sum_i e_{L(x)}(W_i) \leq \delta_{l-1}(n-k)^{r-1}$. Consider the vertex l-partition $W_1 \cup \ldots \cup W_{l-1} \cup X$ of G. Our goal now is to prove that

$$e_G(X) + \sum_{i=1}^{l-1} e_G(W_i) < \delta_l \, n^r. \quad (7)$$

Since $|G| > t_r(n,l) - \varepsilon_l n^r$, we conclude from (4) that

$$e_G(X) < \varepsilon_l n^r. \tag{8}$$

Now

$$\sum_{i=1}^{l-1} e_G(W_i) \le \sum_{z\in X}\sum_{i=1}^{l-1} e_{L(z)}(W_i) + \sum_{i=1}^{l-1} e_{G[N(x)]}(W_i). \tag{9}$$

We will bound each of the two sums on the RHS separately, in the next two claims.

Claim 3:

$$\sum_{z\in X}\sum_{i=1}^{l-1} e_{L(z)}(W_i) < 2r\delta_{l-1}n^r.$$

Proof: For $z \in X$, let $\bar{L}(z) = \{S \cup z : S \in L(z)\}$ and

$$B = \{z \in X : e_{\bar{L}(z)}(X) \ge \sqrt{\varepsilon_l}n^{r-1}\}.$$

Then from (8) we obtain

$$\frac{1}{r}|B|\sqrt{\varepsilon_l}n^{r-1} \le e_G(X) < \varepsilon_l n^r.$$

This implies that $|B| < (r\sqrt{\varepsilon_l})n$. Now suppose that $\sum_{z\in X}\sum_{i=1}^{l-1} e_{L(z)}(W_i) \ge 2r\delta_{l-1}n^r$. Then

$$\sum_{z\in X-B}\sum_{i=1}^{l-1} e_{L(z)}(W_i) = \sum_{z\in X}\sum_{i=1}^{l-1} e_{L(z)}(W_i) - \sum_{z\in B}\sum_{i=1}^{l-1} e_{L(z)}(W_i).$$

Since $\delta_{l-1} \gg \varepsilon_l$, this is greater than

$$2r\delta_{l-1}n^r - (|B|ln^{r-1}) > (2r\delta_{l-1} - rl\sqrt{\varepsilon_l})n^r > (2r-1)\delta_{l-1}n^r.$$

Consequently, there exists $z_0 \in X - B$ for which

$$\sum_{i=1}^{l-1} e_{L(z_0)}(W_i) > (2r-1)\delta_{l-1}n^{r-1}.$$

Since $z_0 \notin B$, we have $e_{\bar{L}(z_0)}(X) \le \sqrt{\varepsilon_l}n^{r-1}$. The same $(r-1)$-set in $L(z_0)$ can be counted as many as $(r-1)/2$ times in $\sum_i e_{L(z_0)}(W_i)$. Hence the family of $(r-1)$-sets counted by $\sum_i e_{L(z_0)}(W_i)$ has size at least $2(2r-1)\delta_{l-1}n^{r-1}/(r-1)$. Let G' be the family of $(r-1)$-sets $S' \subset N(x)$ counted by $\sum_i e_{L(z_0)}(W_i)$. Then

$$|G'| \ge \left(\frac{2(2r-1)}{r-1}\delta_{l-1} - \sqrt{\varepsilon_l}\right)n^{r-1} > (3\delta_{l-1} - \sqrt{\varepsilon_l})n^{r-1}.$$

Since $\delta_{l-1} \gg \varepsilon_l$, this is at least $2\delta_{l-1}n^{r-1}$. Let

$$L'(x) = \{S \in L(x) : |S \cap W_i| \le 1 \text{ for every } i \in [l-1]\}.$$

Then $G' \cap L'(x) = \emptyset$, since every set in G' contains at least two elements from some W_i. By the choice of $W_1, \ldots, W_{l-1}$, $|L'(x)| \geq |L(x)| - \delta_{l-1}(n-k)^{r-1}$. Now Claim 2 implies that

$$\begin{aligned} |L'(x) \cup G'| &\geq (|L(x)| - \delta_{l-1}(n-k)^{r-1}) + 2\delta_{l-1}n^{r-1} \\ &> t_{r-1}(n-k, l-1) + (2\delta_{l-1} - \varepsilon_{l-1} - \delta_{l-1})(n-k)^{r-1}. \end{aligned}$$

Since $\delta_{l-1} \gg \varepsilon_{l-1}$, this is greater than $t_{r-1}(n-k, l-1)$. Consequently, there is a copy of some member of $\mathcal{K}_l^{(r-1)}$ contained in $L'(x) \cup G'$. Let S be its core. Adding vertices x and z_0 to this copy yields a copy of some member of $\mathcal{K}_{l+1}^{(r)}$, with core $S \cup x$. This contradiction completes the proof. □

