The Turán problem for hypergraphs of fixed size

Peter Keevash

Department of Mathematics
Caltech, Pasadena, CA 91125, USA.
keevash@caltech.edu

Submitted: Oct 22, 2004; Accepted: Jun 3, 2005; Published: Jun 14, 2005
Mathematics Subject Classifications: 05D05

Abstract

We obtain a general bound on the Turán density of a hypergraph in terms of the number of edges that it contains. If $\mathcal{F}$ is an r-uniform hypergraph with f edges we show that $\pi(\mathcal{F}) < \frac{f-2}{f-1} - (1+o(1))(2r!^{2/r} f^{3-2/r})^{-1}$, for fixed $r \geq 3$ and $f \to \infty$.

Given an r-uniform hypergraph $\mathcal{F}$, the Turán number of $\mathcal{F}$ is the maximum number of edges in an r-uniform hypergraph on n vertices that does not contain a copy of $\mathcal{F}$. We denote this number by $ex(n,\mathcal{F})$. It is not hard to show that the limit $\pi(\mathcal{F}) = \lim_{n\to\infty} ex(n,\mathcal{F})/\binom{n}{r}$ exists. It is usually called the *Turán density* of $\mathcal{F}$. There are very few hypergraphs with $r > 2$ for which the Turán density is known, and even fewer for the exact Turán number. We refer the reader to [10, 11, 12, 13, 14, 15, 16] for recent results on these problems.

A general upper bound on Turán densities was obtained by de Caen [3], who showed $\pi(K_s^{(r)}) \leq 1 - \binom{s-1}{r-1}^{-1}$, where $K_s^{(r)}$ denotes the complete r-uniform hypergraph on s vertices. A construction showing $\pi(K_s^{(r)}) \geq 1 - \left(\frac{r-1}{s-1}\right)^{r-1}$ was given by Sidorenko [17] (see also [18]); better bounds are known for large r. We refer the reader to Sidorenko [18] for a full discussion of this problem. For a general hypergraph $\mathcal{F}$ Sidorenko [19] (see also [20]) obtained a bound for the Turán density in terms of the number of edges, showing that if $\mathcal{F}$ has f edges then $\pi(\mathcal{F}) \leq \frac{f-2}{f-1}$. In this note we improve this as follows.

Theorem 1 *Suppose $\mathcal{F}$ is an r-uniform hypergraph with f edges.*
(i) If $r = 3$ and $f \geq 4$ then $\pi(\mathcal{F}) \leq \frac{1}{2}(\sqrt{f^2 - 2f - 3} - f + 3)$.
(ii) For a fixed $r \geq 3$ and $f \to \infty$ we have $\pi(\mathcal{F}) < \frac{f-2}{f-1} - (1+o(1))(2r!^{2/r} f^{3-2/r})^{-1}$.

We start by describing our main tool, which is Sidorenko's analytic approach. See [20] for a survey of this method. Consider an r-uniform hypergraph $\mathcal{H}$ on n vertices. It is convenient to regard the vertex set V as a finite measure space, in which each vertex v has $\mu(\{v\}) = 1/n$, so that $\mu(V) = 1$. We write $h : V^r \to \{0,1\}$ for the symmetric function

$h(x_1, \cdots, x_r)$ which takes the value 1 if $\{x_1, \cdots, x_r\}$ is an edge of $\mathcal{H}$ and 0 otherwise. Then $\int h \, d\mu^r = r! e(\mathcal{H}) n^{-r} = d + O(1/n)$, where $d = \binom{n}{r}^{-1} e(\mathcal{H})$ is the density of $\mathcal{H}$.

