Hypergraph Turán Problems

Peter Keevash

Abstract

One of the earliest results in Combinatorics is Mantel's theorem from 1907 that the largest triangle-free graph on a given vertex set is complete bipartite. However, a seemingly similar question posed by Turán in 1941 is still open: what is the largest 3-uniform hypergraph on a given vertex set with no tetrahedron? This question can be considered a test case for the general hypergraph Turán problem, where given an r-uniform hypergraph F, we want to determine the maximum number of edges in an r-uniform hypergraph on n vertices that does not contain a copy of F. To date there are very few results on this problem, even asymptotically. However, recent years have seen a revitalisation of this field, via significant developments in the available methods, notably the use of stability (approximate structure) and flag algebras. This article surveys the known results and methods, and discusses some open problems.

Acknowledgements

Research supported in part by ERC grant 239696 and EPSRC grant EP/G056730/1. Thanks to Dan Hefetz, Dhruv Mubayi, Richard Mycroft and Oleg Pikhurko for helpful comments and corrections.

1 Introduction

The *Turán number* $\mathrm{ex}(n, F)$ is the maximum number of edges in an F-free r-graph on n vertices.[1] It is a long-standing open problem in Extremal Combinatorics to develop some understanding of these numbers for general r-graphs F. Ideally, one would like to compute them exactly, but even asymptotic results are currently only known in certain cases. For ordinary graphs ($r = 2$) the picture is fairly complete. The first step was taken by Turán [190], who solved the case when $F = K_t$ is a complete graph on t vertices. The most obvious examples of K_t-free graphs are $(t-1)$-partite graphs. On a given vertex set, the $(t-1)$-partite graph with the most edges is complete and *balanced*, in that the part sizes are as equal as possible (any two sizes differ by at most 1). Turán's theorem is that this construction always gives the largest K_t-free graph on a given vertex set, and furthermore it is unique (up to isomorphism). This result inspired the development of Extremal Graph Theory, which is now a substantial field of research (see [19]). For general graphs F we still do not know how to compute the Turán number exactly, but if we are satisfied with an approximate answer the theory becomes quite simple: it is enough to know the chromatic number of F. Erdős and Stone [62] showed that if $\chi(F) = t$ then $\mathrm{ex}(n, F) \le \mathrm{ex}(n, K_t) + o(n^2)$. As noted in [58], since $(t-1)$-partite graphs are F-free, this implies that $\mathrm{ex}(n, F) = \mathrm{ex}(n, K_t) + o(n^2)$. When F is not bipartite this gives an asymptotic result for the Turán number. When F is bipartite we can only

[1] An *r-graph* (or *r-uniform hypergraph*) G consists of a vertex set and an edge set, each edge being some r-set of vertices. We say G is *F-free* if it does not have a (not necessarily induced) subgraph isomorphic to F.

deduce that $\text{ex}(n, F) = o(n^2)$; in general it is a major open problem to determine even the order of magnitude of Turán numbers for bipartite graphs. However, we will not consider these so-called 'degenerate' problems here.

By contrast with the graph case, there is comparatively little understanding of the hypergraph Turán problem. Having solved the problem for $F = K_t$, Turán [191] posed the natural question of determining $\text{ex}(n, F)$ when $F = K_t^r$ is a complete r-graph on t vertices. To date, no case with $t > r > 2$ of this question has been solved, even asymptotically. Erdős [54] offered \$500 for the solution of any case and \$1000 for a general solution. A comprehensive survey of known bounds on these Turán numbers was given by Sidorenko [180], see also the earlier survey of de Caen [41]; a survey of more general Turán-type problems was given by Füredi [79]. Our focus will be on fixed F and large n, rather than the 'covering design' problems which occur for small n (see [180]). Despite the lack of progress on the Turán problem for complete hypergraphs, there are certain hypergraphs for which the problem has been solved asymptotically, or even exactly, and most of these results have been obtained since the earlier surveys. These special cases may only be scratching the surface of a far more complex general problem, but they are nevertheless interesting for the rich array of different ideas that have been developed for their solutions, ideas that one may hope can be applied or developed to much greater generality. Thus we feel it is most helpful to organise this survey around the methods; we conclude with a summary of the results for easy reference.

The contents by section are as follows: 1: Introduction, 2: Basic arguments, 3: Hypergraph Lagrangians, 4: Link graphs and multigraphs, 5: Stability, 6: Counting, 7: Flag algebras, 8: The remaining exact results, 9: Bounds for complete hypergraphs, 10: The infinitary perspective, 11: Algebraic methods, 12: Probabilistic methods, 13: Further topics, 14: Summary of results.

We use the following notation. Suppose G is an r-graph. We write $V(G)$ for the vertex set of G and $E(G)$ for the edge set of G. We write $v(G) = |V(G)|$ and $e(G) = |E(G)|$. We often identify G with its edge set, so that $|G|$ means $|E(G)|$. For $X \subseteq V(G)$, the induced subhypergraph $G[X]$ has vertex set X and edge set all edges of G that are contained in X. We often abbreviate 'subhypergraph' to 'subgraph'. A *k-set* is a set of size k. Usually G has n vertices, and asymptotic notations such as $o(1)$ refer to the limit for large n.

2 Basic arguments

We start with a simple but important averaging argument of Katona, Nemetz and Simonovits [101]. Suppose G is an r-graph on n vertices with $\theta\binom{n}{r}$ edges. We say that G has *density* $d(G) = \theta$, as this is the fraction of all possible r-sets that are edges. Now fix any $r \leq m < n$ and consider restricting G to subsets of its vertex set of size m. It is easy to check that the average density of these restrictions is also θ. Taking $m = n - 1$, for any fixed r-graph we see that $\binom{n}{r}^{-1}\text{ex}(n, F) \leq \binom{n-1}{r}^{-1}\text{ex}(n-1, F)$. Indeed, if $\theta = \binom{n-1}{r}^{-1}\text{ex}(n-1, F)$ then we cannot have an F-free r-graph on n vertices with density more than θ, as the averaging argument would give a restriction to $n - 1$ vertices with at least the same density, contradicting the definition of $\text{ex}(n-1, F)$. Thus the ratios $\binom{n}{r}^{-1}\text{ex}(n, F)$ form a decreasing sequences of real numbers in $[0, 1]$. It follows that they have a limit, which is called the *Turán density*, and denoted $\pi(F)$.

Determining the Turán density is equivalent to obtaining an asymptotic result $\mathrm{ex}(n,F) \sim \pi(F)\binom{n}{r}$, provided that we are in the 'non-degenerate' case when $\pi(F) > 0$. An r-graph F is degenerate if and only if it is r-partite, meaning that the vertices of F can be r-coloured so that every edge has exactly one vertex of each colour. One direction of this implication is clear: if F is not r-partite then the complete r-partite r-graph on n vertices gives a non-zero lower bound for the Turán density. It has about $(n/r)^r$ edges, so we obtain the bound $\pi(F) \geq r!/r^r$. The other direction is a result of Erdős [50]: if F is r-partite then $\pi(F) = 0$, and in fact $\mathrm{ex}(n,F) < n^{r-c}$ for some $c = c(F) > 0$. We note for future reference that this argument shows that there are no Turán densities in the range $(0, r!/r^r)$. We will return to the question of what values may be taken by Turán densities in Section 13.1.

A similar averaging argument establishes the important 'supersaturation' phenomenon discovered by Erdős and Simonovits [59]. Informally, this states that once the density of an r-graph G exceeds the Turán density of F, we not only find a copy of F, but in fact a constant fraction of all $v(F)$-sets from $V(G)$ span a copy of F.

Lemma 2.1 (Supersaturation) *For any r-graph F and $a > 0$ there are $b, n_0 > 0$ so that if G is an r-graph on $n > n_0$ vertices with $e(G) > (\pi(F)+a)\binom{n}{r}$ then G contains at least $b\binom{n}{v(F)}$ copies of F.*

Proof Fix k so that $\mathrm{ex}(k,F) \leq (\pi(F)+a/2)\binom{k}{r}$. There must be at least $\frac{1}{2}a\binom{n}{k}$ k-sets $K \subseteq V(G)$ inducing an r-graph $G[K]$ with $e(G[K]) > (\pi(F)+\frac{1}{2}a)\binom{k}{r}$.[2] Otherwise, we would have $\sum_K e(G[K]) \leq \binom{n}{k}(\pi(F)+\frac{1}{2}a)\binom{k}{r} + \frac{1}{2}a\binom{n}{k}\binom{k}{r} = (\pi(F)+a)\binom{n}{k}\binom{k}{r}$. But we also have $\sum_K e(G[K]) = \binom{n-r}{k-r}e(G) > \binom{n-r}{k-r}(\pi(F)+a)\binom{n}{r} = (\pi(F)+a)\binom{n}{k}\binom{k}{r}$, so this is a contradiction. By choice of k, each of these k-sets contains a copy of F, so the number of copies of F in G is at least $\frac{1}{2}a\binom{n}{k}/\binom{n-v(F)}{k-v(F)} = \frac{1}{2}a\binom{n}{v(F)}/\binom{k}{v(F)}$, i.e. at least $b = \frac{1}{2}a\binom{k}{v(F)}^{-1}$ fraction of all $v(F)$-sets span a copy of F. $\square$

Supersaturation can be used to show that 'blowing up' does not change the Turán density. The *t-blowup* $F(t)$ of F is defined by replacing each vertex x of F by t 'copies' $x^1, \cdots, x^t$ and each edge $x_1 \cdots x_r$ of F by the corresponding complete r-partite r-graph of copies, i.e. all $x_1^{a_1} \cdots x_r^{a_r}$ with $1 \leq a_1, \cdots, a_r \leq t$. Then we have the following result.

Theorem 2.2 (Blowing up) $\pi(F(t)) = \pi(F)$.

First we point out a special case that will be used in the proof. When $F = K_r^r$ consists of a single edge we trivially have $\mathrm{ex}(n,F) = 0$ and so $\pi(F) = 0$. Also $F(t) = K_r^r(t)$ is the complete r-partite r-graph with t vertices in each part. Then the result of Erdős mentioned above gives $\pi(F(t)) = 0$.

Proof By supersaturation, for any $a > 0$ there is $b > 0$ so that if n is large and G is an r-graph on n vertices with $e(G) > (\pi(F)+a)\binom{n}{r}$ then G contains at least

[2] In fact, large deviation estimates imply that almost all k-sets $K \subseteq V(G)$ have this property when k is large.

$b\binom{n}{v(F)}$ copies of F. Consider an auxiliary $v(F)$-graph H on the same vertex set as G where edges of H correspond to copies of F in G. For any $T > 0$, if n is large enough we can find a copy K of $K_{v(F)}^{v(F)}(T)$ in H. We colour each edge of K by one of $v(F)!$ colours, corresponding to which of the $v(F)!$ possible orders the vertices of F are mapped to the parts of K. Now a standard result of Ramsey theory implies that for large enough T there is a monochromatic copy of $K_{v(F)}^{v(F)}(t)$, which gives a copy of $F(t)$ in G. □

One application of blowing up is to deduce the Erdős-Stone theorem from Turán's theorem: if $\chi(H) = t$ then H is contained in $K_t(s)$ for some s, so $\pi(H) = \pi(K_t) = \frac{t-2}{t-1}$.

Another useful perspective on blowing up is a formulation in terms of homomorphisms. Given r-graphs F and G we say $f : V(F) \to V(G)$ is a *homomorphism* if it preserves edges, i.e. $f(e) \in E(G)$ for all $e \in E(F)$. Note that f need not be injective; if it is then F is a subgraph of G. We say that G is *F-hom-free* if there is no homomorphism from F to G. Clearly, G is F-hom-free if and only if $G(t)$ is F-free for every t. We can make analogous definitions to the Turán number and density for homomorphic copies of F: we let $\mathrm{ex}_{hom}(n, F)$ be the maximum number of edges in an F-hom-free r-graph on n vertices, and $\pi_{hom}(F) = \lim_{n\to\infty} \binom{n}{r}^{-1} \mathrm{ex}_{hom}(n, F)$. Then blowing up implies that $\pi_{hom}(F) = \pi(F)$.

We can in principle approximate $\pi(F)$ to any desired accuracy by an exhaustive search of small examples. For suppose that $m < n$ and we have found that H is a largest F-hom-free r-graph on m vertices with $\theta\binom{m}{r}$ edges. Then averaging gives $\pi(F) \le \theta$. On the other hand, for any t, the blowup $H(t)$ is an F-free r-graph on tm vertices with $t^r \cdot \theta\binom{m}{r}$ edges, so $\pi(F) \ge \lim_{t\to\infty} \binom{tm}{r}^{-1} t^r \theta \binom{m}{r} = \theta \prod_{i=1}^{r-1}(1 - i/m)$. Thus by examining all r-graphs on m vertices one can approximate $\pi(F)$ to within an error of $O(r^2/m)$. Simple brute force search becomes infeasible even for quite small values of m on very powerful computers. However, more sophisticated search techniques can be much faster, and in some cases they give the best known bounds: see Section 7.

3 Hypergraph Lagrangians

The theory in this section was developed independently by Sidorenko [173] and Frankl and Füredi [69], generalising work of Motzkin and Straus [135] and Zykov [193]. Suppose G is an r-graph on $[n] = \{1, \cdots, n\}$. Recall that the t-blowup $G(t)$ of G is obtained by replacing each vertex by t copies. More generally, we can have different numbers of copies of each vertex: for any vector $t = (t_1, \cdots, t_n)$ we let $G(t)$ be obtained by replacing vertex i with t_i copies, where as before, each edge is replaced by the corresponding complete r-partite r-graph of copies. Then $e(G(t)) = p_G(t) := \sum_{e\in E(G)} \prod_{i\in e} t_i$. Note that $p_G(t)$ is a polynomial where for each edge e of G we have the monomial $\prod_{i\in e} t_i$ in variables corresponding to the vertices of e.

Now suppose that F is an r-graph and G is F-hom-free. We will derive an expression for the best lower bound on $\pi(F)$ that can be obtained from blowups of G. Note that $G(t)$ is an F-free r-graph on $|t| := \sum_{i=1}^n t_i$ vertices with density $d(G(t)) =$

$\binom{|t|}{r}^{-1} p_G(t_1, \cdots, t_n)$. Then $\pi(F) \ge \lim_{m\to\infty} d(G(tm)) = r! p_G(t_1/|t|, \cdots, t_n/|t|)$. Thus we want to maximise $p_G(x)$ over the set S of all $x = (x_1, \cdots, x_n)$ with $x_i \ge 0$ for $1 \le i \le n$ and $|x| = 1$. (Sometimes S is called the *standard simplex*.) We denote this maximum by $\lambda(G) = \max_{x\in S} p_G(x)$: it is known as the *Lagrangian* of G. Note that the maximum is achieved by some $x \in S$, as S is compact and $p_G(x)$ is continuous. Also, x can be approximated to arbitrary precision by vectors $(t_1/|t|, \cdots, t_n/|t|)$ with integral t_i. We deduce that $\pi(F) \ge b(G) := r!\lambda(G)$, where we refer to $b(G)$ as the *blowup density* of G.[3] We have the following approximate bound for the blowup density by the usual density: $b(G) \ge r! p_G(1/n, \cdots, 1/n) = r! n^{-r} e(G) = d(G) - O(1/n)$. Since $\pi(F)$ is the limit supremum of $d(G)$ over F-hom-free G, we deduce that $\pi(F)$ is also the supremum of $b(G)$ over F-hom-free G.

We say that G is *dense* if every proper subgraph G' satisfies $b(G') < b(G)$. This is equivalent to saying that the maximum of $p_G(x)$ over $x \in S$ is only achieved by vectors x with $x_i > 0$ for $1 \le i \le n$, i.e. lying in the interior of S. Then $\pi(F)$ is clearly also the supremum of $b(G)$ over F-hom-free dense G. We say that G *covers pairs* if for every pair of vertices i, j in G there is an edge of G containing both i and j. We claim that if G is dense then G covers pairs. This can be seen from the following simple variational argument. Suppose on the contrary that there is no edge containing both i and j for some pair i, j. Then if we consider $p_G(x)$ with any fixed values for the other variables x_k, $k \ne i, j$ we obtain some linear function $ax_i + bx_j + c$ of x_i and x_j. However, a linear function cannot have an internal strict maximum, so the maximum value of $p_G(x)$ can be achieved with one of x_i or x_j equal to 0. This contradicts the assumption that G is dense, so we deduce that G covers pairs.

We can now derive several results from the theory above. First we recover the results for ordinary graphs ($r = 2$). Note that only complete graphs cover pairs, so only complete graphs can be dense. We have $b(K_t) = 1 - 1/t$, so complete graphs are dense. Suppose that G is a K_t-free graph on n vertices. Then $b(G) = b(G')$ for some dense subgraph G', which must be K_s for some $s < t$. We deduce that $2e(G)/n^2 = 2p_G(1/n, \cdots, 1/n) \le b(G) = b(G') = 1 - 1/s \le \frac{t-2}{t-1}$. This gives Turán's theorem in the case when n is divisible by $t - 1$. (This argument is due to Motzkin and Strauss [135].) Also, K_t is F-hom-free if and only if $\chi(F) > t$. Since $\pi(F)$ is the supremum of $b(G)$ for F-hom-free dense G we deduce the Erdős-Stone theorem.

Next we give some hypergraph results. Let H^r_t be the r-graph obtained from the complete graph K_t by extending each edge with a set of $r - 2$ new vertices. More precisely, H^r_t has vertices x_i for $1 \le i \le t$ and y^k_{ij} for $1 \le i < j \le t$ and $1 \le k \le r - 2$ and edges $x_i x_j y^1_{ij} \cdots y^{r-2}_{ij}$ for $1 \le i < j \le t$. We will refer to H^r_t as an *extended complete graph*. (Sometimes 'expanded' is used, but we will use this terminology in a different context later.) Natural examples of H^r_{t+1}-free r-graphs are the blowups $K^r_t(s)$ of the complete r-graph on t vertices. To see that these are H^r_{t+1}-free note that K^r_t is H^r_{t+1}-hom-free, as any map from H^r_{t+1} to K^r_t will map some pair x_i, x_j to the same vertex, so cannot be a homomorphism. On the other

[3] Given the simple relationship between $b(G)$ and $\lambda(G)$ it is arguably unnecessary to give them both names. However, the name *Lagrangian* is widely used, so should be mentioned here, whereas *blowup density* is more descriptive and often notationally more convenient.

hand, if G covers pairs and has at least $t+1$ vertices then it cannot be H^r_{t+1}-hom-free: to define a homomorphism $f : V(H^r_{t+1}) \to V(G)$ we arbitrarily choose distinct vertices as $f(x_1), \cdots, f(x_{t+1})$, then for each $1 \le i < j \le t+1$ we fix an edge e_{ij} containing $f(x_i)f(x_j)$ and map y^k_{ij}, $1 \le k \le r-2$ to $e_{ij} \setminus \{f(x_i), f(x_j)\}$. It follows that $\pi(H^r_{t+1}) = b(K^r_t) = r!t^{-r}\binom{t}{r} = \prod_{i=1}^{r-1}(1 - i/t)$.