Claim 4:

$$\sum_{i=1}^{l-1} e_{G[N(x)]}(W_i) < r\delta_{l-1}n^r.$$

Proof: Suppose to the contrary that $\sum_{i=1}^{l-1} e_{G[N(x)]}(W_i) \geq r\delta_{l-1}n^r$. For each edge $S \in G[N(x)]$ counted by this sum, choose an $(r-1)$-set $S' \subset S$ such that S' contains at least two vertices in W_i for some $i \in [l-1]$. The same r-set can be counted as many as $r/2$ times in $\sum_i e_{G[N(x)]}(W_i)$. Therefore, the total number of $(r-1)$-sets S' chosen is greater than

$$\frac{\sum_i e_{G[N(x)]}(W_i)}{n(r/2)} > 2\delta_{l-1}n^{r-1}.$$

As argued in Claim 3, none of these $(r-1)$-sets S' appear in $L'(x)$. Consequently, the $(r-1)$-graph H of edges in $L'(x)$ together with the sets S' satisfies

$$|H| > |L'(x)| + 2\delta_{l-1}n^{r-1} > |L(x)| - \delta_{l-1}(n-k)^{r-1} + 2\delta_{l-1}n^{r-1}.$$

By Claim 2 and $\delta_{l-1} > \varepsilon_{l-1}$, this implies that $|H| > t_{r-1}(n-k, l-1)$, which leads to a contradiction as in the proof of Claim 3. □

Now apply Claims 3 and 4 to (9) and use (8). This gives

$$\begin{aligned} e_G(X) + \sum_{i=1}^{l-1} e_G(W_i) &< \varepsilon_l n^r + (2r\delta_{l-1})\, n^r + (r\delta_{l-1})\, n^r \\ &= (\varepsilon_l + 3r\delta_{l-1})\, n^r \\ &< \delta_l\, n^r, \end{aligned}$$

where the last inequality holds since $\varepsilon_l \ll \delta_{l-1} \ll \delta_l$. Consequently (7) holds, and the proof is complete. □

5 Open Problems and Concluding Remarks

The family of r-graphs $\mathcal{K}_{l+1}^{(r)}$ is somewhat similar to the graph K_{l+1}, and this similarity was exploited in the proofs of Theorem 1. However, several well-known proofs of Turán's graph theorem do not seem to easily extend.

• Turán's original proof, which loosely formed the basis of our second proof, doesn't seem to work immediately. In his proof, the induction is performed by removing all the vertices of the smaller clique, but this seems problematic for hypergraphs. Hence we removed only one vertex. Nevertheless, it seems likely that his original proof can also be extended.

• Erdős' 1970 proof of Turán's theorem, on which our first proof is loosely based, seems not to extend in its entirety. In particular, Erdős proved that if G is a K_{l+1}-free graph with degree sequence $d_1 \geq \ldots \geq d_n$, then there exists an l-partite K_{l+1}-free graph G' whose degree sequence $d'_1 \geq \ldots \geq d'_n$ satisfies $d'_i \geq d_i$. From this it is an easy step to derive Turán's theorem. Although we tried to prove this stronger statement, we did not succeed. It would be interesting to decide if this remains true for hypergraphs.

• Caro and Wei gave a proof of Turán's theorem using probabilistic methods (see also Alon-Spencer [1]). It would be interesting to extend this proof to $\mathcal{K}_{l+1}^{(r)}$.

• Li and Li [16] proved Turán's theorem by looking at ideals in polynomials. This is perhaps the most striking and surprising proof of Turán's theorem. In its current form, it does not extend to hypergraphs. In order to conjecture an extension, we briefly describe the proof below. Let G be a graph with vertex set $[n]$. The graph polynomial of G is the homogeneous polynomial on n variables

$$p_G(x_1, \ldots, x_n) = \prod_{i<j,\, ij \notin G} (x_i - x_j).$$

Let $I(n,l) \subset \mathbf{R}(x_1, \ldots, x_n)$ be the ideal of polynomials f such that the identification of any l variables in f results in $f \equiv 0$. It is easy to see that if G is K_l-free, then $p_G \in I(n,l)$. Let $\mathcal{T}_{l-1}$ be the set of all $(l-1)$-partite graphs with vertex set contained in $[n]$, and let $\hat{P}(n,l)$ be the ideal generated by $\{p_G : G \in \mathcal{T}_{l-1}\}$. Since each $G \in \mathcal{T}_{l-1}$ is K_l-free, $\hat{P}(n,l) \subset I(n,l)$. The main result of [16] is that $\hat{P}(n,l) = I(n,l)$. Since the degree of p_G is related to $|G|$, this result allows us to relate the number of edges in a K_l-free graph to the number of edges in a K_l-free graph that is also $(l-1)$-partite, and we obtain Turán's theorem as a consequence.