Now consider a fixed forbidden r-uniform hypergraph $\mathcal{F}$ with f edges on the vertex set $\{1, \cdots, m\}$. We associate to vertex i the variable x_i, and to an edge $e = \{i_1, \cdots, i_r\}$ the function $h_e(x) = h(x_{i_1}, \cdots, x_{i_r})$, where x denotes the vector $(x_1, \cdots, x_m)$. The configuration product of $\mathcal{F}$ with respect to h is the function $h_{\mathcal{F}}(x) = \prod_{e \in \mathcal{F}} h_e(x)$. Then

$$\int h_{\mathcal{F}} d\mu^m = n^{-m}\text{hom}(\mathcal{F}, \mathcal{H}) = n^{-m}\text{mon}(\mathcal{F}, \mathcal{H}) + O(n^{-1}) = n^{-m}\text{aut}(\mathcal{F})\text{sub}(\mathcal{F}, \mathcal{H}) + O(n^{-1}),$$

where $\text{hom}(\mathcal{F}, \mathcal{H})$ is the number of homomorphisms (edge-preserving maps) from $\mathcal{F}$ to $\mathcal{H}$, $\text{mon}(\mathcal{F}, \mathcal{H})$ is the number of these that are monomorphisms (injective homomorphisms), $\text{aut}(\mathcal{F})$ is the number of automorphisms of $\mathcal{F}$ and $\text{sub}(\mathcal{F}, \mathcal{H})$ is the number of $\mathcal{F}$-subgraphs of $\mathcal{H}$. Also, Erdős-Simonovits supersaturation [6] implies that for any $\delta > 0$ there is $\epsilon > 0$ and an integer n_0 so that for any r-uniform hypergraph $\mathcal{H}$ on $n \geq n_0$ vertices with $\binom{n}{r}^{-1} e(\mathcal{H}) > \pi(\mathcal{F}) + \delta$ we have $n^{-m}\text{sub}(\mathcal{F}, \mathcal{H}) > \epsilon$. It follows that

$$\pi(\mathcal{F}) \; = \; \inf_{\epsilon > 0} \; \liminf_{|V| \to \infty} \; \max_{h: V^r \to \{0,1\}, \; \int h_{\mathcal{F}} \; d\mu^m < \epsilon} \; \int h \; d\mu^r. \tag{1}$$

We say that $\mathcal{F}$ is a forest if we can order its edges as $e_1, \cdots, e_f$ so that for every $2 \leq i \leq f$ there is some $1 \leq j \leq i-1$ so that $e_i \cap \left(\cup_{t=1}^{i-1} e_t\right) \subset e_j$. Sidorenko [20] showed that if $\mathcal{F}$ is a forest with f edges then

$$\int h_{\mathcal{F}} \; d\mu^m \geq \left(\int h \; d\mu^r \right)^f. \tag{2}$$

Now we need a lemma on when a hypergraph contains a forest of given size.

Lemma 2 *(i) An r-uniform hypergraph with at least $r!(t-1)^r$ edges contains a forest with t edges.*
(ii) Let $\mathcal{F}$ be a 3-uniform hypergraph. Then either (a) $\mathcal{F}$ contains a forest with 3 edges, or (b) $\pi(\mathcal{F}) = 0$, or (c) $\mathcal{F} \subset K_4^{(3)}$, or (d) $\mathcal{F} = \mathcal{F}_5 = \{abc, abd, cde\}$.

Proof. (i) This is immediate from the result of Erdős and Rado [5] that such a hypergraph contains a sunflower with t petals, i.e. edges $e_1, \cdots, e_t$ for which all the pairwise intersections $e_i \cap e_j$ are equal. A sunflower is in particular a forest.

(ii) Consider a 3-uniform hypergraph $\mathcal{F}$ that does not contain a forest with 3 edges. We can assume that $\mathcal{F}$ is not 3-partite (Erdős [4] showed that this implies $\pi(\mathcal{F}) = 0$) so $\mathcal{F}$ has at least 3 edges. Clearly $\mathcal{F}$ cannot have two disjoint edges, as then adding any other edge gives a forest.

Suppose there is a pair of edges that share two points, say $e_1 = abc$ and $e_2 = abd$. Any other edge must contain c and d, or together with e_1 and e_2 we have a forest. Consider another edge $e_3 = cde$. If there are no other edges then either $\mathcal{F} = \mathcal{F}_5$ or $\mathcal{F} \subset K_4^{(3)}$ (if e

equals a or b). If there is another edge $e_4 = cdf$ then the same argument shows that e_1 and e_2 both contain e and f, i.e. $\mathcal{F} = K_4^{(3)}$ and there can be no more edges.