The previous result is due to Mubayi [138], who also gave an exact result for the following family of r-graphs including H^r_t. Let $\mathcal{H}^r_t$ be the set of r-graphs F that have at most $\binom{t}{2}$ edges, and have some set T of size t such that every pair of vertices in T is contained in some edge. We extend our earlier definitions to a family $\mathcal{F}$ of r-graphs in the obvious way: we say G is *$\mathcal{F}$-free* if it does not contain any F in $\mathcal{F}$, and then we can define $\text{ex}(n, \mathcal{F})$ and $\pi(\mathcal{F})$ as before. Mubayi [138] showed that the unique largest $\mathcal{H}^r_{t+1}$-free r-graph on n vertices is the balanced blowup of K^r_t. This was subsequently refined by Pikhurko [155], who showed that for large n, the unique largest H^r_{t+1}-free r-graph on n vertices is the balanced blowup of K^r_t.

More generally, suppose F is any r-graph that covers pairs. For any $t \ge v(F)$ we define a hypergraph H^F_t as follows. We label the vertices of F as $v_1, \cdots, v_{v(F)}$. We add new vertices $v_{v(F)+1}, \cdots, v_t$. Then for each pair of vertices v_i, v_j not both in F we add another $r-2$ new vertices u^k_{ij}, $1 \le k \le r-2$ and the edge $v_iv_ju^1_{ij} \dots u^{r-2}_{ij}$. Thus every pair of vertices in F is contained in an edge of H^F_t (although H^F_t does not cover pairs because of the new vertices u^k_{ij}). As an example, if we take F to be the r-graph with no vertices then $H^F_t = H^r_t$ as defined above. The following theorem generalises Mubayi's density result.

Theorem 3.1 *If F is an r-graph that covers pairs and $t \ge v(F)$ satisfies $\pi(F) \le b(K^r_t) = \prod_{i=1}^{r-1}(1 - i/t)$ then $\pi(H^F_{t+1}) = b(K^r_t)$.*

Proof The same argument used for H^r_{t+1} shows that K^r_t is H^F_{t+1}-hom-free, so $\pi(H^F_{t+1}) \ge b(K^r_t)$. For the converse, it suffices to show that any H^F_{t+1}-hom-free dense G satisfies $b(G) \le b(K^r_t)$. This holds by monotonicity if G has at most t vertices, so we can assume G has at least $t+1$ vertices. Now we claim that G is F-hom-free. To see this, note that since F covers pairs, any homomorphism f from F to G is injective, i.e. maps F to a copy of F in G. Then f can be extended to a homomorphism from H^F_{t+1} to G, by the same argument used for H^r_{t+1}. This contradicts our choice of G, so G is F-hom-free. Then $b(G) \le \pi(F) \le b(K^r_t)$. □

The argument of the above theorem is due to Sidorenko [174] where it is given in the special case when F is the 3-graph with 3 edges on 4 vertices and $t = 4$. We will see later (Section 6) that $\pi(F) \le 1/3$. Since $b(K^3_4) = 3/8 > 1/3$ we deduce that in this case $\pi(H^F_4) = 3/8$. Another simple application is to the case when F consists of a single edge. Then $\pi(F) = 0$, so $\pi(H^F_{t+1}) = b(K^r_t)$ for all $t \ge r$. The corresponding exact result for this configuration when n is large is given by Mubayi and Pikhurko [141]; we will call it the *generalised fan*, as they call the case $t = r$ a *fan*. As it has not been explicitly pointed out in the earlier literature, we remark that for every r-graph F that covers pairs, the theorem above gives an infinite family of r-graphs for which we can determine the Turán density.

Sidorenko [174] also applied his method to give the asymptotic result for a construction based on trees that satisfy the Erdős-Sós conjecture. This conjecture (see

[51]) states that if T is a tree on k vertices and G is a graph on n vertices with more than $(k-2)n/2$ edges then G contains T. Although this conjecture is open in general, it is known to hold for many families of trees (e.g. Sidorenko proves it in this case when some vertex is adjacent to at least $(k-2)/2$ leaves, and a proof for large k has been claimed by Ajtai, Komlós, Simonovits and Szemerédi). Suppose that T is a tree satisfying the Erdős-Sós conjecture. Let F be the r-graph obtained from T by adding a set S of $r-2$ new vertices to every edge of T (note that it is the same set for each edge). Let F' be the r-graph obtained from F by adding an edge for each uncovered pair consisting of that pair and $r-2$ new vertices (i.e. $F' = H^F_{v(F)}$). We call F' an *extended tree*. The result is $\pi(F') = b(K^r_{k+r-3})$, provided that $k \geq M_r$, where M_r is a small constant that can be explicitly computed. For example $M_3 = 2$, so when $r = 3$ we have $\pi(F') = b(K^r_k)$ for all $k \geq 2$.

We conclude this section with an application of general optimisation techniques to Turán problems given by Bulò and Pelillo [31]. Suppose that G is a k-graph and consider minimising the polynomial $h(x) = p_{\overline{G}}(x) + a\sum_i x_i^k$ over x in S, where $\overline{G}$ is the complementary k-graph whose edges are the non-edges of G. The intuition for this function is that the first term is minimised when x is supported on a clique of G, whereas the second term is minimised when $x = (1/n, \cdots, 1/n)$, so in combination one might expect a maximum clique to be optimal. It is shown in [31] that this is the case when $0 < a < \frac{1}{k(k-1)}$. One can immediately deduce a bound on the Turán number of K^k_{t+1}. Indeed, the minimum of $h(x)$ is achieved by putting weight $1/t$ on the vertices of a K^k_t, giving value at^{1-k}. On the other hand, substituting $x = (1/n, \cdots, 1/n)$ gives an upper bound of $|\overline{G}|n^{-k} + an^{1-k}$. This gives $|\overline{G}|n^{-k} + an^{1-k} \geq at^{1-k}$, so $\mathrm{ex}(n, K^k_{t+1}) \leq \binom{n}{k} - at(n/t)^k + an$ for any $0 < a < \frac{1}{k(k-1)}$. We will see later that this is not as good as bounds obtained by other methods, but the technique is interesting and perhaps more widely applicable.

4 Link graphs and multigraphs

This section explores the following constructive strategy that can be employed for certain Turán problems. Given an r-graph G and a vertex x of G, the *link* (or *neighbourhood*) $G(x)$ is the $(r-1)$-graph consisting of all $S \subseteq V(G)$ with $|S| = r-1$ and $S \cup \{x\} \in E(G)$. Suppose we are considering the Turán problem for an r-graph F of the following special form: there is some $X \subseteq V(F)$ such that every edge e of F is either contained in X or has exactly one point in X. Then the strategy for finding F is to first find a copy of the subgraph $F[X]$, then extend it to F by consideration of the links of the vertices in X.

Our first example will be to the following question of Katona. Say that an r-graph G is *cancellative* if whenever A, B, C are edges of G with $A \cup C = B \cup C$ we have $A = B$. For example, an (ordinary) graph G is cancellative if and only if it is triangle free. Katona asked for the maximum size of a cancellative 3-graph of n vertices. This was answered as follows by Bollobás [20].

Theorem 4.1 *The unique largest cancellative 3-graph on n vertices is 3-partite.*

It is not hard to see that a 3-partite 3-graph is cancellative. The largest 3-partite 3-graph on n vertices is clearly complete and balanced (meaning as before that the 3 parts are as equal as possible). We denote this 3-graph by $S_3(n)$ and write $s_3(n) = e(S_3(n))$. We will sketch a short proof given by Keevash and Mubayi [106] using link graphs. In this application of the method, the subgraph $F[X]$ described above will just be a single edge $e = xyz$. The links $G(x)$, $G(y)$ and $G(z)$ are pairwise edge-disjoint graphs, for if, say, we had edges xab and yab then $xab \cup xyz = yab \cup xyz$ contradicts G being cancellative. We consider their union U restricted to $V(G) \setminus \{x, y, z\}$ as a 3-edge-coloured graph. Any triangle in U must be 'rainbow' (use all 3 colours), as if, say, ab and ac both have colour x and bc has colour y then $xab \cup bcy = xac \cup bcy$ contradicts G being cancellative.

Proof Suppose G is a cancellative 3-graph on n vertices. For simplicity we just show the inequality $e(G) \le s_3(n)$, though the uniqueness statement also follows easily. We use induction on n. The result is obvious for $n \le 4$ so suppose $n \ge 5$. If any triple of vertices is incident to at most $s_3(n) - s_3(n-3)$ edges then we can delete it and apply induction. Thus we can assume that every triple is incident to more than $s_3(n) - s_3(n-3) = t_3(n) - n + 1$ edges, where $t_3(n)$ denotes the number of edges in the balanced complete 3-partite 'Turán graph' on n vertices. Now consider an edge $e = xyz$. Note that there are at most $n - 3$ edges that intersect e in 2 vertices, otherwise there would be some w that forms an edge with 2 pairs of e, but G is cancellative. Since e is incident to at least $t_3(n) - n + 2$ edges (including itself), the number of edges in U is at least $t_3(n) - n + 2 - (n-3) - 1 = t_3(n-3) + 1$. By Turán's theorem U contains a K_4; let its vertex set be $abcd$. This K_4 is 3-edge-coloured in such a way that every triangle is rainbow, which is only possible when it is properly 3-edge-coloured, i.e. each colour is a matching of two edges. Finally we consider the 7-set $S = xyzabcd$. The colouring of $abcd$ implies that every pair of vertices in S is contained in an edge of G, so have disjoint links. But by averaging, the total size of the links of vertices in S is at least $\frac{7}{3}(t_3(n) - n + 2) > \binom{n}{2}$, contradiction. □

A similar argument was applied in [106] to give a new proof of a theorem of Frankl and Füredi [66]. Note that a 3-graph is cancellative if and only if it does not contain either of the following 3-graphs: $F_4 = \{123, 124, 134\}$, $F_5 = \{123, 124, 345\}$. Thus we can write Bollobás' theorem as $\text{ex}(n, \{F_4, F_5\}) = s_3(n)$. This was improved in [106] to the 'pure' Turán result $\text{ex}(n, F_5) = s_3(n)$ for large n. (This was the first hypergraph Turán theorem.) The original proof required $n \ge 3000$; this was improved to $n \ge 33$ in [106], where the extremal example $S_3(n)$ was also characterised. Very recently, Goldwasser [86] has determined $\text{ex}(n, F_5)$ and characterised the extremal examples for all n: $S_3(n)$ is the unique extremal example for $n > 10$, the 'star' (all triples containing some fixed vertex) is the unique extremal example for $n < 10$, and both $S_3(n)$ and the star are extremal for $n = 10$. A new proof of the asymptotic form of the Frankl-Füredi theorem had previously been given by Mubayi and Rödl [143]. That paper applied the link method to obtain several other bounds on Turán densities. They also gave 5 specific 3-graphs each of which has Turán density 3/4. One of these, denoted $F(3, 3)$, is obtained by taking an edge abc, three additional vertices d, e, f, and all edges with one vertex from abc and two from def.

We remark that induction arguments as in the above proof are often very useful for Turán problems. Above it was convenient to consider deleting triples, but usually one considers deleting a single vertex. Then in order to prove the statement $e(G) \le$

$f(n)$ for an F-free r-graph G on n vertices one can assume that the minimum degree of a vertex in G is more than $f(n) - f(n-1)$. (This argument gives one of the simplest proofs of Turán's theorem.) A caveat is that this induction argument depends on being able to prove a base case, which is not always convenient, as the desired bound may not even be true for small n. Then the following proposition is a more convenient method for obtaining a minimum degree condition. (The proof is to repeatedly delete vertices with degree less than the stated bound: a simple calculation shows that this process terminates and that the final graph has many vertices.)

Proposition 4.2 *For any $\delta, \epsilon > 0$ and $n_0 \geq r \geq 2$ there is n_1 so that any r-graph on $n \geq n_1$ vertices with at least $(\delta + 2\epsilon)\binom{n}{r}$ edges contains an r-graph on $m \geq n_0$ vertices with minimum degree at least $(\delta + \epsilon)\binom{m-1}{r-1}$.*

Our next example using links is by de Caen and Füredi [42], who were the originators of the method. They gave a surprisingly short proof of a conjecture of Sós [183] on the Turán number of the Fano plane. The Fano plane is an ubiquitous object in combinatorics. It is the unique 3-graph on 7 vertices in which every pair of vertices is contained in exactly one edge. It can be constructed by identifying the vertices with the non-zero vectors of length 3 over $\mathbb{F}_2$ (the field with two elements), and the edges with triples $\{x, y, z\}$ with $x + y = z$. It is easy to check that the Fano plane is not bipartite, in that for any partition of its vertex set into two parts, at least one of the parts must contain an edge. Thus a natural construction of a Fano-free 3-graph on n vertices is to take the balanced complete *bipartite* 3-graph: the vertex set has two parts of size $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$, and the edges are all triples that intersect both parts. Sós conjectured that this construction gives the exact value for the Turán number of the Fano plane. The following result from [42] verifies this conjecture asymptotically (see Section 5 for the exact result).

Theorem 4.3 *$ex(n, Fano) \sim \frac{3}{4}\binom{n}{3}$.*

As in the previous example, the construction starts with a single edge $e = xyz$. We combine the links to create a 3-edge-coloured *link multigraph* $L = G(x) + G(y) + G(z)$, where $+$ denotes multiset union; note that unlike the previous example the links need not be edge-disjoint. To find a Fano plane we need to find the same object which appeared in the previous proof: a properly 3-edge-coloured K_4. (Since an edge may have more than one colour, this means we can select a colour for each edge to obtain the required colouring.) To prove an asymptotic result we can assume that G has at least $(3/4 + 2\epsilon)\binom{n}{3}$ edges for some small $\epsilon > 0$, and then that G has minimum degree at least $(3/4 + \epsilon)\binom{n}{2}$ by Proposition 4.2. However, the argument now seems to get stuck at the point of using this lower bound on the links to find a properly 3-edge-coloured K_4.

The key idea is to instead start with a copy of K_4^3, with the intention of using one of its edges as the edge e above. Since G has edge density more than $3/4$, averaging shows that it has a 4-set $wxyz$ of density more than $3/4$, i.e. spanning K_4^3 (we will see better bounds later on the Turán density of K_4^3). Now we consider the 4-edge-coloured link multigraph L, obtained by restricting $G(w)+G(x)+G(y)+G(z)$ to $V(G) \setminus \{w, x, y, z\}$. It suffices to find a properly 3-edge-coloured K_4. This can be achieved by temporarily forgetting the colours of the edges and just counting multiplicities. Thus we consider L as a multigraph with edge multiplicities at most 4 and at least $(3+4\epsilon)\binom{n}{2}$ edges. Such a multigraph must have a 4-set $abcd$ that spans at least 21 edges in L: this is a special case of a theorem of Füredi and Kündgen [80]. Finally we put the colours back. We may consider the bipartite graph B in which one part B_1 is the 4-set $wyxz$, the other part B_2 is the 3 matchings of size 2 formed by $abcd$, and edges in B correspond to edges of G in the obvious way, e.g. we join w to $\{ab, cd\}$ if wab and wcd are edges. Since $L[abcd]$ is at most 3 edges from being complete, the same is true of B. This implies (e.g. using Hall's theorem) that B has a matching that covers B_2. This gives the proper 3-edge-colouring of $abcd$ required to prove the theorem.

5 Stability

Many extremal problems have the property that there is a unique extremal example, and moreover any construction of close to maximum size is structurally close to this extremal example. For example, in the Turán problem for the complete graph K_t, Turán's theorem determines $\mathrm{ex}(n, K_t)$ and describes the unique extremal example as the balanced complete $(t-1)$-partite graph on n vertices. More structural information is given by the Erdős-Simonovits Stability Theorem [182], which may be informally stated as saying that any K_t-free graph G on n vertices with $e(G) \sim \mathrm{ex}(n, K_t)$ is structurally close to the extremal example. More precisely, we have the following statement.

Theorem 5.1 *For any $\epsilon > 0$ there is $\delta > 0$ such that if G is a K_t-free graph with at least $(1-\delta)ex(n, K_t)$ edges then there is a partition of the vertices of G as $V_1 \cup \cdots \cup V_{t-1}$ with $\sum_i e(V_i) < \epsilon n^2$.*

As well as being an interesting property of extremal problems, this phenomenon gives rise to a surprisingly useful tool for proving exact results. This stability method has two stages. First one proves a stability theorem, that any construction of close to maximum size is structurally close to the conjectured extremal example. Armed with this, we can consider any supposed better construction as being obtained from the extremal example by introducing a small number of imperfections into the structure. The second stage is to analyse any possible imperfection and show that it must lead to a suboptimal configuration, so in fact the conjectured extremal example must be optimal.

This approach can be traced back to work of Erdős and Simonovits in the 60's in extremal graph theory (see [182]). More recently it was applied independently by Keevash and Sudakov [110] and by Füredi and Simonovits [84] to prove the conjecture of Sós mentioned above in an exact form: for large n the unique largest Fano-free 3-graph on n vertices is the balanced complete bipartite 3-graph. Since then it has been applied to many problems in hypergraph Turán theory and more broadly in combinatorics as whole. We will discuss the other Turán applications later; we refer the reader to [107] for an application in extremal set theory and some further references using the method.

To understand how the method works in more detail, it is helpful to consider the 'baby' case of the Turán problem for the 5-cycle C_5. This is not hard to handle by other means, but is sufficiently simple to illustrate the method without too many technicalities. Since $\chi(C_5) = 3$, the Erdős-Stone theorem gives $\mathrm{ex}(n, C_5) \sim \mathrm{ex}(n, K_3) = \lfloor n^2/4 \rfloor$. In fact $\mathrm{ex}(n, C_5) = \lfloor n^2/4 \rfloor$ for $n \geq 5$. We will sketch a proof of this equality for large n. This is a special case of a theorem of Simonovits, that if F is a graph with $\chi(F) = t$ and $\chi(F \setminus e) < t$ for some edge e of F then $\mathrm{ex}(n, F) = \mathrm{ex}(n, K_t)$ for large n. (We say that such graphs F are *critical*: examples are cliques and odd cycles.) The first step is the following stability result (we state it informally, the precise statement is similar to that in Theorem 5.1).

Lemma 5.2 *Suppose G is a C_5-free graph on n vertices with $e(G) \sim n^2/4$. Then G is approximately complete bipartite.*

Proof (Sketch.) First we claim that we can assume G has minimum degree $\delta(G) \sim n/2$. This is a similar statement to that in Proposition 4.2, although we cannot apply that result, as we cannot remove too many vertices if we want to obtain the structure of G. The solution is to use the same vertex deletion argument and use the bound from the Erdős-Stone theorem to control the number of vertices deleted. The calculation is as follows, for some small $\delta > 0$. If $e(G) > (1-\delta)n^2/4$ then we can delete at most $\delta^{1/2}n$ vertices of proportional degree less than $1/2 - \delta^{1/2}$, otherwise we arrive at a C_5-free graph G' on $n' = (1-\delta^{1/2})n$ vertices with $e(G') > e(G) - \sum_{i=(1-\delta^{1/2})n+1}^{n} (1/2 - \delta^{1/2})i > (1-\delta)n^2/4 - \delta^{1/2}n^2/2 + \delta^{1/2}\left(\binom{n+1}{2} - \binom{n'+1}{2}\right) = n'^2/4 - \delta n^2/2 + \delta(1-\delta^{1/2}/2)n^2 - O(n) > (1+\delta)n'^2/4$, contradiction.