Here is our proposed extension to 3-graphs. For a 3-graph G with vertex set $[n]$, define the hypergraph polynomial by

$$p_G(x_1,\ldots,x_n) = \prod_{i<j<k,\, ijk\notin G} (x_i - x_j)(x_i - x_k)(x_j - x_k).$$

In order to capture the information given by a pair of vertices with codegree zero, we need the differentiation operator, where $\partial^{(j)} f/\partial x_i$ denotes the partial derivative of f with respect to x_i, taken j times. The reason for this is that we need to speak about roots of polynomials with high multiplicities. Let

$$DI(n,l) = \left\{ p \in \mathbf{R}(x_1,\ldots,x_n) : \frac{\partial^{(j)} p(x_1,\ldots,x_n)}{\partial x_i} \in I(n,l) \quad \text{for every } i,j \text{ with } i\in[n], j\le n-3 \right\}.$$

Once again it is easy to see that if G is $\mathcal{K}_l^{(3)}$-free, then $p_G \in DI(n,l)$. Let $\mathcal{T}_{l-1}^{(3)}$ be the set of all $(l-1)$-partite 3-graphs with vertex set contained in $[n]$, and let $\hat{P}^{(3)}(n,l)$ be the ideal generated by $\{p_G : G \in \mathcal{T}_{l-1}^{(3)}\}$. Since every 3-graph in $\mathcal{T}_{l-1}^{(3)}$ is $\mathcal{K}_l^{(3)}$-free, we have $\hat{P}^{(3)}(n,l) \subset DI(n,l)$.

Conjecture 6. $\hat{P}^{(3)}(n,l) = DI(n,l)$.

An easy consequence of this conjecture is the upper bound in Theorem 1, since $|G|$ is again related to the degree of p_G as in the graph case. A referee pointed out that Conjecture 6 could be posed, with obvious modifications, for r-graphs with $r > 3$ as well.

Our approach to determining $\pi(H_{l+1}^{(r)})$ was to first determine the Turán density for the larger (but finite) family $\mathcal{K}_{l+1}^{(r)}$, and then use supersaturation. It would be nice to proceed directly.

Conjecture 7. [1] *Let $l \ge r \ge 2$. Then for $n > n_0(l,r)$, we have $\mathrm{ex}(n, H_{l+1}^{(r)}) = t_r(n,l)$, and the unique extremal example is $T_r(n,l)$.*

6 Acknowledgments

The author thanks the referees for their helpful comments which improved the presentation.

References

[1] N. Alon, J. Spencer, The probabilistic method. Second edition. With an appendix on the life and work of Paul Erdős. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience [John Wiley & Sons], New York, 2000.

[1] Pikhurko has recently proved this conjecture, along with a stability theorem for $H_{l+1}^{(r)}$

[2] D. de Caen, Z. Füredi, The maximum size of 3-uniform hypergraphs not containing a Fano plane. J. Combin. Theory Ser. B 78 (2000), no. 2, 274–276.

[3] P. Erdős, On extremal problems of graphs and generalized graphs, Israel J. Math, 2, (1964) 183–190.

[4] P. Erdős, On the graph theorem of Turán. (Hungarian) Mat. Lapok 21 (1970), 249–251 (1971).

[5] P. Erdős, M. Simonovits, Supersaturated graphs and hypergraphs, Combinatorica 3 (1983), no. 2, 181–192.

[6] P. Erdős, Problems and results on graphs and hypergraphs: similarities and differences. Mathematics of Ramsey theory, 12–28, Algorithms Combin., 5, Springer, Berlin, 1990.

[7] P. Frankl, Asymptotic solution of a Turán-type problem, Graphs and Combinatorics, 6 (1990), 223-227.

[8] Z. Füredi, Turán type problems. Surveys in combinatorics, 1991 (Guildford, 1991), 253–300, London Math. Soc. Lecture Note Ser., 166, Cambridge Univ. Press, Cambridge, 1991.

[9] Z. Füredi, O. Pikhurko, M. Simonovits, The Turan Density of the Hypergraph $\{abc, ade, bde, cde\}$, Electron. J. Combin., 10 (2003) 7pp.

[10] Z. Füredi, O. Pikhurko, M. Simonovits, On Triple Systems with Independent Neighborhoods, to appear, Comb. Prob. & Comput.

[11] P. Frankl, V. Rödl, Hypergraphs do not jump, Combinatorica 4 (1984), no. 2-3, 149–159.

[12] Z. Füredi, M. Simonovits, Triple systems not containing a Fano configuration, submitted.

[13] P. Keevash, D. Mubayi, Stability results for cancellative hypergraphs, J. Combin. Theory Ser. B 92 (2004), no. 1, 163–175.

[14] P. Keevash, B. Sudakov, The exact Turán number of the Fano plane, to appear, Combinatorica.

[15] P. Keevash, B. Sudakov, On a hypergraph Turán problem of Frankl, to appear, Combinatorica.

[16] R. S. Li, W. W. Li, Independence numbers of graphs and generators of ideals, Combinatorica 1 (1981), no. 1, 55–61.

[17] T. S. Motzkin, E. G. Straus, Maxima for graphs and a new proof of a theorem of Turán, Canad. J. Math. 17, 1965, 533–540.

[18] D. Mubayi, V. Rödl, On the Turán number of triple systems, J. Combin. Theory Ser. A 100, (2002), 136–152.

[19] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, 1968 Theory of Graphs (Proc. Colloq., Tihany, 1966) pp. 279–319