The other possibility is that every pair of edges intersect in exactly one point. Then there are at most 2 edges containing any point, or we would have a forest with 3 edges. Consider three edges, which must have the form $e_1 = abc$, $e_2 = cde$, $e_3 = efa$. There can be at most one more edge $e_4 = bdf$. But this forms a 3-partite hypergraph (with parts ad, be, cf), a case we have already excluded. This proves the lemma. $\square$

Proof of Theorem. Let $\mathcal{F}$ be an r-uniform hypergraph with f edges that contains a forest $\mathcal{T}$ with t edges. Label the edges $e_1, \cdots, e_f$, where $e_1, \cdots, e_t$ are the edges of $\mathcal{T}$. Suppose that $\mathcal{H}$ is an r-uniform hypergraph on a vertex set V of size n. Define the measure μ and the function $h : V^r \to \{0,1\}$ as before. Observe the inequality

$$h_{\mathcal{F}}(x) \geq h_{\mathcal{T}}(x) + \sum_{i=t+1}^{f} h_{e_1}(x)(h_{e_i}(x) - 1).$$

This holds, as the second term is non-positive (since $h_e(x) \in \{0,1\}$), so it could only fail for some x if $h_{\mathcal{F}}(x) = 0$ and $h_{\mathcal{T}}(x) = 1$. But then we have $h_{e_1}(x) = \cdots = h_{e_t}(x) = 1$ and $h_{e_i}(x) = 0$ for some $i > t$, and the term $h_{e_1}(x)(h_{e_i}(x) - 1) = -1$ cancels $h_{\mathcal{T}}(x)$, so the inequality holds for all x. Integrating gives

$$\int h_{\mathcal{F}}(x)\ d\mu^m \geq \int h_{\mathcal{T}}(x)\ d\mu^m + \sum_{i=t+1}^{f} \int h_{e_1}(x)h_{e_i}(x) - h_{e_1}(x)\ d\mu^m \geq p^t + (f-t)(p^2 - p),$$

where we write $p = \int h\ d\mu^r$ and apply the inequality (2) for the forests $\mathcal{T}$ and $\{e_1, e_i\}$, $t+1 \leq i \leq f$. By equation (1) we deduce that the Turán density $\pi = \pi(\mathcal{F})$ satisfies $\pi^t + (f-t)(\pi^2 - \pi) \leq 0$.

Writing $g(x) = x^{t-1} + (f-t)(x-1)$ we either have $\pi = 0$ or $g(\pi) \leq 0$. Now $g(0) = -(f-t) \leq 0$, $g(1) = 1$ and $\frac{dg}{dx} = (t-1)x^{t-2} + f - t \geq 0$ for $0 < x < 1$ so g has exactly one root α in $[0,1]$, and $\pi \leq \alpha$.

First we consider the case $r = 3$. If $f \geq 5$ then by the lemma we can take $t = 3$. Solving the quadratic $g(x) = x^2 + (f-3)(x-1) = 0$ gives $\pi \leq \alpha = \frac{1}{2}(\sqrt{f^2 - 2f - 3} - f + 3)$. This also holds when $f = 4$, as then by the lemma we may suppose that $\mathcal{F} = K_4^{(3)}$. Chung and Lu [2] showed that $\pi(K_4^{(3)}) \leq \frac{3+\sqrt{17}}{12}$ which is less than $\frac{1}{2}(\sqrt{5} - 1)$.