Thus we can assume that $\delta(G) \sim n/2$. Next we choose a 4-cycle $abcd$ in G. These are plentiful, as G has edge density about 1/2, whereas the 4-cycle is bipartite, so has zero Turán density. Now we note that the neighbourhoods $N(a)$ and $N(b)$ cannot share a vertex x other than c or d, otherwise $axbcd$ is a 5-cycle. Furthermore each of these neighbourhoods does not contain a path of length 3, e.g. if $wxyz$ is a path of length 3 in $N(a)$ then $awxyz$ is a 5-cycle. Thus each is very sparse, e.g. $N(a)$ cannot have average degree at least 6, as it is not hard to show that it would then have a subgraph of minimum degree at least 3, and so a path of length 3. Thus we have found two disjoint sets of size about $n/2$ containing only $O(n)$ edges. □

The second step is to refine the approximate structure and deduce an exact result.

Theorem 5.3 *$ex(n, C_5) = \lfloor n^2/4 \rfloor$ for large n.*

Proof (Sketch.) Suppose G is a maximum size C_5-free graph on n vertices. We claim that we can assume G has minimum degree $\delta(G) \geq \lfloor n/2 \rfloor$. For suppose we have proved the result under this assumption for all $n \geq n_0$. Then suppose that n is much larger then n_0 and repeatedly delete vertices while the minimum degree condition fails. A similar calculation to that in the lemma shows that this process terminates with a C_5-free graph G' on $n' \geq n_0$ vertices with $\delta(G') \geq \lfloor n'/2 \rfloor$, and moreover if any vertices were deleted we have $e(G') > \lfloor n'^2/4 \rfloor$, contradiction. Thus we can assume $\delta(G) \geq \lfloor n/2 \rfloor$.

By the lemma G is approximately complete bipartite. Consider a bipartition $V(G) = A \cup B$ that is optimal, in that $e(A) + e(B)$ is minimised. Then $e(A) + e(B) <$

ϵn^2, for some small $\epsilon > 0$. Also, A and B each have size about $n/2$, say $(1/2 \pm \epsilon^{1/2})n$, otherwise $e(G) < |A||B| + \epsilon n^2 < n^2/4$, contradicting G being maximum size. Write $d_A(x) = |N(x) \cap A|$ and $d_B(x) = |N(x) \cap B|$ for any vertex x. Note that for any $a \in A$ we have $d_A(a) \leq d_B(a)$, otherwise we could improve the partition by moving a to B. Similarly, $d_B(b) \leq d_A(b)$ for any $b \in B$.

Next we claim that the 'bad degrees' must be small, e.g. that $d_A(a) < cn$ for all $a \in A$ where $c = 2\epsilon^{1/2}$. For suppose this fails for some a. Then $N(a) \cap A$ and $N(a) \cap B$ both have size at least cn. Moreover they span a bipartite graph with no path of length 3, so only $O(n)$ edges. This gives $(cn)^2 - O(n) > e(A) + e(B)$ 'missing edges' between A and B, so $e(G) < |A||B| \leq n^2/4$, contradiction.

Finally we claim that there are no 'bad edges', i.e. that (A, B) gives a bipartition of G. For suppose that aa' is an edge in A. Then $|N_B(a) \cap N_B(a')| > d(a) - cn + d(a') - cn - |B| > (1/2 - 5\epsilon^{1/2})n$. But there is no path $ba''b'$ with $b, b' \in B' = N_B(a) \cap N_B(a')$ and $a'' \in A' = A \setminus \{a, a'\}$, so A' and B' span a bipartite graph with only $O(n)$ edges - a very emphatic contradiction! □

The above proof illustrates a template that is followed by many (but not all) applications of the stability method. In outline, the steps are: (i) G has high minimum degree, (ii) G has approximately correct structure, so consider an optimal partition, (iii) the bad degrees are small, (iv) there are no bad edges. For example, the deduction in [110] of the exact result for the Fano plane from the stability result follows this pattern. It is instructive to note that as well as the link multigraph method of the previous section, considerable use is made of an additional property of the Fano plane: there is a vertex whose deletion leaves a 3-partite 3-graph (any vertex has this property). It is an intriguing problem to understand what properties of an r-graph F make it amenable to either of the two steps of the stability method, i.e. whether a stability result holds, and whether it can be used to deduce an exact result.

A variant form of the stability approach is to prove a statement analogous to the following theorem of Andrásfai, Erdős and Sós [7]: any triangle-free graph G on n vertices with minimum degree $\delta(G) > 2n/5$ is bipartite. The approach taken in [84] to the exact result for the Fano plane is to prove the following statement: if $\delta > 0$, n is large, and G is a Fano-free 3-graph on n vertices with $\delta(G) > (3/4 - \delta)\binom{n}{2}$ then G is bipartite. One might think that this is a stronger type of statement than the stability result, but in fact it is equivalent in difficulty: it follows by exactly the same proof as that of the refinement argument sketched above (the second stage of the stability method). It would be interesting, and probably rather more difficult, to determine the smallest minimum degree for which this statement holds. For example, in the Andrásfai-Erdős-Sós theorem the bound $2n/5$ is tight, as shown by the blowup of a 5-cycle; what is the analogous 'second-best' construction for the Fano plane? (Or indeed, for other hypergraph Turán results...?)

We conclude this section by mentioning a nice application of the stability method to showing the 'non-principality' of Turán densities. If $\mathcal{F}$ is a set of graphs then it is clear that $\pi(\mathcal{F}) = \min_{F \in \mathcal{F}} \pi(F)$; one can say that the Turán density for a set of graphs is 'principal', in that it is determined by just one of its elements. However, Balogh [10] showed that this is not the case for hypergraphs, confirming a conjecture of Mubayi and Rödl [143]. Mubayi and Pikhurko [142] showed that even a set of two hypergraphs may not be principal. Let F denote the Fano plane and let F' be the 'cone' of K_t, i.e. the 3-graph on $\{0, \cdots, t\}$ with edges $0ij$ for $1 \le i < j \le t$. Note that F' is not contained in the blowup of K_t^3, so $\pi(F') \to 1$ as $t \to \infty$. Thus we can choose t so that $\min\{\pi(F), \pi(F')\} = \pi(F) = 3/4$. Now we claim that $\pi(\{F, F'\}) < 3/4$. To see this, consider a large Fano-free 3-graph G with edge density $3/4 - o(1)$. Then G is approximately a complete bipartite 3-graph. But the complete bipartite 3-graph contains many copies of F', and these cannot all be destroyed by the approximation, so G contains a copy of F', as required.

6 Counting

The arguments in this section deduce bounds on the edge density of an F-free r-graph G from counts of various small subgraphs. The averaging argument discussed in Section 2 gives the essence of this idea, but this uses F-freeness only to say that r-graphs containing F have a count of zero in G. However, there are various methods that can extract information about other counts. Arguments using the Cauchy-Schwartz inequality or non-negativity of squares play an important role here. We will illustrate this in the case when F is the (unique) 3-graph on 4 vertices with 3 edges (we called this F_4 in Section 4). First consider what can be obtained from a basic averaging argument. If G is an F-free 3-graph then every 4-set of G has at most 2 edges, i.e. density at most 1/2. It follows that the density of G is at most 1/2.

For an improvement, consider the sum $S = \sum_{xy} \binom{d(x,y)}{2}$, where the sum is over unordered pairs of vertices xy and $d(x, y)$ denotes the number of edges containing xy. Note that S counts the number of unordered pairs ab such that axy and bxy are edges. Since we are assuming that every 4-set has at most 2 edges, this is exactly the number of 4-sets $abxy$ with 2 edges. On the other hand, the Cauchy-Schwartz inequality (more precisely, convexity of the function $f(t) = \binom{t}{2}$) gives $S \ge \binom{n}{2}\binom{d}{2}$, where d is the average value of $d(x, y)$. We have $\binom{n}{2}d = \sum_{xy} d(x, y) = 3e(G) = 3\theta\binom{n}{3}$, where θ denotes the edge density of G, so $S \ge \frac{1}{4}\theta^2 n^4 + O(n^3)$. For an upper bound on S, we can double-count pairs (T, e) where T is a 4-set with 2 edges and e is an edge of T. This gives $2S \le (n - 3)e(G)$, so $S \le \frac{1}{12}\theta n^4 + O(n^3)$. Combining the bounds on S we obtain $\theta \le 1/3 + O(1/n)$ as a bound on the density of G.

It is remarkable that the Turán problem is still open for such a seemingly simple 3-graph – we do not even know its Turán density! For many years the above argument (due to de Caen [38]) gave the best known upper bound. More recently it has been improved in a series of papers [134, 136, 186, 144, 158]; the current best bound is 0.2871 due to Baber and Talbot [9]. The best known lower bound is $\pi(F) \ge 2/7 = 0.2857\cdots$, due to Frankl and Füredi [67]. It is worth noting their unusual iterative construction. For most Turán problems, the known or conjectured optimal construction is obtained by dividing the vertex set into a fixed number of parts and defining edges by consideration only of intersection sizes with the parts. However, this construction is obtained by first blowing up a particular 3-graph on 6 vertices, then iteratively substituting copies of the construction inside each of the 6 parts. The 3-graph in question has the 10 edges 123, 234, 345, 451, 512, 136, 246, 356, 416, 526. This has blowup density $3!\frac{10}{6^3} = \frac{5}{18}$, so the overall density is given by the geometric series $\frac{5}{18}\sum_{i \ge 0}(1/36)^i = \frac{2}{7}$. If this is indeed the optimal construction, as conjectured in [136], then its complicated nature gives some indication as to why it has so far eluded proof. It is also an intriguing example from the point of view of the finiteness questions discussed in [123, 156] (see Section 7).

Our next example of a counting argument, due to Sidorenko [175], gives a bound on the Turán density that depends only on the number of edges. If F is an r-graph with f edges then the bound is $\pi(F) \le \frac{f-2}{f-1}$. Before proving this we set up some notation. Suppose that G is an r-graph on n vertices. Label the vertices of F as $\{v_1, \cdots, v_t\}$ for some t and the edges as $\{e_1, \cdots, e_f\}$. Let $x = (x_1, \cdots, x_t)$ denote an arbitrary t-set of vertices in G. We think of the map $\pi_x : v_i \mapsto x_i$ as a potential embedding of F in G. Write $e_i(x)$ for the indicator function that is 1 if $\pi_x(e_i)$ is an edge of G or 0 otherwise. Then $F(x) = \prod_{i=1}^f e_i(x)$ is 1 if π_x is a homomorphism from F to G or 0 otherwise. Now suppose that G is F-free. Then $F(x)$ can only be non-zero if $x_i = x_j$ for some $i \neq j$. If we choose x randomly this has probability $O(1/n)$, so $\mathbb{E}F(x) = O(1/n)$.

Now comes a trick that can only be described as pulling a rabbit out of a hat. We claim that the following inequality holds pointwise: $F(x) \ge e_1(x) + \sum_{i=2}^f e_1(x)(e_i(x)-1)$. To see this, note that since $F(x)$ and all the $e_i(x)$ are $\{0,1\}$-valued, the only case where the inequality is not obvious is when $F(x) = 0$ and $e_1(x) = 1$. But then $F(x) = 0$ implies that $e_i(x) = 0$ for some i, and then the term $e_1(x)(e_i(x)-1)$ gives a -1 to cancel $e_1(x)$, so the inequality holds. Taking expected values gives $\mathbb{E}F(x) \ge \sum_{i=2}^f \mathbb{E}e_1(x)e_i(x) - (f-2)\mathbb{E}e_1(x)$. Now $\mathbb{E}e_1(x) = \theta + O(1/n)$, where θ is the edge density of G. Also, the Cauchy-Schwartz inequality implies that $\mathbb{E}e_1(x)e_i(x) \ge \theta^2 + O(1/n)$ for each i (this is similar to the lower bound on S in the previous example). We deduce that $(f-1)\theta^2 - (f-2)\theta \le O(1/n)$, i.e. $\theta \le \frac{f-2}{f-1} + O(1/n)$, as required.

It is an interesting question to determine the extent to which this bound can be improved. When $f = 3$ no improvement is possible for general r, as the bound of $1/2$ is achieved by the triangle for $r = 2$, or the 'expanded triangle' for even r (see [111]). For $r = f = 3$ there are only two non-degenerate cases to consider, namely the 3-graphs F_4 and F_5 discussed in Section 4. The Turán number of F_5 is given by the complete 3-partite 3-graph, which has density $2/9$. Then from the discussion of F_4 above we see that when $r = f = 3$ the bound of $1/2$ is quite far from optimal. We do not know if $1/2$ can be achieved when $f = 3$ and $r \ge 5$ is odd. A refinement of Sidorenko's argument given in [102] shows that the bound $\pi(F) \le \frac{f-2}{f-1}$ can be improved if r is fixed and f is large. One might think that the worst case is when F is a complete r-graph, which would suggest an improvement to $\pi(F) \le 1 - \Omega(f^{-(r-1)/r})$ (see [102] for further discussion). Another general bound obtained by the Local Lemma will be discussed later in the section on probabilistic methods.

The above two examples show that the 'right' counting argument can be surprisingly effective, but they do not give much indication as to how one can find this argument. For a general Turán problem there are so many potential inequalities that might be useful that one needs a systematic approach to understand their capabilities. The most significant steps in this direction have been taken by Razborov [158], who has obtained many of the sharpest known bounds on Turán densities using his theory of flag algebras [156]. We will describe this in Section 7, but first we note that earlier steps in this direction appear in the work of de Caen [39] and Sidorenko [172].

We briefly describe the quadratic form method in [172], as it has some additional features that are not exploited by other techniques. Suppose G is an r-graph on $[n]$. Let y be the vector in $\mathbf{R}^{\binom{n}{r-1}}$, where co-ordinates correspond to $(r-1)$-sets of vertices, and the entry for a given $(r-1)$-set is its degree, i.e. the number of edges containing it. Let u be the all-1 vector. Writing $(\cdot,\cdot)$ for the standard inner product, we have $(u,u)=\binom{n}{r-1}$, $(y,u)=r|G|$ and $(y,y)=r|G|+p$, where p is the number of ordered pairs of edges that intersect in $r-1$ vertices. Let $\overline{G}=\{[n]\setminus e : e \in E(G)\}$ be the complementary $(n-r)$-graph and define $\overline{y}$, $\overline{u}$, $\overline{p}$ similarly. Note that the degree of an $(n-r-1)$-set in $\overline{G}$ is equal to the number of edges of G in the complementary $(r+1)$-set. Note also that $\overline{p}=p$. Let Q_i be the number of $(r+1)$-sets that contain at least i edges. Thus $\overline{y}$ has Q_i-Q_{i+1} co-ordinates equal to i. For any $t\geq 1$ we have the inequality $(\overline{y}-(r+1)\overline{u},\overline{y}-t\overline{u}) \leq \sum_{i=0}^{t}(r+1-i)(t-i)(Q_i-Q_{i+1})$. Then come some algebraic manipulations which we will just summarise: (i) use summation by parts, (ii) substitute $Q_0=(\overline{u},\overline{u})=\binom{n}{r+1}$, (iii) rewrite in terms of y and u. The resulting inequality is $(y,y)+((1+t/r)(n-r)+1)(y,u)+\sum_{i=1}^{t}(t+r+2-2i)Q_i \leq 0$.

Now we come to the crux of the method, which is an inequality for (y,u) for general vectors y, u satisfying a quadratic inequality as above. The inequality is that if $(y,y)-2a(y,u)+b(u,u)\leq 0$ then $(y,u)\geq(a-\sqrt{a^2-b})(u,u)$. To see this add $-(y-su,y-su)\leq 0$ to the first inequality, which gives $2(s-a)(y,u)\leq(s^2-b)(u,u)$, so $(y,u)\geq\frac{b-s^2}{2(a-s)}(u,u)$ for $s\leq a$; the optimal choice is $s=a-\sqrt{a^2-b}$ (note that the first inequality implies $a^2-b\geq 0$). For example, let us apply the inequality to 3-graphs in which every 4-set spans at least t edges, where $t\in\{1,2,3\}$ (it turns out that we should choose the same t above). Then $Q_i=\binom{n}{4}$ for $0\leq i\leq t$. We apply the general inequality with $a=\frac{1}{2}((1+t/3)(n-3)+1)\sim\frac{1}{2}(1+t/3)n$ and $b=\sum_{i=1}^{t}(t+5-2i)\binom{n}{4}\binom{n}{2}^{-1}\sim\frac{1}{3}tn^2$. Then $s/n\sim(1+t/3)/2-\sqrt{(1+t/3)^2/4-t/3}=t/3$, so $|G|=\frac{1}{3}(y,u)\geq\frac{tn}{9}\binom{n}{2}\sim\frac{t}{3}\binom{n}{3}$. In particular we have re-proved the earlier example (in complementary form). An interesting additional feature of this method is that one can obtain a small improvement by exploiting integrality of the vectors y and u. Instead of adding the inequality $-(y-su,y-su)\leq 0$ above, one can use $-(y-\lfloor s\rfloor u, y-(\lfloor s\rfloor+1)u)\leq 0$. This does not alter the asymptotic bound, but can be used to improve Turán bounds for small n, which in turn can be used in other asymptotic arguments.

7 Flag algebras

A systematic approach to counting arguments is provided by the theory of flag algebras. This is abstract and difficult to grasp in full generality, but for many applications it can be boiled down to a form that is quite simple to describe. Our discussion here will be mostly based on the nice exposition given in [9]. The starting point is the following description of the averaging bound. Given r-graphs H and G we write $i_H(G)$ for the 'induced density' of H in G. This is defined as the probability that a random $v(H)$-set in $V(G)$ induces a subgraph isomorphic to H; we think of H as fixed and G as large. For example, if H is a single edge then $i_H(G)=d(G)$ is the edge density of G. Fix some $\ell\geq r$, and let $\mathcal{G}_\ell$ denote the set of all r-graphs on ℓ vertices (up to isomorphism). Then we can write $d(G)=\sum_{H\in\mathcal{G}_\ell} i_H(G)d(H)$. Now suppose F is an r-graph and let $\mathcal{F}_\ell$ denote the set of all F-free r-graphs on ℓ vertices.

If G is F-free then $i_H(G) = 0$ for $H \in \mathcal{G}_\ell \setminus \mathcal{F}_\ell$, so $d(G) = \sum_{H \in \mathcal{F}_\ell} i_H(G) d(H)$. In particular we have the averaging bound $d(G) \le \max_{H \in \mathcal{F}_\ell} d(H)$, but this is generally rather weak. The idea of the method is to generate further inequalities on the densities $i_H(G)$ that improve this bound. If we have an inequality $\sum_{H \in \mathcal{F}_\ell} c_H i_H(G) \ge 0$ then we have $d(G) \le \sum_{H \in \mathcal{F}_\ell} i_H(G)(d(H) + c_H) \le \max_{H \in \mathcal{F}_\ell}(d(H) + c_H)$, which may be an improvement if some coefficients c_H are negative.

These inequalities can be generated by arguments similar to that used on the sum $S = \sum_{xy} \binom{d(x,y)}{2}$ in Section 6 when F is the 3-graph with 4 vertices and 3 edges. In this context, one should view that argument as generating an inequality for a 4-vertex configuration (two edges) from two 3-vertex configurations (edges) that overlap in two points. We will use this as a running example to illustrate the flag algebra definitions. In general, we will consider overlapping several pairs of r-graphs along a common labelled subgraph. To formalise this, we define a *type* σ to consist of an F-free r-graph on k vertices together with a bijective labelling function $\theta : [k] \to V(\sigma)$ for some $k \ge 0$ (if $k < r$ then σ has no edges, and if $k = 0$ it has no vertices). Then we define a σ-*flag* to be an F-free r-graph H containing an induced copy of σ, labelled by θ. In our example we take $\sigma = xy$ to be a 3-graph with 2 verties and no edges, labelled as $\theta(1) = x$ and $\theta(2) = y$. Then we take $H = e^\sigma$ to be a single edge xyz, with the same labelling $\theta(1) = x$ and $\theta(2) = y$.

Next we define induced densities for flags. Let Φ be the set of all injective maps $\phi : [k] \to V(G)$. For any fixed $\phi \in \Phi$ we define $i_{H,\phi}(G)$ as the probability that a random $v(H)$-set S in $V(G)$ containing the image of ϕ induces a σ-flag isomorphic to H. Note that $i_{H,\phi}(G)$ can only be non-zero if $\phi([k])$ induces a copy of σ that can be identified with σ in such a way that $\phi = \theta$. If this holds, then $i_{H,\phi}(G)$ is the probability that $G[S]$ induces a copy of the underlying r-graph of H consistent with the identification of $\phi([k])$ with σ. We can relate flag densities to normal densities by averaging over $\phi \in \Phi$; we have $\mathbb{E}_\phi i_{H,\phi}(G) = p_\sigma(H) i_H(G)$, where we also let H denote the underlying r-graph of the σ-flag H, and $p_\sigma(H)$ is the probability that a random injective map $\phi : [k] \to V(H)$ gives a copy of the type σ. In our example, for any $u, v \in V(G)$ we consider the function ϕ defined by $\phi(1) = u$ and $\phi(2) = v$; we denote this function by uv. Then $i_{e^\sigma,uv}(G)$ is the probability that a random vertex $w \in V(G) \setminus \{u, v\}$ forms an edge with uv, i.e. $i_{e^\sigma,uv}(G) = \frac{d(u,v)}{n-2}$. Then we have $\mathbb{E}_{uv} i_{e^\sigma,uv}(G) = \frac{1}{n(n-1)} \sum_u \sum_{v \ne u} \frac{d(u,v)}{n-2} = \frac{6e(G)}{n(n-1)(n-2)} = d(G) = i_e(G)$; note that $p_\sigma(e) = 1$.

Given two σ-flags H and H' we have the approximation $i_{H,\phi}(G) i_{H',\phi}(G) = i_{H,H',\phi}(G) + o(1)$, where we define $i_{H,H',\phi}(G)$ to be the probability that when we independently choose a random $v(H)$-set S and a random $v(H')$-set S' in $V(G)$ subject to $S \cap S' = \phi([k])$ we have $G[S] \cong H$ and $G[S'] \cong H$ as σ-flags. Here the $o(1)$ term tends to zero as $v(G) \to \infty$: this expresses the fact that random embeddings of H and H' are typically disjoint outside of $\phi([k])$. Note that we can compute $i_{H,H',\phi}(G)$ by choosing a random ℓ-set L containing $S \cup S'$ for some $\ell \ge v(H) + v(H') - k$ and conditioning on the σ-flag J induced by L. Writing $\mathcal{F}^\sigma_\ell$ for the set of σ-flags on ℓ vertices we have $i_{H,H',\phi}(G) = \sum_{J \in \mathcal{F}^\sigma_\ell} i_{H,H',\phi}(J) i_{J,\phi}(G)$. Thus we can express $i_{H,H',\phi}(G)$ as a linear combination of flag densities $i_{J,\phi}(G)$, where the coefficients $i_{H,H',\phi}(J)$ are given by a finite computation.

In our running example we consider $H = H' = e^\sigma$. Then $i_{e^\sigma,e^\sigma,uv}(G)$ is the probability that a random pair w, w' of vertices in $V(G) \setminus \{u, v\}$ each form an edge with uv, i.e. $i_{e^\sigma, e^\sigma, uv}(G) = \binom{d(u,v)}{2}\binom{n-2}{2}^{-1}$. We can also compute this by conditioning on $J = G[L]$ for a random 4-set L containing uv. Let J denote the σ-flag on 4 vertices where 2 vertices are labelled x and y, and there are 2 edges, both containing x and y. Then $i_{e^\sigma, e^\sigma, uv}(G) = i_{J,uv}(G)$; note that this uses our particular choice of F, otherwise we would have additional terms corresponding to the 3-edge and 4-edge configurations. Averaging over uv we get $\frac{1}{n(n-1)} \sum_u \sum_{v \neq u} \binom{d(u,v)}{2}\binom{n-2}{2}^{-1} = p_\sigma(J) i_J(G) = \frac{1}{6} i_J(G)$, i.e. $i_J(G) = \binom{n}{4}^{-1} \sum_{uv} \binom{d(u,v)}{2}$, similarly to the calculation of $S = \sum_{uv} \binom{d(u,v)}{2}$ in Section 6.

Now we consider several σ-flags $F_1, \cdots, F_m$, assign them some real coefficients $a_1, \cdots, a_m$ and consider the inequality $(\sum_{i=1}^m a_i i_{F_i,\phi}(G))^2 \geq 0$. Expanding the square and using the identity for overlapping flags we obtain $\sum_{i,j=1}^m a_i a_j i_{F_i,F_j,\phi}(G) \geq o(1)$. Then we convert this into an inequality for densities of subgraphs (rather than flags) by averaging over $\phi \in \Phi$. Provided that we have chosen $\ell \geq v(F_i) + v(F_j) - k$ we can compute the average $\mathbb{E}_\phi i_{F_i,F_j,\phi}(G)$ by choosing a random ℓ-set L and conditioning on the subgraph $H \in \mathcal{F}_\ell$ induced by L. Let Φ_H be the set of all injective maps $\phi : [k] \to V(H)$. Then $\mathbb{E}_\phi i_{F_i,F_j,\phi}(G) = \sum_{H \in \mathcal{F}_\ell} b_{ij}(H) i_H(G)$, where $b_{ij}(H) = \mathbb{E}_{\phi \in \Phi_H} i_{F_i,F_j,\phi}(H)$. In any application H and $F_1, \cdots, F_m$ are fixed small r-graphs, so these coefficients $b_{ij}(H)$ can be easily computed. Finally we have an inequality of the required form: $\sum_{H \in \mathcal{F}_\ell} c_H i_H(G) \geq o(1)$, where $c_H = \sum_{i,j=1}^m a_i a_j b_{ij}(H)$.

Continuing the previous example, let $\sigma = xy$, let F_0 be the σ-flag on 3 vertices with no edges, taken with coefficient $a_0 = -1$, and let $F_1 = e^\sigma$ be the σ-flag on 3 vertices with one edge, taken with coefficient $a_1 = 2$. There are three 3-graphs in $\mathcal{F}^4$, which we label H_0, H_1 and H_2 according to the number of edges. The coefficients $b_{ij}(H_k)$ may be computed as follows: $b_{00}(H_0) = 1$, $b_{00}(H_1) = 1/2$, $b_{00}(H_2) = 1/6$, $b_{01}(H_0) = 0$, $b_{01}(H_1) = 1/4$, $b_{01}(H_2) = 1/3$, $b_{11}(H_0) = 0$, $b_{11}(H_1) = 0$, $b_{11}(H_2) = 1/6$. Then we obtain the coefficients $c_{H_0} = 1$, $c_{H_1} = -1/2$, $c_{H_2} = -1/2$, so the inequality $i_{H_0}(G) - i_{H_1}(G)/2 - i_{H_2}(G)/2 \geq o(1)$. We also have the averaging identity $d(G) = i_{H_1}(G)/4 + i_{H_2}(G)/2$. Adding the inequality $o(1) \leq i_{H_0}(G)/3 - i_{H_1}(G)/6 - i_{H_2}(G)/6$ gives $d(G) \leq i_{H_0}(G)/3 + i_{H_1}(G)/12 + i_{H_2}(G)/3 \leq 1/3 + o(1)$, so we recover the previous bound.

The reader may now be thinking that this is more obscure than the earlier argument and we are no nearer to a systematic approach! Indeed, there is no obvious way to choose σ and the σ-flags $F_1, \cdots, F_m$; in results so far these have been obtained by guesswork and computer experimentation. However, once these have been fixed an optimal inequality can be determined by solving a semidefinite program. (We refer the reader to [122] for background information on semidefinite programming that we will use below.) We first remark that once we have chosen σ and the σ-flags $F_1, \cdots, F_m$ we can sum several inequalities of the form $(\sum_{i=1}^m a_i i_{F_i,\phi}(G))^2 \geq 0$. Equivalently, we can fix a positive semidefinite m by m matrix $Q = (q_{ij})_{i,j=1}^m$ and use the inequality $\sum_{i,j=1}^m q_{ij} i_{F_i,\phi}(G) i_{F_j,\phi}(G) \geq 0$. Averaging over ϕ gives $\sum_{H \in \mathcal{F}_\ell} c_H(Q) i_H(G) \geq o(1)$, where $c_H(Q) = \sum_{i,j=1}^m q_{ij} b_{ij}(H)$. We can write $c_H(Q) = Q \cdot B(H)$, where $B(H)$ is the matrix $(b_{ij}(H))_{i,j=1}^m$ (note that it is symmetric) and $X \cdot Y = tr(XY)$ denotes inner product of symmetric matrices. We obtain a bound $\pi(F) \leq V + o(1)$, where V is the solution of the following optimisation problem in the variables $\{q_{ij}\}_{i,j=1}^m$ and $\{i_H(G)\}_{H \in \mathcal{F}_\ell}$:

$$V = \inf_{Q \geq 0} \sup_{(i_H(G)) \in S} \sum_{H \in \mathcal{F}_\ell} (d(H) + c_H(Q)) i_H(G).$$

Here we write $Q \geq 0$ to mean that Q is positive semidefinite and S for the set of vectors $(i_H(G))$ with $i_H(G) \geq 0$ for all H and $\sum_{H \in \mathcal{F}_\ell} i_H(G) = 1$. We can also compute V by the same expression in which the infimum and supremum are interchanged: this follows from Sion's minmax theorem, as S is compact and convex, and the set of positive semidefinite matrices is convex. Now the inner infimum over $Q \geq 0$ has a 'hidden constraint' in the term $\sum_H c_H(Q) i_H(G) = Q \cdot \sum_H i_H(G) B(H)$; this term is unbounded below unless $\sum_H i_H(G) B(H) \geq 0$, in which case it is minimised at 0 when $Q = 0$. Here we are using the fact that $X \geq 0$ if and only if $X \cdot Y \geq 0$ for all $Y \geq 0$. Thus we can express our bound in the following semidefinite program:

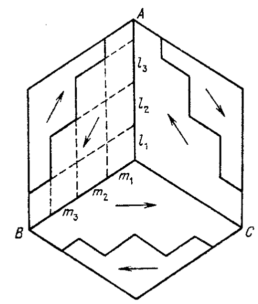
$$V = \max \sum_{H \in \mathcal{F}_\ell} d(H) i_H(G) \quad \text{s.t.} \quad (i_H(G)) \in S \quad \text{and} \quad \sum_H i_H(G) B(H) \geq 0.$$

This formulation is amenable to solution by computer, at least when ℓ is small, so that $\mathcal{F}_\ell$ is not too large. Razborov [158] has applied this method to re-prove many results on Turán densities that were obtained by other methods, and to obtain the sharpest known bounds for several Turán problems that are still open. At this point we recall the question of Turán mentioned earlier: what is the largest 3-graph on n vertices with no tetrahedron? Turán proposed the following construction: form a balanced partition of the n vertices into sets V_0, V_1, V_2, and take the edges to be those triples that either have one vertex in each part, or have two vertices in V_i and one vertex in V_{i+1} for some i, where $V_3 := V_0$. One can check that this construction gives a lower bound $\pi(K_3^4) \geq 5/9$. Until recently, the best known upper bound was $\pi(K_4^3) \leq \frac{3+\sqrt{17}}{12} = 0.593592\cdots$, given by Chung and Lu [32]. Razborov [158] announced computations that suggest the bound $\pi(K_4^3) \leq 0.561666$. These computations were also verified in [9], so the bound is probably correct. Here we should stress that, unlike some computer-aided mathematical arguments where there is potential to doubt whether they are really 'proofs', the flag algebra computations described here can in principle be presented in a form that can be verified (very tediously) by hand. However, there is not much point in doing this for a bound that is very unlikely to be tight, so this bound is likely to remain unrigorous!

The main result of [158] is the following asymptotic result for the tetrahedron problem under an additional assumption: if G is a 3-graph on n vertices in which no 4-set of vertices spans exactly 1 edge or exactly 4 edges then $e(G) \leq (5/9 + o(1))\binom{n}{3}$. This is given by an explicit flag computation that can be verified (laboriously) by hand.[4] This was subsequently refined by Pikhurko [154] using the stability method to give an exact result: for large n, the Turán construction for no tetrahedron gives the unique largest 3-graph on n vertices in which no 4-set of vertices spans exactly 1 edge or exactly 4 edges. It is interesting to contrast the uniqueness and stability of this restricted problem with the full tetrahedron problem, where there are exponentially many constructions that achieve the best known bound, see [26, 117, 64, 78].

[4] It is easy to verify that the matrices used in the argument are positive definite without the floating point computations referred to in [158]; for example one can just verify that all symmetric minors have positive determinant (symmetric Gaussian elimination is a more efficient method).

We briefly describe the elegant Fon-der-Flaass construction [64] here. Suppose that Γ is an oriented graph with no induced oriented 4-cycle. Let G be the 3-graph on the same vertex set in which abc is an edge if it induces a subgraph of Γ which either has an isolated vertex or a vertex of outdegree 2. It is not hard to see that any 4-set contains at least one edge of G, so the complement of G is K_4^3-free. The picture (from [64]) illustrates a suitable class of orientations of a complete tripartite graph. The parts A, B, C are represented by line segments. They are partitioned into subparts, represented by subsegments, e.g. A is partitioned into parts labelled by $\ell_1, \ell_2, \cdots$. The direction of edges is represented by an arrow in the appropriate rhombus. A weakened form of Turán's conjecture raised in [64] is whether all constructions of this type have density at least 4/9. Some progress towards this was recently made by Razborov [159].



It is natural to ask whether all hypergraph Turán problems can be solved using flag algebras (at least in principle, given enough computation). A related general question posed by Lovász [123] is whether every linear inequality $\sum_H c_H i_H(G) \geq 0$ valid for all graphs G can be expressed as a finite sum of squares. Razborov [156] posed a similar question in terms of a certain 'Cauchy-Schwartz' calculus. Both these questions were recently answered in the negative by Hatami and Norine [92]. Furthermore, they proved that the problem of determining the validity of a linear inequality is undecidable. Their proof is a reduction to Matiyasevich's solution to Hilbert's tenth problem, which shows undecidability of the problem of determining whether a multivariate polynomial with integer coefficients is non-negative for every assignment of integers to its variables. There does not seem to be a direct consequence of their results for deciding inequalities of the form $a\pi(F) \leq b$ for r-graphs F and integers a and b, but perhaps this problem may also be 'difficult', or even undecidable.

8 The remaining exact results

Exact results for hypergraph Turán numbers are so rare that we can finish off a description of the known results in this section. By exact results we mean that the Turán number is determined for large n (it would be of course be nice to know it for all n, but then this section would already be finished!) We have mentioned earlier the exact results for F_5 (which implies that for cancellative 3-graphs), the Fano plane, extended complete graphs and generalised fans. Sidorenko [175, 176] and Frankl [65] considered the Turán problem for the following $2k$-graph which we call the *expanded triangle* C_3^{2k}. The vertex set is $K_1 \cup K_2 \cup K_3$ where K_1, K_2 and K_3 are disjoint k-sets, and there are three edges $K_1 \cup K_2$, $K_1 \cup K_3$ and $K_2 \cup K_3$. Thus the expanded triangle is obtained from a graph triangle by expanding each vertex into a k-set. Suppose that G is a $2k$-graph on n vertices with no expanded triangle. There is a natural auxiliary graph J on k-sets of vertices, where we join two k-sets in J if their union is an edge of G. Then J is triangle-free, and applying Mantel's theorem gives the bound $\pi(C_3^{2k}) \leq 1/2$. For a construction, consider a partition of n vertices into two roughly equal parts, and take the edges to be all

$2k$-sets that intersect each part in an odd number of vertices: we call this *complete oddly bipartite*. To see that this does not contain the expanded triangle, consider an attempted embedding and look at the intersection sizes of the k-sets K_1, K_2 and K_3 with one of the parts. Some two of these have the same parity, so combine to form an edge with an even intersection with this part. This gives a lower bound that matches the upper bound asymptotically, so $\pi(C_3^{2k}) = 1/2$.

Keevash and Sudakov [111] proved an exact result, confirming a conjecture of Frankl, that for large n, the unique largest $2k$-graph on n vertices with no expanded triangle is complete oddly bipartite. One should note that the number of edges is this construction is maximised by a partition that is slightly unbalanced (by an amount of order $\sqrt{n}$). Finding the optimal partition sizes is in fact an open problem, equivalent to finding the minima of binary Krawtchouck polynomials. Nevertheless, known bounds on this problem are sufficient to allow an application of the stability method, with the conclusion that the optimal construction is complete oddly bipartite, even if we do not know the exact part sizes. Sidorenko also considered the *expanded clique* C_r^{2k}, obtained by expanding each vertex of K_r into a k-set. Applying Turán's theorem to the auxiliary graph J on k-sets gives $\pi(C_r^{2k}) \leq \frac{r-2}{r-1}$. On the other hand, Sidorenko only gave an asymptotic matching lower bound in the case when r is of the form 2^p+1. The construction is to partition n vertices into 2^p parts, labelled by the vector space $\mathbb{F}_2^p$, and take edges to be all $2k$-sets whose labels have a non-zero sum. This is C_r^{2k}-free, as for any r k-sets, the labels of some two will have the same sum (by the pigeonhole principle), so form an edge whose labels sum to zero. This raised the question of what happens for r not of this form. One might think that a combinatorial problem of this nature will not depend on a number theoretic condition, so there ought to be other constructions. However, we showed in [111] that this is not the case, in a somewhat different application of the stability method. We studied structural properties of a putative C_r^4-free 4-graph with density close to $\frac{r-2}{r-1}$ and showed that they give rise to certain special proper edge-colourings of K_{r-1}. It then turns out that these special edge-colourings have a natural $\mathbb{F}_2$ vector space structure on the set of colours, so we get a contradiction unless r is of the form 2^p+1.