Now consider the case when $r \geq 3$ is fixed and $f \to \infty$. By the lemma we can take $t = (f/r!)^{1/r}$. Write $\alpha = 1 - \epsilon$. Since $g(\alpha) = 0$ we have $(f-t)\epsilon = (1-\epsilon)^{t-1} < 1$, so $\epsilon < 1/(f-t)$. From the Taylor expansion of $(1-\epsilon)^{t-1}$ we have $(f-t)\epsilon > 1 - (t-1)\epsilon + \binom{t-1}{2}\epsilon^2 - \binom{t-1}{3}\epsilon^3$. Also $\binom{t-1}{3}\epsilon^3 < \frac{1}{6}\left(\frac{t-1}{f-t}\right)^3 < \frac{1}{6}(t/f)^3$ (since $f > t^2$) so $\binom{t-1}{2}\epsilon^2 - (f-1)\epsilon + 1 - \frac{1}{6}(t/f)^3 < 0$.

Writing $\Delta = (f-1)^2 - 4\binom{t-1}{2}(1-\frac{1}{6}(t/f)^3)$ for the discriminant of this quadratic we have

$$\begin{aligned}
\epsilon &> \frac{f-1-\Delta^{1/2}}{(t-1)(t-2)} = \frac{2(1-\frac{1}{6}(t/f)^3)}{f-1+\Delta^{1/2}} \\
&= \frac{2}{f-1}\left(1+\left(1-2(t-1)(t-2)(1-\frac{1}{6}(t/f)^3)(f-1)^{-2}\right)^{1/2}\right)^{-1} + O(t^3/f^4) \\
&= \frac{2}{f-1}\left(1+1-(t-1)(t-2)(f-1)^{-2}+O(t^4/f^4)\right)^{-1} + O(t^3/f^4) \\
&= \frac{1}{f-1}(1+\frac{1}{2}(t-1)(t-2)(f-1)^{-2}+O(t^4/f^4)) + O(t^3/f^4) \\
&= \frac{1}{f-1} + \frac{(t-1)(t-2)}{2(f-1)^3} + O(t^3/f^4).
\end{aligned}$$

Since $\alpha = 1-\epsilon$ and $t = (f/r!)^{1/r}$ we have

$$\pi \le \alpha < \frac{f-2}{f-1} - (1+o(1))(2r!^{2/r}f^{3-2/r})^{-1}.$$

This proves the theorem. $\square$

Remarks. (1) For a graph G we have $e(G) \ge \binom{\chi(G)}{2}$ with equality if and only if G is complete. The Erdős-Stone theorem [7] implies that $\pi(G) = \frac{\chi(G)-2}{\chi(G)-1} < 1 - \frac{1+o(1)}{\sqrt{2e(G)}}$. It is natural to think that complete hypergraphs should also have the highest Turán density among all hypergraphs with the same number of edges. Were this true de Caen's bound would give $\pi(\mathcal{F}) < 1 - \Omega(f^{-(r-1)/r})$ for an r-uniform hypergraph $\mathcal{F}$ with f edges.

(2) If $\mathcal{F}$ has 3 edges then Sidorenko's bound $\pi(\mathcal{F}) \le 1/2$ is tight when $\mathcal{F} = K_3^{(2)}$ is a triangle, or more generally when $\mathcal{F}$ is the $2k$-uniform hypergraph with edges $\{P_1 \cup P_2, P_2 \cup P_3, P_3 \cup P_1\}$, where P_1, P_2, P_3 are disjoint sets of size k (see [8, 14]). If $\mathcal{F}$ is 3-uniform and has 3 edges then the lemma shows that $\pi(\mathcal{F}) \le \max\{\pi(\mathcal{F}_4), \pi(\mathcal{F}_5)\}$, where $\mathcal{F}_4$ denotes the 3-edge subgraph of $K_4^{(3)}$ and $\mathcal{F}_5 = \{abc, abd, cde\}$. Frankl and Füredi [9] showed that $\pi(\mathcal{F}_5) = 2/9$ and Mubayi [15] showed $\pi(\mathcal{F}_4) < 1/3 - 10^{-6}$, so we see that $\pi(\mathcal{F}) < 1/3 - 10^{-6}$, and Sidorenko's bound is not tight. It would be interesting to determine if it is ever tight for a hypergraph with edges of odd size.