It would be interesting to give better bounds for other r. We present here a new construction showing that $\pi(C_4^4) \geq 9/14 = 0.6428\cdots$; this is not far from the upper bound of $2/3$ (which we know is not sharp). Partition a set V of n vertices into sets A, B, C, where A is further partitioned as $A = A_1 \cup A_2$. We will optimise the sizes of these sets later. We say that $S \subseteq V$ has *type* ijk if $|S \cap A| = i$, $|S \cap B| = j$, $|S \cap A| = k$. Let G be the 4-graph in which 4-tuples of the following types are edges: (i) all permutations of 310, (ii) 121 and 112 (but not 211), (iii) 400 with 3 vertices in one A_i and 1 vertex in the other. We claim that G is C_4^4-free. To see this, suppose for a contradiction that we can choose 4 pairs of vertices such that any pair of these pairs forms an edge of G. We can naturally label each pair as AA, BB, CC, AB, AC or BC. Since every pair of pairs forms an edge, we see that each label apart from AA can occur at most once, and at most one of AA, BB, CC can occur. Now consider cases according to how many times AA occurs. If AA does not occur, then BB and CC can account for at most one pair, so the other 3 pairs must be AB, AC, BC; however, AB and AC do not form an edge. If AA occurs once or twice then BB, CC and BC cannot occur, so we must have AB and AC; however, again these do not form an edge. If AA occurs at least 3 times then some two AA's do not form an edge, as the restriction of G to A is C_3^4-free. In all cases we have a contradiction, so G is C_4^4-free. Computations show that the optimal set sizes are $|A_1| = |A_2| = (7-\sqrt{21})/28$ and $|B| = |C| = (7+\sqrt{21})/28$, and that then G has density $9/14$. Since this relatively simple construction gives a density quite close to the easy upper bound, we conjecture that it is optimal.

The question of improving the auxiliary graph bound described above gives rise to the following 'coloured Turán problem' that seems to be of independent interest. (A similar problem is also discussed in Section 13.4.) Suppose H is a C_4^4-free 4-graph on n vertices with $\alpha\binom{n}{4}$. Let J be the K_4-free auxiliary graph on pairs: J has $N = \binom{n}{2}$ vertices and $\sim \alpha\binom{N}{2}$ edges. If α is close to $2/3$ then for a 'typical' triangle xyz in J the common neighbourhoods N_{xy}, N_{xz}, N_{yz} partition most of $V(J)$ into 3 independent sets. These must account for almost all of the missing edges. Going back to the original set of n vertices, we can interpret N_{xy}, N_{xz}, N_{yz} as a 3-edge-colouring of (most of) the $\binom{n}{2}$ pairs, such that any choice of a pair of disjoint pairs of the same colour gives a non-edge of H. Counting then implies that almost all non-edges are properly 3-edge-coloured. It is not hard to see that this is impossible, but the question is to quantify the extent to which this property is violated. For example, what is the minimum number of 4-cycles in which which precisely one pair of opposite edges has both edges of the same colour?

Now we will return to cancellative r-graphs. Bollobás conjectured that the natural generalisation of his theorem for cancellative 3-graphs should hold, namely that the largest cancellative r-graph on n vertices should be r-partite. This was proved by Sidorenko [173] in the case $r = 4$. Note that there are four configurations that are forbidden in a cancellative 4-graph, there is the 4-graph expanded triangle mentioned above, and another three which are formed by taking two edges as $abcx$, $abcy$ and a third edge that contains xy and intersects abc in either 0, 1 or 2 vertices. Sidorenko showed that even just forbidding these last three configurations but allowing the expanded triangle one obtains the same result, that the largest such 4-graph is 4-partite. This was further refined by Pikhurko [152] who showed that it is enough to forbid just one configuration, the *generalised triangle* $\{abcx, abcy, uvxy\}$: for large n, the largest 4-graph with no generalised triangle is 4-partite.

Sidorenko's argument is an instructive application of hypergraph Lagrangians. We will sketch the proof that if G is a cancellative 4-graph then $e(G) \le (n/4)^4$, which is tight when n is divisible by 4. Since $e(G)/n^4 = p_G(1/n, \cdots, 1/n) \le \lambda(G)$ it suffices to prove that the Lagrangian $\lambda(G)$ is at most 4^{-4}. We can assume that G covers pairs; then it follows that any two vertices in G have disjoint link 3-graphs (or we get one of Sidorenko's three forbidden configurations). Recall that $\lambda(G) = \max_{x \in S} p_G(x)$ where S is the set of all $x = (x_1, \cdots, x_n)$ with $x_i \ge 0$ for all i and $\sum_i x_i = 1$. Suppose that the maximum occurs at some x with $x_i > 0$ for $1 \le i \le m$ and $x_i = 0$ for $i > m$ (without loss of generality). We can discard all $i > m$ and then regard the maximum as being at an interior point of the corresponding region S_m defined for the vector $x = (x_1, \cdots, x_m)$. Next comes an ingredient from the theory of optimisation we have not yet mentioned: the gradient of $p_G(x)$ is normal to the constraint plane $\sum_i x_i = 1$, i.e. the partial derivatives $\partial_i p_G(x)$, $1 \le i \le m$ are all equal to some constant c. We can compute it by $c = c\sum_i x_i = \sum x_i \partial_i p_G(x) = r p_G(x) = r\lambda(G)$. Also, since vertices have disjoint links, any monomial $x_a x_b x_c$ occurs at most once in $\sum_i \partial_i p_G(x)$, so $mr\lambda(G) = \sum_i \partial_i p_G(x) \le \max_{x \in S} \sum_{1 \le a<b<c<m} x_a x_b x_c = m^{-3}\binom{m}{3}$. This gives the required bound if $m = 4$ or $m \ge 6$, and we cannot have $m = 5$ if G covers pairs, so we are done.

Pikhurko's proof is an ingenious combination of Sidorenko's argument with the stability method. And what happens for larger r? Shearer [170] showed that the Bollobás conjecture is false for $r \ge 10$. The intermediate values are still open. Suppose now that we alter the general problem and just forbid the configurations analogous to those in Sidorenko's result; thus we consider r-graphs such that there do not exist two edges that share an $(r-1)$-set T and a third edge containing the two vertices not in T. When $r = 5$ and $r = 6$ this problem was solved by Frankl and Füredi [70]: the extremal constructions are the blowups of the small Witt designs, the $(11, 5, 4)$-design for $r = 5$ and the $(12, 6, 5)$-design for $r = 6$. They obtain the bounds $e(G) \le \frac{6}{11^4} n^5$ when $r = 5$, with equality only when $11|n$, and $e(G) \le \frac{11}{12^5} n^6$ when $r = 6$, with equality only when $12|n$. (The exact result for all large n remains open.) The proofs involve some intricate computations with hypergraph Lagrangians. They make the following appealing conjecture that would have greatly simplified some of these computations were it known. Consider the problem of maximising the Lagrangian $\lambda(G)$ for r-graphs G on m edges. Is the maximum attained when G is an initial segment of the colexicographic order?

Some further exact results can be grouped under the general umbrella of 'books'. The *r-book with p pages* is the r-graph obtained by taking $p \le r$ edges that share a common $(r-1)$-set T, and one more edge that is disjoint from T and contains the vertices not in T. For example, the generalised triangle of Pikhurko's result mentioned above is the 4-book with 2 pages. Füredi, Pikhurko and Simonovits [83] gave an exact result for the 4-book with 3 pages: for large n the unique extremal 4-graph is obtained by a balanced partition into two parts and taking edges as all 4-sets with 2 vertices in each part. Next consider r-graphs that do not have an r-book with r pages. A nice reformulation of this property is to say that such r-graphs G have *independent neighbourhoods*: for any $(r-1)$-set T, the neighbourhood $N(T) = \{x \in V(G) : T \cup \{x\} \in E(G)\}$ does not contain any edges of G. Füredi, Pikhurko and Simonovits [82] gave an exact result for 3-graphs with independent neighbourhoods: for large n, the unique extremal 3-graph is obtained by taking a partition into two parts A, B and taking edges as all 3-sets with 2 vertices in A and 1 vertex in B (the optimal class sizes are $|A| = 2n/3$, $|B| = n/3$ when n is divisible by 3). Füredi, Mubayi and Pikhurko [81] gave an exact result for 4-graphs with independent neighbourhoods: for large n, the unique extremal 4-graph is complete oddly bipartite (the same construction as for the expanded triangle). There is a conjecture in [81] for general r that the largest r-graphs with independent neighbourhoods are obtained by a partition into two parts A, B and taking edges as all r-sets that intersect B in an odd number of vertices, but are not contained in B. The results mentioned above confirm this for $r = 3$ and $r = 4$. However, the conjecture was disproved for $r \ge 7$ by Bohman, Frieze, Mubayi and Pikhurko [17]. The conjecture would have implied that r-graphs with independent neighbourhoods have edge density at most 1/2. In fact, the construction in [17]. shows that the maximum edge density is roughly $1 - \frac{2\log r}{r}$, which approaches 1 for large r. The authors of [17] believe that the conjecture is true for $r = 5$ and $r = 6$.

There is one more exact result (to the best of this author's knowledge). We ask the expert readers to take note, as it seems to be have been overlooked in earlier bibliographies on this subject. The motivating problem is the Turán problem for K_5^3, where Turán conjectured that the complete bipartite 3-graph gives the extremal construction. This was disproved for $n = 9$ by Surányi (the affine plane over $\mathbb{F}_3$) and for all odd $n \geq 9$ by Kostochka and Sidorenko (see construction 5 in [180]). However, they did not disprove the asymptotic conjecture, so it may be that $\pi(K_5^3) = 3/4$. Zhou [192] obtained an exact result when one forbids a larger class of 3-graphs that includes K_5^3. Say that two vertices x, y in a 3-graph G are *t-connected* if there are vertices a, b, c such that every triple with 2 vertices from abc and 1 from xy is an edge. Say that xyz is a *t-triple* if xyz is an edge and each pair in xyz is t-connected. For example, K_5^3 is a t-triple. The result of [192] is that the unique largest 3-graph on n vertices with no t-triple is complete bipartite. Note that the 3-graph $F(3,3)$ mentioned earlier is an example of a t-triple, so the result of [143] strengthens the asymptotic form of Zhou's result (but not the exact result).

9 Bounds for complete hypergraphs

We return now to the original question of Turán, concerning the Turán numbers for the complete hypergraphs K_t^r. None of the Turán densities $\pi(K_t^r)$ with $t > r > 2$ has yet been determined, so here we have a more modest goal of giving reasonable bounds. Most of these can be found in an excellent survey of Sidorenko [180], to which we refer the reader for full details. We will not reproduce this here, but instead outline the ideas behind the main bounds, summarise the other bounds, and also mention a few more recent developments. For the purpose of this section it is convenient to change to the 'complementary' notation that was preferred by many early writers on Turán numbers. They define the 'Turán number' $T(n, k, r)$ to be the minimum number of edges in an r-graph G on n vertices such that any subset of k vertices contains at least one edge of G. Note that G has this property if and only if the 'complementary' r-graph of r-sets that are not edges of G is K_k^r-free; thus $T(n, k, r) + \mathrm{ex}(n, K_k^r) = \binom{n}{r}$. They also define the density $t(k, r) = \lim_{n\to\infty} \binom{n}{r}^{-1} T(n, k, r)$; thus $t(k, r) + \pi(K_k^r) = 1$.

We start with the lower bound on $t(k, r)$, which is equivalent to an upper bound on $\pi(K_k^r)$. The trivial averaging argument gives $t(k, r) \geq \binom{k}{r}^{-1}$. In general, the best known bound is $t(k, r) \geq \binom{k-1}{r-1}^{-1}$, due to de Caen [38]. This follows from his exact bound of $T(n, k, r) \geq \frac{n-k+1}{n-r+1}\binom{k-1}{r-1}^{-1}\binom{n}{r}$. This in turn is deduced from a hypergraph generalisation of a theorem of Moon and Moser that relates the number of cliques of various sizes in a graph. Suppose that G is an r-graph on n vertices and let N_k be the number of copies of K_k^r in G. Then the inequality is

$$N_{k+1} \geq \frac{k^2 N_k}{(k-r+1)(k+1)}\left(\frac{N_k}{N_{k-1}} - \frac{(r-1)(n-k)+k}{k^2}\right), \tag{9.1}$$

provided that $N_{k-1} \neq 0$. Given this inequality, the bound on $T(n, k, r)$ follows from some involved calculations; the main step is to show by induction on k that $N_k \geq N_{k-1}\frac{r^2\binom{k}{r}}{k^2\binom{n}{r-1}}(e(G) - F(n, k, r))$, where $F(n, k, r) = (r^{-1}(n-r+1) - \binom{k-1}{r-1}^{-1}(n-k+1))\binom{n}{r-1}$. Inequality (9.1) is proved by the following double counting argument.

Let P be the number of pairs (S,T) where S and T are each sets of k vertices, such that S spans a K^r_k, T does not span a K^r_k, and $|S \cap T| = k-1$. For an upper bound on P, we enumerate the N_{k-1} copies of K^r_{k-1} and let a_i be the number of K^r_k's containing the ith copy. Since $\sum_{i=1}^{N_{k-1}} a_i = kN_k$ we have $P = \sum_{i=1}^{N_{k-1}} a_i(n-k+1-a_i) \le (n-k+1)kN_k - N_{k-1}^{-1}k^2N_k^2$. For a lower bound, we enumerate the copies of K^r_k as $B_1, \cdots, B_{N_k}$, and let b_i be the number of K^r_{k+1}'s containing the ith copy. For each B_j, there are $n-k-b_j$ ways to choose $x \notin B_j$ such that $B_j \cup \{x\}$ does not span a K^r_{k+1}. Given such an x, there is some $C \subseteq B_j$ of size $k-1$ such that $C \cup \{x\}$ is not an edge. Then for each $y \in B_j \setminus C$ the pair $(B_j, B_j \cup x \setminus y)$ is counted by P. This gives $P \ge \sum_{j=1}^{N_k}(n-k-b_j)(k-r-1) = (k-r-1)((n-k)N_k - (k+1)N_{k+1}$. Combining with the lower bound and rearranging gives the required inequality.

Next we consider the upper bound on $t(k,r)$, which is equivalent to a lower bound on $\pi(K^r_k)$. The best general construction is due to Sidorenko [171]; it implies the bound $t(k,r) \le \left(\frac{r-1}{k-1}\right)^{r-1}$. For comparison with the lower bound, note that $\left(\frac{r-1}{k-1}\right)^{r-1}\binom{k-1}{r-1} = \prod_{i=1}^{r-1}\frac{k-i}{k-1}\frac{r-1}{r-i}$; if k is large compared to r, the ratio of the bounds is roughly $(r-1)^{r-1}(r-1)!^{-1}$, which is exponential in r, but independent of k. To explain the construction, we will rephrase it here using the following simple fact.

The lorry driver puzzle. A lorry driver needs to follow a certain closed route. There are several petrol stations along the route, and the total amount of fuel in these stations is sufficient for the route. Show that there is some starting point from which the route can be completed.

The construction is to divide n vertices into $k-1$ roughly equal parts $A_1, \dots, A_{k-1}$, and say a set B of size r is an edge of G if there is some j such that $\sum_{i=1}^{s}|B \cap A_{j+i}| \ge s+1$ for each $1 \le s \le r-1$ (where $A_i := A_{i-k+1}$ if $i > k-1$). To interpret this in the lorry driver framework, consider any set K of size k, imagine that each element of K represents a unit of fuel, and that it takes $\frac{k}{k-1}$ units of fuel to drive from A_i to A_{i+1}. Then K contains enough fuel for a complete circuit, so the lorry driver puzzle tells us that there is some starting point from which a complete circuit is possible. (For completeness we now give the solution to the puzzle. Imagine that the driver starts with enough fuel to drive around the route and consider the journey starting from an arbitrary point, in which she still picks up all the fuel at any station, even though she doesn't need it. Then the point at which the fuel reserves are lowest during this route can be used as a starting point for another route which satisfies the requirements.) Let B be the set of the first r elements of K that are encountered on this circuit (breaking ties arbitrarily). Since $r \ge (r-1)\frac{k}{k-1}$, the lorry can advance distance $r-1$ using just the fuel from B. This implies that B is an edge, as $\lceil s\frac{k}{k-1}\rceil = s+1$ for $1 \le s \le r-1$. Thus any set of size k contains an edge, as required.

It is not obvious how to estimate the number of edges in the construction without tedious calculations, so we will give a simple combinatorial argument here. It is convenient to count edges together with an order of the vertices in each edge, thus counting each edge $r!$ times. We can form an ordered edge $B = x_1 \dots x_r$ using the following three steps: (i) choose the starting index j, (ii) assign each x_ℓ to one of the parts A_{j+i}, $1 \le i \le r-1$, (iii) choose a vertex for each x_ℓ within its assigned part. Clearly there are $k-1$ choices in step (i) and $\left(\frac{n}{k-1}\right)^r + O(n^{r-1})$ choices in step (iii). In step (ii) there are $(r-1)^r$ ways to assign the parts if we ignore the required inequalities on the intersection sizes (i.e. that there should be enough fuel for the lorry). Now we claim that given any assignment, there is exactly one cyclic permutation that satisfies the required inequalities. More precisely, if we assign b_i of the x_ℓ's to A_{j+i} for $1 \le i \le r-1$, then there is exactly one c with $1 \le c \le r-1$ such that the shifted sequence $b'_i = b_{c+i}$ (where $b_i := b_{i-r+1}$ for $i > j+r-1$) satisfies $\sum_{i=1}^{s} b'_i \ge s+1$ for each $1 \le s \le r-1$. To see this consider a lorry that makes a circuit of A_{j+i}, $1 \le i \le r-1$, where as before each of the x_ℓ's is a unit of fuel, but now it takes one unit of fuel to advance from A_i to A_{i+1}, and the lorry is required to always have one spare unit of fuel. Clearly a valid starting point for the lorry is equivalent to a shifted sequence satisfying the required inequalities. As in the solution to the original puzzle, we imagine that the driver starts with enough fuel to drive around the route and consider the journey starting from an arbitrary point. Then the point at which the fuel reserves are lowest during this route is a starting point for a route where there is always one spare unit of fuel. Furthermore, this is the unique point at which the fuel reserves are lowest, and so it gives the unique cyclic permutation satisfying the required inequalities. We deduce that there are $(r-1)^{r-1}$ valid assignments in step (ii). Putting everything together, the number of edges is $r!^{-1} \cdot (k-1) \cdot (r-1)^{r-1} \cdot (1+O(1/n)) \left(\frac{n}{k-1}\right)^r \sim \left(\frac{r-1}{k-1}\right)^{r-1} \binom{n}{r}$.

Having discussed the general case in detail, we now summarise some better bounds that have been found in specific cases. One natural case to focus on is $t(r+1,r) = 1 - \pi(K^r_{r+1})$. For large r, a construction of Sidorenko [181] gives the best known upper bound, which is $t(r+1,r) \le (1+o(1))\frac{\log r}{2r}$. Other known bounds are effective for small r; these are $t(r+1,r) \le \frac{1+2\ln r}{r}$ by Kim and Roush [114] and $t(2s+1,2s) \le 1/4 + 2^{-2s}$ by de Caen, Kreher and Wiseman [43]. On the other hand, the known lower bounds are very close to the bound $t(r+1,r) \ge 1/r$ discussed above in the general case. Improvements to the second order term were given by Chung and Lu [32], who showed that $t(r+1,r) \ge \frac{1}{r} + \frac{1}{r(r+3)} + O(r^{-3})$ when r is odd, and by Lu and Zhao [129], who obtained some improvements when r is even, the best of which is $t(r+1,r) \ge \frac{1}{r} + \frac{1}{2r^3} + O(r^{-4})$ when r is of the form $6k+4$. Thus the known upper and lower bounds are separated by a factor of $(1/2+o(1))\log r$. As a first step towards closing this gap, de Caen [41] conjectured that $r \cdot t(r+1,r) \to \infty$ as $r \to \infty$.