(3) How many edges in an r-uniform hypergraph guarantee a forest with t edges? An answer to this question may lead to an improvement in our theorem, and it also seems interesting in its own right. Erdős and Rado [5] conjectured that for any t there is a constant C so that any r-uniform hypergraph with C^r edges contains a sunflower with t edges. We can obtain a bound of this form for forests, indeed, we claim that any r-uniform hypergraph $\mathcal{F}$ with $(2^t)^r$ edges contains a forest with t edges. For if we fix any edge e, then the other edges have 2^r possible intersections with it, so we can find a hypergraph $\mathcal{F}' \subset \mathcal{F}\backslash e$ with $(2^{t-1})^r$ edges, all of which have the same intersection with e. By induction we can find a forest with $t-1$ edges in $\mathcal{F}'$, and adding e gives a forest of size t in $\mathcal{F}$.

Actually, it is not hard to improve this bound to $2\binom{r}{r/2}^{t-2}$. For we only need the intersections $\{e \cap e' : e' \in \mathcal{F}\}$ to form a chain, and the subsets of e can be partitioned into $\binom{r}{r/2}$ chains (see, for example, [1] page 10). Thus we need only lose a factor $\binom{r}{r/2}$ at each induction step, and after $t-2$ steps we get down to a 2-edge forest.

However, this bound does not help in our application, as we are interested in the case when r is fixed and t is large. We have an upper bound of $r!t^r$ from Erdős and Rado, and and noting that $K_{r+t-2}^{(r)}$ does not contain a forest with t edges we obtain a lower bound of $\binom{r+t-2}{r} \sim t^r/r!$, so we have a constant $r!^2$ factor of uncertainty.

References

[1] B. Bollobás, **Combinatorics**, Set systems, hypergraphs, families of vectors and combinatorial probability, Cambridge University Press, Cambridge, 1986.

[2] F. Chung and L. Lu, An upper bound for the Turán number $t_3(n, 4)$, *J. Combin. Theory Ser. A* **87** (1999), 381–389.

[3] D. de Caen, Extension of a theorem of Moon and Moser on complete subgraphs, *Ars Combin.* **16** (1983), 5–10.

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

[5] P. Erdős and R. Rado, Intersection theorems for systems of sets, *J. London Math. Soc.* **35** 1960, 85–90.

[6] P. Erdős and M. Simonovits, Supersaturated graphs and hypergraphs, *Combinatorica* **3** (1983), 181–192.

[7] P. Erdős and A. H. Stone, On the structure of linear graphs, *Bull. Amer. Math. Soc.* **52** (1946), 1087–1091.

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

[9] P. Frankl, Z. Füredi, A new generalization of the Erdős-Ko-Rado theorem, *Combinatorica* **3** (1983), 341–349.

[10] Z. Füredi, O. Pikhurko and M. Simonovits, On triple systems with independent neighborhoods, *Combin. Probab. Comput.*, to appear.

[11] Z. Füredi and M. Simonovits, Triple systems not containing a Fano Configuration, *Combin. Probab. Comput.*, to appear.

[12] P. Keevash, The Turán problem for projective geometries, *J. Combin. Theory Ser. A*, to appear.

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

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

[15] D. Mubayi, On hypergraphs with every four points spanning at most two triples, *Electron. J. Combin.* **10** (2003), Note 10, 4 pp. (electronic).

[16] D. Mubayi and V. Rödl, On the Turán number of Triple Systems, *J. Comb. Theory Ser. A*, **100** (2002), 136–152.

[17] A. F. Sidorenko, Systems of sets that have the T-property, *Vestnik Moskov. Univ. Ser. I Mat. Mekh.* 1981, 19–22.

[18] A. F. Sidorenko, What we know and what we do not know about Turán numbers, *Graphs Combin.* **11** (1995), 179–199.

[19] A. F. Sidorenko, Extremal combinatorial problems in spaces with continuous measure, *Issled. Operatsiĭi ASU* **34** (1989), 34–40.

[20] A. F. Sidorenko, An analytic approach to extremal problems for graphs and hypergraphs, Extremal problems for finite sets (Visegrád, 1991), 423–455, Bolyai Soc. Math. Stud., 3, János Bolyai Math. Soc., Budapest, 1994.