We have already discussed the known bounds for K^3_4 in Section 7. For K^4_5 the following nice construction was given by Giraud [85]. Suppose M is an m by m matrix with entries equal to 1 or 0. We define a 4-graph G on $n = 2m$ vertices corresponding to the rows and columns of M. Any 4-set of rows or 4-set of columns is an edge. Also, any 4-set of 2 rows and 2 columns inducing a 2 by 2 submatrix with even sum is an edge. We claim that any 5-set of vertices of G contains an edge. This is clear if we have at least 4 rows or at least 4 columns, so suppose without loss of generality that we have 3 rows and 2 columns. Then in the induced 3 by 2 submatrix we can choose 2 rows whose sums have the same parity, i.e. a 2 by 2 submatrix with even sum, which is an edge. To count edges in G, first note that we have $2\binom{m}{4}$ from 4-sets of rows and 4-sets of columns. Also, for any pair i,j of columns, we can divide the rows into two classes O_{ij} and E_{ij} according to whether the entries in columns i and j have odd or even sum. Then the number of 2 by 2 submatrices using columns

i and j with even sum is $\binom{|O_{ij}|}{2} + \binom{|E_{ij}|}{2} \ge 2\binom{m/2}{2}$. Furthermore, for some values of m there is a construction that achieves equality for every pair i, j: take a Hadamard matrix, i.e. a matrix with ± 1 entries in which every pair of columns is orthogonal, then replace the -1 entries by 0.

This shows that $t(5,4) \le \lim_{m\to\infty} \binom{2m}{4}^{-1} \left(2\binom{m}{4} + 2\binom{m/2}{2}\binom{m}{2}\right) = 5/16$; equivalently $\pi(K_5^4) \ge 11/16 = 0.6875$. Sidorenko [180] conjectured that equality holds. Markström [131] gave an upper bound $\pi(K_5^4) \le \frac{1753}{2380} = 0.73655\cdots$. This was achieved by an extensive computer search to find all extremal 4-graphs for $n \le 16$. Based on this evidence, he made the stronger conjecture that this construction (modified according to divisibility conditions) is always optimal for $n \ge 12$. Markström [133] has also compiled a web archive of small constructions for various hypergraph Turán problems. For K_5^3 the Turán numbers were computed for $n \le 13$ by Boyer, Kreher, Radziszowski and Sidorenko [25]. The collinear triples of points of the projective plane of order 3 form the unique 3-graph on 13 vertices such that every 5-set contains at least one edge. It follows that $t(5,3) \ge 52/\binom{13}{3} = 2/11$, i.e. $\pi(K_5^3) \le 9/11$. As we mentioned in Section 8, Turán conjectured that $\pi(K_5^3) = 3/4$, corresponding to the complete bipartite 3-graph.

More generally, Turán conjectured that $\pi(K_{t+1}^3) = 1 - (2/t)^2$. Together with Mubayi we found the following family of examples establishing the lower bound (previously unpublished). It is convenient to work in the complementary setting; thus we describe 3-graphs of density $(2/t)^2$ such that every $(t+1)$-set contains at least one edge. Let D be any directed graph on $\{1, \dots, t\}$ that is the vertex-disjoint union of directed cycles (we allow cycles of length 2, but not loops). Let $V_1, \cdots, V_t$ be a balanced partition of a set V of n vertices. Let G be the 3-graph on V where the edges consist of all triples that are either contained within some V_i or have 2 points in V_i and 1 point in V_j, for every directed edge (i,j) of D. Then G has $t\binom{n/t}{3} + t\binom{n/t}{2}n/t \sim (2/t)^2\binom{n}{3}$ edges. Also, if $S \subseteq V$ does not contain any edge of G, then S has at most 2 points in each part, and whenever it has 2 points in a part it is disjoint from the next part on the corresponding cycle, so we must have $|S| \le t$. Thus G has the required properties.

10 The infinitary perspective

A new perspective on extremal problems can be obtained by stepping outside of the world of hypergraphs on finite vertex sets, and viewing them as approximations to an appropriate 'limit object'. This often leads to more elegant formulations of results from the finite world, after one has put in the necessary theoretical ground work to make sense of the 'limit'. That alone may justify this perspective for those of a theoretical bent, though others will ask whether it can solve problems not amenable to finite methods. Since the theory itself is quite a recent development, it is probably too soon to answer this latter question, other than to say that elegant reformulations usually lead to progress in mathematics.

We will approach the subject by first returning to flag algebras (see Section 7), which we will now describe in the theoretical framework of [156]. Recall that the aim when applying flag algebras to Turán problems was to generate a 'useful' inequality of the form $\sum_H c_H i_H(G) \ge 0$, valid for any F-free r-graph G. We can package the coefficients c_H as a 'formal sum' $\sum_H c_H H$ in $\mathbf{R}\mathcal{F}$, by which we mean the real vector space of formal finite linear combinations of F-free r-graphs. We can think of any F-free r-graph G as acting on $\mathbf{R}\mathcal{F}$ via the map $\sum_H c_H H \mapsto \sum_H c_H i_H(G)$; we will identify this map with G. Our goal will be to understand F-free r-graphs purely as appropriate maps on formal finite linear combinations. First we note that certain elements always evaluate to zero, so they should be factored out. If H is an F-free r-graph and $\ell \geq v(H)$ then $i_H(G) = \sum_{J \in \mathcal{F}_\ell} i_H(J) i_J(G)$, so the linear combination $H - \sum_{J \in \mathcal{F}_\ell} i_H(J) J$ is mapped to zero by G. Let $\mathcal{K}$ be the subspace generated by all such combinations in the kernel, and let $\mathcal{A} = \mathbf{R}\mathcal{F}/\mathcal{K}$ be the quotient space. Now we make $\mathcal{A}$ into an algebra by defining a multiplication operator: we let $HH' = \sum_{J \in \mathcal{F}_{v(H)+v(H')}} i_{H,H'}(J) J$, where $i_{H,H'}(J)$ is the probability that when $V(J)$ is randomly partitioned as $S \cup S'$ with $|S| = v(H)$ and $|S'| = v(H')$ we have $J[S] \cong H$ and $J[S'] \cong H'$. (One needs to prove that this is well-defined.) Then we have $G(HH') = G(H)G(H') + o(1)$ when $v(G)$ is large, so the map G is an 'approximate homomorphism' from $\mathcal{A}$ to $\mathbf{R}$. One final property to bear in mind is that $G(H) := i_H(G)$ is always non-negative. (A similar construction gives rise to an algebra $\mathcal{A}^\sigma = \mathbf{R}\mathcal{F}^\sigma/\mathcal{K}^\sigma$ for any type σ; we have just described the case when $|\sigma| = 0$ for simplicity.)

Now we come to the point of the above discussion: it gives an approximate characterisation of F-free r-graphs, in the following sense. Given an r-graph G, we can identify the map $G : \mathbf{R}\mathcal{F} \to \mathbf{R}$ defined above with the vector $(i_H(G))_{H \in \mathcal{F}} \in [0,1]^{\mathcal{F}}$; we will also identify this vector with G. The space $[0,1]^{\mathcal{F}}$ is compact in the product topology, so any sequence of r-graphs contains a convergent subsequence. Let Φ be the set of homomorphisms ϕ from $\mathcal{A}$ to $\mathbf{R}$ such that $\phi(H) \geq 0$ for every H in $\mathcal{F}$. The following key result is Theorem 3.3 in [156]: for any convergent sequence of F-free r-graphs, the limit is in Φ; conversely, any element of Φ is the limit of some sequence of F-free r-graphs. (For simplicity, we have stated this result just for F-free r-graphs, but there is a much more general form that applies to flags in theories.) This result establishes a correspondence between the finite world inequalities $\sum_H c_H i_H(G) \geq o(1)$ for r-graphs G (which we were interested in above) and inequalities $\phi(\sum_H c_H H) \geq 0$ for ϕ in Φ in the infinitary world. In particular, the Turán density $\pi(F) = \limsup_{G \in \mathcal{F}} d(G)$ can be rewritten as $\pi(F) = \max_{\phi \in \Phi} \phi(e)$. Note the maximum value $\pi(F)$ is achieved by some extremal homomorphism $\phi \in \Phi$ (this is because Φ is compact and $\phi \mapsto \phi(e)$ is continuous). This permits 'differential methods' (see section 4.3 of [156]), i.e. deriving inequalities from the fact that any small perturbations of ϕ must reduce $\phi(e)$, which are potentially very powerful. For example, perturbation with respect to a single vertex is analogous to the deletion argument in Proposition 4.2, but general perturbations do not have any obvious analogue in the finite setting.

Graph limits were first studied by Borgs, Chayes, Lovász, Sós and Vesztergombi (2003 unpublished and [24]) and by Lovász and Szegedy [126]. A substantial theory has been developed since then, of which we will only describe a couple of ingredients here: a convenient description of limit objects and the equivalence of various notions of convergence. The starting point is a very similar notion of convergence to that used by [156]. Let $t_H(G)$ denote the *homomorphism density* of H in G, defined as the probability that a random map from V_H to V_G is a homomorphism. Say that a sequence $G_1, G_2, \dots$ is left-convergent if $t_H(G_i)$ converges for every H. (It is not hard to see via inclusion-exclusion that using induced densities is equivalent.) The limit objects can be described as *graphons*, which are symmetric measurable[5] functions $W : [0,1]^2 \to [0,1]$. For such a function we can define the homomorphism density of H in W as $t_H(W) = \int \prod_{ij \in E(H)} W(x_i, x_j) dx$, where $x = (x_1, \cdots, x_{v(H)})$ and the integral is over $[0,1]^{v(H)}$. We can recover $t_H(G)$ as a case of this definition by defining a graphon W_G as a step function based on the adjacency matrix of G. Label $V(G)$ by $[n]$, partition $[0,1]^2$ into n^2 squares of side $1/n$, and set W_G equal to 1 on subsquare (i,j) if ij is an edge, otherwise 0. Then $t_H(G) = t_H(W_G)$. The main result of [126] is that for any left-convergent sequence (G_i) there is a graphon W such that $t_H(G_i) \to t_H(W)$ for every graph H, and conversely, any graphon W can be obtained in this way from a left-convergent sequence (G_i). This gives an intuitive picture of a graph limit as an 'infinite adjacency matrix'. A more formal justification of this intuition is given by the *W-random graph* $G(n,W)$. This is a random graph on $[n]$ defined by choosing independent $X_1, \cdots, X_n$ uniformly in $[0,1]$ and connecting vertices i and j with probability $W(X_i, X_j)$. Corollary 2.6 of [126] shows that $G(n,W)$ converges to W with probability 1.

An alternative description of convergence is given by the cut-norm on graphons, defined by $\|W\|_\square = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W(x,y)\ dx\ dy \right|$. First we need to take account of the lack of uniqueness in graphons. Suppose $\phi : [0,1] \to [0,1]$ is a measure-preserving bijection and define W^ϕ by $W^\phi(x,y) = W(\phi(x), \phi(y))$. Then W^ϕ is equivalent to W in the sense that $t_H(W^\phi) = t_H(W)$ for any H. We define the cut-distance between two graphons as $\delta_\square(W, W') = \inf_\phi \|W^\phi - W'\|_\square$, where the infimum is over all measure-preserving bijections. Then another equivalent definition of convergence for (G_i), given in [24], is to say that $\delta_\square(W_{G_i}, W) \to 0$ for some graphon W. Furthermore, W is essentially unique, in that if $G_n \to W$ and $G_n \to W'$ then $\delta_\square(W, W') = 0$. The equivalence classes $[W] = \{W' : \delta_\square(W, W') = 0\}$ are named *graphits* by Pikhurko [153], in a paper that introduces an analytic approach to stability theorems. Theorem 15 in [153] contains a characterisation of stability that can be informally stated as follows: an extremal graph problem is stable if and only if there is a unique graphit that can be obtained by limits of approximately extremal graphs. (We say that an r-graph F is *stable* if for any $\epsilon > 0$ there is $\delta > 0$ and n_0 such that for any two F-free r-graphs G and G' on $n > n_0$ vertices, each having at least $(\pi(F) - \delta)\binom{n}{r}$ edges, we can obtain G' from G by adding or deleting at most ϵn^r edges.) The analytic proof of the Erdős-Simonovits stability theorem given in [153] is much more complicated than a straightforward approach, but may point the way to other stability results that cannot be obtained by simpler methods.

The limit theory above also has close connections with the theory of regularity for graphs and hypergraphs, which are explored by Lovász and Szegedy [127]. We start with the Szemerédi's Regularity Lemma, which is a fundamental tool in modern graph theory. Our discussion here will be rather brief; for an extensive survey we refer the reader to [116]. Roughly speaking, the regularity lemma allows any graph G to be approximated by a weighted graph R, in which the size of R depends only

[5]We will assume in this discussion that the reader is familiar with the basics of measure theory. A careful exposition for the combinatorial reader that fills in much of this background is given in [153]. Note also that we are using a more restricted definition of 'graphon' than the original definition given in [126].

the desired accuracy of approximation, but is independent of the size of G. A precise statement of the lemma (in its simplest form) is as follows: for any $\epsilon > 0$, there is a number $m = m(\epsilon)$, such that for any number n, and any graph G on n vertices, there is a partition of $V(G)$ as $V_1 \cup \cdots \cup V_r$ for some $r \le m$, so that all but at most ϵn^2 pairs of vertices belong to induced bipartite subgraphs $G_{ij} := G(V_i, V_j)$ that are 'ϵ-regular'. We have not yet defined 'ϵ-regular': this is a notion that captures the idea that the bipartite subgraph $G(V_i, V_j)$ looks like a random bipartite graph. The formal definition is as follows. Suppose G is a bipartite graph with parts A and B. The density of G is $d(G) = \frac{|E(G)|}{|A||B|}$. We say that G is *ϵ-regular* if for any $A' \subseteq A$, $B' \subseteq B$ with $|A'| > \epsilon|A|$, $|B'| > \epsilon|B|$, writing G' for the bipartite subgraph of G induced by A' and B' we have $d(G') = d(G) \pm \epsilon$. Note that the definition fits well with the randomness heuristic: standard large deviation estimates imply that a random bipartite graph is ϵ-regular with high probability.

After applying Szemerédi's Regularity Lemma, we can use the resulting partition to define an approximation of G. This is the *reduced graph* R, defined on the vertex set $[r] = \{1, \cdots, r\}$. The vertices of R correspond to the parts $V_1, \cdots, V_r$ (which are also known as *clusters*). The edges of R correspond to pairs of clusters that induce bipartite subgraphs that look random and are sufficiently dense: we fix a 'density parameter' d, and include an edge ij in R with weight $d_{ij} := d(G_{ij})$ whenever G_{ij} is ϵ-regular with $d_{ij} \ge d$. A key property of this approximation of G by R is that it satisfies a 'counting lemma', whereby the number of copies of any fixed graph in G can be accurately predicted by the weighted number of copies of this graph in R. For example, we have the following Triangle Counting Lemma. Suppose $1 \le i, j, k \le r$ and write $T_{ijk}(G)$ for the set of triangles in G with one vertex in each of V_i, V_j and V_k. Write $d_{ijk} = \frac{|T_{ijk}(G)|}{|V_i||V_j||V_k|}$ for the corresponding 'triangle density', i.e. the proportion of all triples with one vertex in each of V_i, V_j and V_k that are triangles. Suppose $0 < \epsilon < 1/2$ and G_{ij}, G_{ik} and G_{jk} are ϵ-regular. Then $d_{ijk} = d_{ij}d_{ik}d_{jk} \pm 20\epsilon$. Note that this corresponds well to the randomness intuition: if the graphs were indeed random, with each edge being independently selected with probability equal to the corresponding density, then the probability of any particular triple uvw being a triangle would be the product of the probabilities for each of its three pairs uv, uw, vw being edges. Furthermore, there is a Counting Lemma for general subgraphs along similar lines, which starts to indicate the connection with the notion of convergence discussed above using subgraph densities.

The following weaker form of the regularity lemma was obtained by Frieze and Kannan [77]. Given a partition P of $V(G)$ as $V_1 \cup \cdots \cup V_r$, for $S, T \subseteq V(G)$ we write $e_P(S,T) = \sum_{i,j=1}^r d_{ij}|V_i \cap S||V_j \cap T|$. Note that $e_P(S,T)$ is the expected value of $e_G(S,T)$ (the number of edges of between S and T) if each G_{ij} were a random bipartite graph of density d_{ij}. The result of [77] is that there is a partition P into $r \le 2^{2/\epsilon^2}$ classes such that $|e_G(S,T) - e_P(S,T)| \le \epsilon n^2$ for all $S, T \subseteq V(G)$. This is rather weaker than the regularity lemma, as it has replaced a uniformity condition holding locally for most pairs of classes by a global uniformity condition. The compensation is that the number of classes needed is much smaller, only an exponential function, as opposed to the tower bound that is necessary in the regularity lemma (see [87]). The weak regularity lemma can be reformulated in analytic language as follows (Lemma 3.1 of [127]): for any graphon W and $\epsilon > 0$ there is a graphon W' that is a step function with at most $2^{2/\epsilon^2}$ steps such that $\|W - W'\|_\square \le \epsilon$. The full regularity lemma, indeed even a stronger form due to Alon, Fischer, Krivelevich and Szegedy [4], can also be obtained from an analytic form. The key fact is that graphits form a compact metric space with the distance $\delta_\square$ defined above (Theorem 5.1 of [127]). This implies the following (Lemma 5.2 of [127]): let $h(\epsilon, r) > 0$ be an arbitrary fixed function; then for any ϵ there is $m = m(\epsilon)$ such that any graphon W can be written as $W = U + A + B$, where U is a step function with $r \le m$ steps, $\|A\|_\square \le h(\epsilon, r)$ and $\|B\|_1 \le \epsilon$. Informally, this says that one can change W by a small function B to obtain a function $U + A$ which corresponds to an extremely regular partition. Regular approximation results of this type were first obtained in [162, 187].

Regularity theory for hypergraphs is much more complicated, so we will only make a few comments here and refer the reader to the references for more information. The theory was first developed independently in different ways by Rödl et al. [148, 163, 162] and Gowers [89]. Alternative perspectives and refinements were given in [8, 94, 187, 188]. The analytic theory discussed above is generalised to hypergraphs by Elek and Szegedy [46, 47] as follows. For any sequence of r-graphs $G_1, G_2, \dots$ which is convergent in the sense that $t_H(G_i)$ converges for every r-graph H, there is a limit object W, called a *hypergraphon*, such that $t_H(G_i) \to t_H(W)$ for every r-graph H. Hypergraphons are functions of $2^r - 1$ variables, corresponding to the non-empty subsets of $[r]$, that are symmetric under permutations of $[r]$. The need for $2^r - 1$ variables (actually $2^r - 2$, as $[r]$ is unnecessary) reflects the fact that a complete theory of hypergraph regularity needs to consider the simplicial r-complex generated by an r-graph, and regularise k-sets with respect to $(k-1)$-sets for each $2 \le k \le r$ (see Section 5 of [88] for further explanation of this point).

In the case of 3-graphs we consider a function $W(x_1, x_2, x_3, x_{12}, x_{13}, x_{23})$ from $[0,1]^6$ to $[0,1]$ that is symmetric under permutations of 123. Given a fixed 3-graph H, we can define the homomorphism density of H in W similarly to above by $t_H(W) = \int \prod_{e \in H} W_e dx$, where $x = (x_1, \cdots, x_{v(H)})$ and the integral is over $[0,1]^{v(H)}$ as before, and W_e is evaluated according to some fixed labelling $e = e_1e_2e_3$ by $W_e = W(x_{e_1}, x_{e_2}, x_{e_3}, x_{e_1e_2}, x_{e_1e_3}, x_{e_2e_3})$. Similarly to the graph case, any 3-graph G can be realised by a hypergraphon W_G that is a step function (which need only depend on the first 3 co-ordinates). Some intuition for hypergraphons can be obtained by consideration of the W-random 3-graph $G(n, W)$. This can be defined as a random 3-graph on $[n]$ by choosing independent random variables X_i, $1 \le i \le n$ and X_{ij}, $1 \le i < j \le n$ uniformly in $[0,1]$ and including the edge ijk for $i < j < k$ with probability $W(X_i, X_j, X_k, X_{ij}, X_{ik}, X_{jk})$. Theorem 12 of [47] shows that $G(n, W)$ converges to W with probability 1. If we approximate W by a step function then this gives us the following informal picture of a regularity partition of a 3-graph G: a piece of the partition is obtained by taking three classes V_i, V_j, V_k, then three 'random-like' bipartite graphs $V_{ij} \subseteq V_i \times V_j$, $V_{ik} \subseteq V_i \times V_k$, $V_{jk} \subseteq V_j \times V_k$, and then a 'random-like' subset of the triangles formed by V_{ij}, V_{ik} and V_{jk}. Further generalisations of the theory from graphs to hypergraphs given in [47] are the equivalence of various definitions of convergence (Theorem 14), and a formulation of regularity as compactness (Theorem 4).

We conclude this section with a concrete situation where hypergraph regularity theory gives some insight into Turán problems. This is via the removal lemma, a straightforward consequence of hypergraph regularity theory that can be easily stated as follows. For any $b > 0$ and r-graph F there is $a > 0$, so that if G is a r-graph on n vertices with fewer than $an^{v(F)}$ copies of F, then one can delete at most bn^r edges from G to obtain an F-free r-graph. This was used by Pikhurko [155, Lemma 4] to show that if the Turán problem for F is stable then so is the Turán problem for any blowup $F(t)$. A sketch of the proof is as follows. Suppose that $0 < 1/n \ll a \ll b \ll c$, and G_1 and G_2 are $F(t)$-free r-graphs on n vertices, each having at least $(\pi(F(t))-b)\binom{n}{r}$ edges. Since G_1 and G_2 are $F(t)$-free, supersaturation implies that they each have at most $an^{v(F)}$ copies of F. The removal lemma implies that one can delete at most bn^r edges from G_1 and G_2 to obtain F-free r-graphs G_1' and G_2', each having at least $(\pi(F) - 2b)\binom{n}{r}$ edges (recall that $\pi(F(t)) = \pi(F)$). Now by stability of F we can obtain G_2' from G_1' by adding or deleting at most cn^r edges. Thus we can obtain G_2 from G_1 by adding or deleting at most $2cn^r$ edges, so $F(t)$ is stable. In particular, this enables the application of the stability method in [155] to the extended complete graph H_t^r (see Section 3); stability of H_t^r follows from stability of $\mathcal{H}_t^r$, which was proved by Mubayi [138].

11 Algebraic methods

Kalai [98] proposed the following conjecture generalising Turán's tetrahedron problem. Suppose that G is a 3-graph on $[n] = \{1, \cdots, n\}$ such that every 4-set of vertices spans at least one edge (thus G is the complement of a K_4^3-free 3-graph). Fix $s \geq 1$ and consider the following matrix $M_s(G)$. The rows are indexed by edges of G. The columns are divided into s blocks, each of which contains $\binom{n}{2}$ columns indexed by all pairs of vertices. The entry in row e and column uv in block i is $\pm x_{i,w}$ if $e = uvw$ for some w or is 0 otherwise, where $\{x_{i,w} : 1 \leq i \leq s, w \in V(G)\}$ are indeterminate variables, and the sign is positive if w lies between u and v, otherwise negative. Let $r_s(G)$ be the rank of $M_s(G)$. The conjecture is that

$$r_s(G) \geq \sum_{i=1}^{s} \left(\binom{n-2i}{2} - \binom{i}{2} \right) = s\binom{n}{2} - 2\binom{s+1}{2}n + 3\binom{s+2}{3}.$$

Note that the sum in the conjecture is maximised when $s = \lfloor n/3 \rfloor$, and the value obtained is the number of edges in (the complementary form of) Turán's conjecture.

What is the motivation for this conjecture? The definition of $M_s(G)$ is reminiscent of the incidence matrices, which have seen many applications in Combinatorics (see [75]). The (pair) *incidence matrix* for G has rows indexed by edges of G, columns by pairs of vertices, and the entry in row e and column uv is 1 if $e = uvw$ for some w or is 0 otherwise. Thus $M_s(G)$ is obtained by concatenating s copies of the incidence matrix and replacing the 1's by certain weights. In the case $s = 1$, we can set all the variables $x_w := x_{1,w}$ equal to 1 without changing the rank: to see this note that the variables cancel if we multiply each column uv by $x_u x_v$ and divide each row uvw by $x_u x_v x_w$. Thus we obtain the *signed incidence matrix*, which is obtained from the incidence matrix by attaching signs according to the order of u, v, w as above.

To understand signed incidence matrices it is helpful to start with graphs. Suppose G is a graph on $[n]$. Then the signed incidence matrix has rows indexed by edges of G, columns indexed by $[n]$, and a row ij with $i < j$ has -1 in column i, 1 in column j, and 0 in the other columns. Note that the set of rows corresponding to a cycle in G can be signed so that the sum is zero, so is linearly dependent. Conversely, it is not hard to show by induction that a set of rows corresponding to an acyclic subgraph is linearly independent. Another way to say this is that the signed incidence matrix is a linear representation of the cycle matroid of G. (We refer the reader to [149] for an introduction to Matroid Theory.) Thus the maximum rank is $n-1$, with equality if and only if G is connected.

Kalai [97] developed a 'hyperconnectivity' theory for graphs using generalised signed incidence matrices. Similarly to the 3-graph case, when G is a graph, we define the matrix $M_s(G)$, which has rows indexed by $E(G)$, s blocks of columns each indexed by $[n]$, and a row uv with $u<v$ has $x_{i,v}$ in column u of block i, $-x_{i,u}$ in column v of block i, and is 0 otherwise. The resulting matrix is then considered to be a linear representation of the *s-hyperconnectivity matroid*. The maximum possible rank is $sn-\binom{s+1}{2}$. Any G achieving this maximum is called *s-hyperconnected*. One result of [97] is that any s-hyperconnected graph is s-connected, in the usual sense that deleting any $s-1$ vertices leaves a connected graph. Another is that K_{s+2} is a circuit, i.e. a minimally dependent set in the matroid, which leads us to an interesting digression on saturation problems.

The *saturation problem* for H is to determine $s(n,H)$, defined as the minimum number of edges in a maximal H-free graph on n vertices. Thus we want an H-free graph G such that adding any new edge to G creates a copy of H, and G has as few edges as possible. Suppose that G is K_{s+2}-saturated. Then for any pair $uv \notin E(G)$ there is a copy of K_{s+2} in $G \cup uv$, which is a circuit, so uv is in the span of G. It follows that G spans the entire s-hyperconnectivity matroid. In particular, G has at least $sn-\binom{s+1}{2}$ edges. This bound is tight, as may be seen from the example K_s+E_{n-s}, i.e. a clique of size s completely joined to an independent set of size $n-s$. More generally, the same argument applies to any G that has the weaker property that there is a sequence $G=G_0,G_1,\dots,G_t=K_n$, where each G_{i+1} is obtained from G_i by adding an edge that creates a copy of K_{s+2} in G_{i+1} that was not present in G_i. Thus Kalai showed that such G also must have at least $sn-\binom{s+1}{2}$ edges, giving a new proof of a conjecture of Bollobás [19, Exercise 6.17]. See the recent survey [63] for more information on saturation problems.

Now we return to consider the meaning of the signed incidence matrix for 3-graphs. First we give another interpretation for graphs. We can think of the signed incidence matrix for K_n as a linear map from $\mathbb{F}^{\binom{n}{2}}$ to $\mathbb{F}^n$, for some field $\mathbb{F}$, acting on row vectors from the right. Then an edge uv with $u<v$ is mapped to the vector $v-u$, where we are identifying edges and vertices with their corresponding basis vectors. Geometrically, this is a *boundary* operation: we think of the line segment from u to v as having boundary points u and v, with the sign indicating the order. Similarly, we can think of the signed incidence matrix for K_n^3 as a linear map from $\mathbb{F}^{\binom{n}{3}}$ to $\mathbb{F}^{\binom{n}{2}}$, where an edge uvw with $u<v<w$ is mapped to $-vw+uw-uv$. It is convenient to write $vu=-uv$. Then we can think of the boundary operation as taking a 2-dimensional triangle uvw to its bounding cycle, oriented cyclically as wv, vu, uw. This cycle has 'no boundary', in that if we apply the boundary operation to $wv+vu+uw$ we get $v-w+w-u+u-v=0$. In general, an oriented cycle has no boundary, which conforms to the geometric picture of it as a closed loop. The

cycles generate the *cycle space*, which is the subspace of $\mathbb{F}^{\binom{n}{3}}$ of vectors that have no boundary, i.e. are mapped to zero by the signed incidence matrix.

Now consider a 3-graph G on $[n]$. We can create a simplicial complex C which has G as its two-dimensional faces, and the complete graph K_n as its one-dimensional faces; i.e. we take the 1-skeleton of the n-simplex and glue in triangles according to the edges of G. We interpret the rows of the signed incidence matrix of G as the boundary cycles of the triangles. These generate the *boundary space* of G, which is a subspace of the cycle space of K_n. The quotient space is the first *homology* space $H_1(C)$: it is a measure of the number of 1-dimensional 'holes' in the complex C. A lower bound on $r_1(G)$ is equivalent to an upper bound on the first Betti number $\beta_1(C) = \dim H_1(C)$, so the case $s = 1$ of Kalai's conjecture can be rephrased as saying that $\beta_1(C) \le n-2$: this was proved by Kalai (unpublished). He also established the corresponding algebraic generalisation of Turán's theorem on complete graphs.

Kalai also introduced a procedure of *algebraic shifting*, which is an intriguing and potentially powerful tool for a variety of combinatorial problems. In general, 'shifting' or 'compression' refers to a commonly employed technique in extremal set theory, where a problem for general families is reduced to the problem for an initial segment in some order; e.g. most proofs of the Kruskal-Katona theorem have this flavour. We refer the reader to [99] for a survey; here we just give a very brief taste of the operation and its properties. Suppose G is a k-graph on $[n]$ and $X = (x_{ij})_{i,j=1}^n$ is a matrix of indeterminates. Let $X^{\wedge k}$ be the $\binom{n}{k}$ by $\binom{n}{k}$ matrix indexed by k-sets, in which the (S,T)-entry is the determinant of the k by k submatrix of X corresponding to S and T. Let $M(G)$ be the submatrix of $X^{\wedge k}$ formed by the rows corresponding to edges of G. Now construct a basis for the column space of $M(G)$ by the greedy algorithm, at each step choosing the first column not in the span of those chosen previously. The k-sets indexing the chosen columns give the *(exterior) shifted family* $\Delta(G)$. This rather obscure process has some remarkable properties. Björner and Kalai [16] showed that it preserves the face numbers and Betti numbers of any simplicial complex. Even for a graph G, the presence of certain edges in $\Delta(G)$ encodes non-trivial information. For example, 23 appears iff G has a cycle, 45 appears iff G is non-planar, and dn appears iff G is d-hyperconnected. It seems computationally hard to compute $\Delta(G)$, although the randomised algorithm of substituting random constants for the variables and using Gaussian elimination is very likely to give the correct result. Potential applications are discussed in Section 6 of [99], but they still are yet to be realised!

Another application of homological methods was given by Csakany and Kahn [37]. A *d-simplex* is a collection of $d+1$ sets with empty intersection, every d of which have nonempty intersection. A few examples serve to illustrate that many common extremal problems have a forbidden configuration that is a simplex: the Erdős-Ko-Rado theorem [57] forbids 2 disjoint sets, which is a 1-simplex; the Ruzsa-Szemerédi $(6,3)$-theorem [165] forbids the *special triangle* $\{123, 345, 561\}$, which is a 2-simplex; the Turán tetrahedron problem forbids the 3-simplex K_4^3. Chvátal [33] posed the problem of determining the largest r-graph on n vertices with no d-simplex (Erdős [52] had posed the triangle problem earlier). He conjectured that when $r \ge d+1 \ge 2$ and $n > r(d+1)/d$ the maximum number of edges is $\binom{n-1}{r-1}$, with equality only for a star (all sets containing some fixed vertex). The known cases are $r = d+1$ (Chvátal [33]), fixed r, d and large n (Frankl and Füredi [68]), $d = 2$ (Mubayi and Verstraëte

[145]), and $\Omega(n) < r < n/2 - O(1)$ (Keevash and Mubayi [107]).

Csakany and Kahn gave new proofs of Chvátal's result (and also a similar result of Frankl and Füredi on the special triangle). They work with homology over the field $\mathbb{F}_2$, which has the advantage that there is no need to worry about signs ($+1 = -1$), so boundary maps are given by incidence matrices. They note that a star G is acyclic, meaning that the boundary map is injective on the space generated by the edges of G. Furthermore, for any acyclic G the size of G is equal to the dimension of its boundary space, which is at most $\binom{n-1}{r-1}$, as this is the dimension of the boundary space of the complete r-graph K_n^r. Thus it suffices to consider the case when G has a non-trivial cycle space. Next they show that all minimal cycles in G are copies of K_{r+1}^r, and that no edge can overlap a K_{r+1}^r in precisely $r-1$ points. Thus each cycle substantially reduces the dimension of the boundary space for the acyclic part of G, and (omitting some substantial details) the result follows after some rank computations.

We mention one final application of algebraic methods with a different flavour. Suppose G is a graph on $[n]$. Assign variables $x_1, \cdots, x_n$ to the vertices and consider the polynomial $f_G(x) = \prod_{ij \in E(G)} (x_i - x_j)$. Thus f_G vanishes iff $x_i = x_j$ for some edge ij. Note that G has independence number $\alpha(G) < k$ iff f_G belongs to the ideal I of polynomials in $\mathbf{Z}[x]$ that vanish on any assignment x with at least k equal variables. Li and Li [121] showed that I is generated by the polynomials $f_G(x)$ for graphs G that are a disjoint union of $k-1$ cliques, and moreover the sizes of the cliques may be taken as equal as possible. One can also show that the degree of any polynomial in I is at least the degree of the generators. Applying this to f_G for any G with $\alpha(G) < k$, the resulting lower bound on the number of edges gives a proof of Turán's theorem (in complementary form). It would be interesting to obtain similar generalisations for other Turán problems.

12 Probabilistic methods

While probabilistic methods are generally very powerful in Combinatorics, they seem to be less effective for Turán problems, perhaps because the extremal constructions tend to be quite orderly. Some exceptions to this are random constructions for the tetrahedron codegree problem (see Section 13.2) and the bipartite link problem (see Section 13.6). For certain bipartite Turán problems the best known constructions are random, although these are in cases where the upper bound is quite far from the lower bound, so it is by no means an indication that the best construction is random. Consider the Turán problem for the complete bipartite graph $K_{r,r}$. Kövari, Sós and Turán [105] obtained the upper bound $\text{ex}(n, K_{r,r}) = O(n^{2-1/r})$. A simple probabilistic lower bound due to Erdős and Spencer [61] is obtained by taking the random graph $G_{n,p}$ and deleting an edge from each copy of $K_{r,r}$. Then the expected number of edges has order $\Theta(pn^2) - \Theta(p^{r^2} n^{2r})$, so choosing $p = \Theta(n^{-2/(r+1)})$ gives a lower bound of order $\Omega(n^{2-2/(r+1)})$. Recently, Bohman and Keevash [18] obtained a small improvement to $\Omega(n^{2-2/(r+1)} (\log n)^{1/(r^2-1)})$ from the analysis of the H-free process. However, there is still a polynomial gap between the bounds.

Lu and Székely [128] applied the Lovász Local Lemma to Turán problems (among others). The general framework is as follows. Suppose that $A_1, \cdots, A_n$ are 'bad' events. A graph G on $[n]$ is a *negative dependency* graph for the events if $\mathbb{P}(A_i \mid$

$\cap_{j \in S} \overline{A_j}) \leq \mathbb{P}(A_i)$ for any i and S such that there are no edges from i to S and $\mathbb{P}(\cap_{j \in S} \overline{A_j}) > 0$. The general form of the local lemma states that if there are $x_1, \cdots, x_n \in [0,1)$ such that $\mathbb{P}(A_i) \leq x_i \prod_{j: ij \in E(G)} (1 - x_j)$ for all i then $\mathbb{P}(\cap_{i=1}^n \overline{A_i}) \geq \prod_{i=1}^n (1 - x_i) > 0$, i.e. there is a positive probability that none of the bad events occur. This is applied to give the following packing result for hypergraphs. Suppose that H_1 and H_2 are r-graphs such that H_i has m_i edges and every edge of H_i intersects at most d_i other edges of H_i, for $i = 1, 2$. Suppose that $n \geq \max\{v(H_1), v(H_2)\}$ and $(d_1 + 1)m_2 + (d_2 + 1)m_1 \leq \frac{1}{e}\binom{n}{r}$, where e is the base of natural logarithms. Then there are edge-disjoint embeddings of H_1 and H_2 on the same set of n vertices. This result is in turn used to deduce several results, including the following Turán bound. Suppose that F is an r-graph such that every edge intersects at most d other edges. Then $\pi(F) \leq 1 - \frac{1}{e(d+1)}$. This may be compared with the result of Sidorenko mentioned above (Section 6) that bounds $\pi(F)$ in terms of the number of edges in F. In many cases the bound in terms of d is an improvement, but the appearance of e makes it seem very unlikely that it is ever tight!

13 Further topics

This section gives a brief taste of a few areas of research closely related to the Turán problem. It is necessarily incomplete, both in the selection of topics and in the choice of references for each topic. The topics by subsection are 13.1: Jumps, 13.2: Minimum degree problems, 13.3: Different host graphs, 13.4: Coloured Turán problems, 13.5: The speed of properties, 13.6: Local sparsity, 13.7: Counting subgraphs.

13.1 Jumps

Informally speaking, 'jumps' refer to the phenomenon that r-graphs of a certain density are often forced to have large subgraphs with a larger density. For example, the Erdős-Stone theorem implies that a large graph of density bigger than $1 - 1/t$ contains blowups $K_{t+1}(m)$ of K_{t+1}, so has large subgraphs of density more than $1 - 1/(t+1)$. Another example is the result of Erdős mentioned earlier (Section 2) that a large r-graph of positive density contains complete r-partite r-graphs $K_r^r(m)$, so has large subgraphs of density more than $r!/r^r$. Formally, the density d is a *jump* for r-graphs if there is some $c > 0$ such that for any $\epsilon > 0$ and $m \geq r$ there is n_0 sufficiently large such that any r-graph on n vertices with density at least $d + \epsilon$ has a subgraph on m vertices with density at least $d + c$. For example, every $d \in [0,1)$ is a jump for graphs, and every $d \in [0, r!/r^r)$ is a jump for r-graphs. Deciding whether $r!/r^r$ is a jump for r-graphs is a long-standing open problem of Erdős [53]. In fact, Erdős made the stronger conjecture that every $d \in [0,1)$ is a jump for r-graphs, but this was disproved by Frankl and Rödl [72]. The distribution of jumps and non-jumps is not at all understood, and very few specific examples are known. Further examples of non-jumps are given by Frankl, Peng, Rödl and Talbot [71] and Peng, e.g. [150]. On the positive side, Baber and Talbot [9] recently applied flag algebras to show that every $d \in [0.2299, 0.2316)$ is a jump for 3-graphs.

We give a brief sketch of the Frankl-Rödl method, as applied in [71] to prove that $5/9$ is not a jump for 3-graphs. One uses the following reformulation: d is a jump for r-graphs if and only if there is a finite family $\mathcal{F}$ of r-graphs with Turán density $\pi(\mathcal{F}) \le d$ and blowup density $b(F) > d$ for all $F \in \mathcal{F}$. Suppose for a contradiction that $5/9$ is a jump for 3-graphs. Choose $\mathcal{F}$ with $\pi(\mathcal{F}) \le 5/9$ and $b(F) > 5/9$ for all $F \in \mathcal{F}$. Let t be large and G be the Turán construction with parts of size t. Now the idea is to add $O(t^2)$ random edges inside each part, obtaining G^* such that $b(G^*) > 5/9$, but $b(H) \le 5/9$ for any small subgraph H of G^* (here we are omitting a lot of the proof). Since $b(G^*) > 5/9$ and $\pi(\mathcal{F}) \le 5/9$, a sufficiently large blowup $G^*(m)$ must contain some $F \in \mathcal{F}$. We can write $F \subseteq H(m)$ for some small subgraph H of G^*. But then $b(F) \le b(H(m)) = b(H) \le 5/9$ contradicts the choice of $\mathcal{F}$, so $5/9$ is not a jump for 3-graphs.

13.2 Minimum degree problems

Turán problems concern the maximum number of edges in an F-free r-graph, but it is also natural to ask about the maximum possible minimum degree. More precisely, there is a minimum s-degree parameter $\delta_s(G)$ for each $0 \le s \le r-1$, defined as the minimum over all sets S of s vertices of the number of edges containing S. Then we can define a generalised Turán number $\mathrm{ex}_s(n, F)$ as the largest value of $\delta_s(G)$ attained by an F-free r-graph G on n vertices. Note that $\delta_0(G) = e(G)$, so $\mathrm{ex}_0(n, F) = \mathrm{ex}(n, F)$ is the usual Turán number. We can also define generalised Turán densities $\pi_s(F) = \lim_{s\to\infty} \mathrm{ex}_s(n, F)\binom{n-s}{r-s}^{-1}$ (it is non-trivial to show that the limit exists). A simple averaging argument shows that $\pi_i(F) \ge \pi_j(F)$ when $i \le j$. The vertex deletion method in Proposition 4.2 shows that $\pi_1(F) = \pi_0(F) = \pi(F)$, so minimum 1-degree problems are not essentially different to Turán problems. However, in general we obtain a rich source of new problems, and it is not apparent how they relate to each other. The case $s = r-1$ was introduced by Mubayi and Zhao [146] under the name of *codegree density*. Their main result is that for $r \ge 3$ there are no jumps for codegree problems. In particular, the set of codegree densities is dense in $[0, 1)$. Moreover, they conjecture that any $d \in [0, 1)$ is the codegree density of some family.

As for Turán problems, there are few known results for codegree problems, even asymptotically. The tetrahedron K_4^3 is again one of the first interesting examples. Here the asymptotically best known construction is to take a random tournament on $[n]$ and say that a triple ijk with $i < j < k$ is an edge if i has one edge coming in and one edge coming out. This shows that the codegree density of the tetrahedron is at least $1/2$. For an upper bound, nothing better is known than the bounds for the usual Turán density, which are also upper bounds on the codegree density by averaging. One known result is for the Fano plane, where Mubayi [137] showed that the codegree density is $1/2$. Turán and codegree problems for other projective geometries were considered in [103, 113, 104]. An exact codegree result for the Fano plane was obtained by Keevash [104]: if G is a Fano-free 3-graph on n vertices, where n is large, and $\delta_2(G) \ge n/2$, then n is even and G is a balanced complete bipartite 3-graph. The argument used a new 'quasirandom counting lemma' for regularity theory, which extends the usual counting lemma by not only counting copies of a particular subgraph, but also showing that these copies are evenly distributed. Even for graphs, this quasirandom counting lemma has consequences that are not immediately obvious; for example, given a tripartite graph G in which each bipartite graph is dense and ϵ-regular (for some small ϵ), for any choice of dense graphs H_1, H_2, H_3 inside the parts V_1, V_2, V_3 of G, there are many copies of $K_3(2)$ in G in which the pairs inside each part are edges of the graphs H_1, H_2, H_3. Results of this type are potentially very powerful in assembling hypergraphs from smaller pieces.

Minimum degree conditions also lead to the study of spanning configurations. Here we look for conditions on a hypergraph G on n vertices that guarantee a particular subgraph F that also has n vertices. The prototype is Dirac's theorem [44] that every graph on $n \geq 3$ vertices with minimum degree at least $n/2$ contains a Hamilton cycle. Other classical minimum degree result for graphs is the Hajnal-Szemerédi theorem [91] that minimum degree $(1 - 1/t)n$ gives a perfect packing by copies of K_t (when t divides n). A generalisation by Kómlos, Sarközy and Szemerédi [115] states that the same minimum degree even gives the $(t-1)$th power of a Hamilton cycle (when n is large). Another generalisation by Kühn and Osthus [120] determines the threshold for packing an arbitrary graph H up to an additive constant (the precise statement is technical). An example result for hypergraphs is a theorem of Rödl, Ruciński and Szemerédi [160] that any r-graph on n vertices with minimum codegree $(1 + o(1))n/2$ has a 'tight' Hamilton cycle, i.e. a cyclic ordering of the vertices such that every consecutive r-set is an edge. We refer the reader to the surveys [119] for graphs and [161] for hypergraphs.

13.3 Different host graphs

A range of new problems open up when we consider additional properties for Turán problems, besides that of not containing some forbidden r-graph. For any host r-graph H and fixed r-graph F, we may define $\mathrm{ex}(H, F)$ as the maximum number of edges in an F-free subgraph of H. Thus the usual Turán number $\mathrm{ex}(n, F) = \mathrm{ex}(K_n^r, F)$ is the case when H is complete. We stick to graphs ($r = 2$) for simplicity.

In principle one can consider any graph H, but some host graphs seem particular natural. An important case is when H is given by a random model, e.g. the Erdős-Rényi random graph $G_{n,p}$. This is motivated by considerations of *resilience* of properties. Here we consider some property of $G_{n,p}$ (i.e. a property that holds with high probability) and ask how resilient it is when some edges are deleted. For example, if p is not 'too small', $G_{n,p}$ not only has a triangle, but one even needs to delete asymptotically half of the edges to destroy all triangles. Equivalently, any triangle-free subgraph of $G_{n,p}$ has asymptotically at most half of its edges. This is tight, as any graph has a bipartite subgraph that contains at least half of its edges. To clarify what 'too small' means, note that if the number of triangles is much smaller than the number of edges then such a result will not hold, as one can delete one edge from each triangle with negligible effect. This suggests $p = n^{-1/2}$ as a threshold for the problem, which is indeed the case (this follows from a result of Frankl and Rödl [74]). There is a large literature on generalising this result, which we do not have space to go into here. A comprehensive generalisation to many extremal problems was recently given independently by Schacht [169] and Conlon and Gowers [36]. Among these results is Turán's theorem for random graphs, that when p is not too small the largest K_{t+1}-free subgraph of $G_{n,p}$ has asymptotically $1 - 1/t$ of its edges; again the threshold for p is the value for which the number of K_t's is comparable with the number of edges. Similar results apply for hypergraph

Turán problems, and to certain extremal problems from number theory, such as Szemerédi's theorem on arithmetic progressions. Another direction of research is that started by Sudakov and Vu [185] on *local resilience*. Here the question is how many edges one needs to delete from each vertex to destroy a certain property of $G_{n,p}$. This is a better question for global properties such as Hamiltonicity, which one can destroy by deleting all edges at one vertex: this is not a significant global change, but a huge local change.

The above only concerns the case when the host graph H is random. Mubayi and Talbot [144] consider Turán problems with colouring conditions, which could also be viewed from the perspective of a constrained host graph. Say that an r-graph G is *strongly t-colourable* if there is a t-colouring of its vertices such that no edge has more than one vertex of the same colour. (They call this 't-partite', but we use this term differently in this paper; our use of 't-partite' is equivalent to their use of 't-colourable'.) Their main result (in our language) is that the asymptotic maximum density of an F-free r-graph on n vertices that is strongly t-colourable is equal to the maximum blowup density $b(G)$ over all hom-F-free r-graphs G on t vertices. For example, the maximum density in a strongly 4-colourable K_4^3-free 3-graph is $8/27$; this is achieved by a construction with 4 parts of sizes $n/3, 2n/9, 2n/9, 2n/9$, with edges equal to all triples with one vertex in the large part and the other two vertices in two different smaller parts. Chromatic Turán problems were considered earlier by Talbot [186] as a tool for obtaining bounds on Turán density of the 3-graph on 4 vertices with 3 edges. (These chromatic bounds were subsequently improved by Markström and Talbot [132].) Here the problem is to estimate the maximum density of an F-free r-graph on n vertices that is t-partite. Mubayi and Talbot solved this problem for the extended complete graph, in the sense that they have a procedure for computing the maximum density, which is in principle finite, although not practical except in small cases. They conjecture that the natural example is optimal, but can only prove this for $r = 2$ and $r = 3$. One example of their result shows that chromatic Turán densities can be irrational: the maximum density of a bipartite K_6^3-free 3-graph is $(13\sqrt{13} - 35)/27 \approx 0.4397$, achieved by blowing up $K_5^3 - e$.

Another case which has received a lot of attention is when $H = Q_n$ is the graph of the n-cube, i.e. $V(H)$ consists of all subsets of $[n]$ and edges join sets that differ in precisely one element. Erdős [55] posed the problem of determining the maximum proportion of edges in a C_4-free subgraph of Q_n. Noting that any consecutive levels of the cube span a C_4-free subgraphs, a lower bound of $1/2$ is obtained by taking the union of the subgraphs spanned by levels $2i$ and $2i+1$ for $0 \le i < n/2$. The best known upper bound is approximately 0.6226, due to Thomason and Wagner [189]. We will not attempt to survey the literature on these problems, but refer the reader to Conlon [34] for a simpler proof of many of the known results and several references. We draw the reader's attention to the problem of deciding whether $\text{ex}(Q_n, C_{10})$ has the same order of magnitude as $e(Q_n)$; this is the only unsolved instance of this problem for cycles (the answer is 'yes' for C_4 and C_6; 'no' for C_8 and longer cycles).

We also remark that even 'vertex Turán problems' in the cube seem to be hard. For example, what is the smallest constant a_d such that there is a set of $\sim a_d 2^n$ vertices in the n-cube that hits every subcube of dimension d? This problem was introduced by Alon, Krech and Szabó [5], who showed $\frac{\log d}{2^{d+2}} \le a_d \le \frac{1}{d+1}$; there is a surprisingly large gap between the upper and lower bounds! A variant on this problem introduced by Johnson and Talbot [96] is to find a particular subset $F \subseteq V(Q_d)$: what is the large constant λ_F such that there exists $S \subseteq V(Q_n)$ of size $|S| \sim \lambda_F 2^n$ such that there is no subgraph embedding $i : Q_d \to Q_n$ with $i(F) \subseteq S$? In particular, they conjecture that $\lambda_F = 0$ for $|F| \le \binom{d}{d/2}$ (it is not hard to see that this can be false for larger F). Bukh (personal communication) observed that this conjecture is equivalent to the following hypergraph Turán problem. For $r \ge s > t$ we define the following r-graph $S^r K_s^t$, which may be regarded as a 'suspension' of K_s^t. The vertex set of $S^r K_s^t$ is the union of disjoint sets S of size s and A of size $r - t$. The edges consist of all r-tuples containing A. The conjecture is that $\lim_{r\to\infty} \pi(S^r K_s^t) = 0$ for any $s > t \ge 2$. Even the case $s = 4$ and $t = 2$ is currently open! The case $s = 3$ and $t = 2$ is straightforward: $S^r K_3^2$ just consists of (any) 3 edges on a set of $r + 1$ vertices, so $\pi(S^r K_3^2) \le 2/(r + 1)$. However, it is an interesting problem to determine the order of magnitude of $\pi(S^r K_3^2)$ for large r: Alon (communication via Bukh) gave a lower bound of order $(\log r)/r^2$. In general, given the apparent difficulty of determining Turán densities exactly, it seems that such problems involving additional limits may be a fruitful avenue for developing the theory.

13.4 Coloured Turán problems

There are a variety of generalisations of the Turán problem that allow additional structures, such as directed edges, multiple edges, or coloured edges. Even for graphs this leads to rich theories and several unsolved problems. Brown, Erdős and Simonovits initiated this field with a series of papers on problems for digraphs and multigraphs. For multigraph problems, we fix some positive integer q and consider multigraphs with no loops and edge multiplicity at most q. Then given a family $\mathcal{F}$ of multigraphs, we want to determine $\mathrm{ex}(n, \mathcal{F})$, defined as the maximum number of edges in a multigraph not containing any F in $\mathcal{F}$. A further generalisation allows directions on the edges; for simplicity we ignore this here. In the case $q = 1$ this is the usual Turán problem. For $q = 2$ (or digraphs), it is shown in [27] that any extremal problem has a blowup construction that is asymptotically optimal. Here a *blowup* is defined by taking some symmetric $t \times t$ matrix A whose entries are integers between 0 and q, dividing a vertex set into t parts, and putting a_{ij} edges between any pair of vertices u, v with u in part i and v in part j (we may have $i = j$).

Similarly to the usual Turán problem, one can define the *blowup density* $b(A)$ which is the density achieved by this construction: formally $b(A)$ is the maximum value of $x^t Ax$ over the standard simplex S of all $x = (x_1, \cdots, x_n)$ with $x_i \ge 0$ for $1 \le i \le n$ and $\sum x_i = 1$. Say that such a matrix is *dense* if any proper principal submatrix has lower blowup density. It is shown in [28] that for any dense matrix A there is a finite family $\mathcal{F}$ such that A is the unique matrix whose blowup gives asymptotically optimal constructions of $\mathcal{F}$-free multigraphs. Furthermore, for $q = 2$ (or digraphs), in [29] they describe an algorithm that determines all optimal matrices for a given family (the algorithm is finite, but not practical). Simpler proofs of these results were given by Sidorenko [178], who also showed that analogous statements do not hold for $q > 2$, thus disproving a conjecture of Brown, Erdős and Simonovits. Brown, Erdős and Simonovits also conjectured that all densities are jumps (as for graphs), but this was disproved by Rödl and Sidorenko [164] for $q \ge 4$.

The conjecture is true for $q = 2$, but is open for $q = 3$.

Another coloured variant on many problems of extremal set theory, including Turán problems, was introduced by Hilton [93] and later by Keevash, Saks, Sudakov and Verstraëte [109]. Given a list of set systems, which we think of as colours, we call another set system *multicoloured* if for each of its sets we can choose one of the colours it belongs to in such a way that each set gets a different colour. Given an integer k and some forbidden configurations, the multicoloured extremal problem is to choose k colours with total size as large as possible subject to containing no multicoloured forbidden configuration. Let f be the number of sets in the forbidden configuration. One possible extremal construction for this problem is to take $f - 1$ colours to consist of all possible sets, and the other colours to be empty. Another construction is to take all k colours to be equal to some fixed family that is of maximum size subject to not containing a forbidden configuration. In [109] we solved the multicolour version of Turán's theorem, by showing that one of these two constructions is always optimal. In other words, if $G_1, \cdots, G_k$ are graphs on the same set of n vertices for which there is no multicoloured K_t, then $\sum_{i=1}^{k} e(G_i)$ is maximised either by taking $\binom{t}{2} - 1$ complete graphs and the rest empty graphs, or by taking all k graphs equal to some fixed Turán graph $T_{t-1}(n)$. Simple calculations show that there is a threshold value k_c so that the first option holds for $k < k_c$ and the second option holds for $k \geq k_c$. This proved a conjecture of Hilton [93] (although we were not aware of this paper at the time of writing). It would be interesting to understand which other extremal problems exhibit this phenomenon of having only two extremal constructions. It is not universal, as shown by an example in [109], but it does hold for several other classical problems of extremal set theory, as shown in [21] and [112].

A related problem posed by Diwan and Mubayi (unpublished) concerns the minimum size of a colour, rather than the total size of colour. Specifically, for any n and a fixed graph F with edges coloured red or blue, they ask for the threshold m such that, given any red graph and blue graph on the same set of n vertices each with more than m edges, one can find a copy of F with the specified colouring. They pose a conjecture when F is a coloured clique, and prove certain cases of their conjecture. Their proof uses a stronger result which replaces the minimum size of a colour by a weighted linear combination of the colours. Such problems have been recently studied in a much more general context by Marchant and Thomason [130], who gave applications to the probability of hereditary graph properties (see Section 13.5).

We conclude this subsection with another coloured generalisation studied by Keevash, Mubayi, Sudakov and Verstraëte [108]. For a fixed graph H, we ask for the maximum number of edges in a properly edge-coloured graph on n vertices which does not contain a rainbow H, i.e. a copy of H all of whose edges have different colours. This maximum is denoted $\mathrm{ex}^*(n, H)$, and we refer to it as the *rainbow Turán number* of H. For any non-bipartite graph H we showed that $\mathrm{ex}^*(n, H) \sim \mathrm{ex}(n, H)$, and for large n we have $\mathrm{ex}^*(n, H) = \mathrm{ex}(n, H)$ when H is critical (e.g. a clique or an odd cycle). Bipartite graphs H are a source of many open problems. The case when $H = C_{2k}$ is an even cycle is particularly interesting because of its connection to additive number theory. We conjecture that $\mathrm{ex}^*(n, C_{2k}) = O(n^{1+1/k})$, which would generalise a result of Ruzsa on B_k^*-sets in abelian groups. (We proved it for

$k = 2$ and $k = 3$.) More generally, there is considerable scope to investigate number theoretic consequences of extremal results on coloured graphs, as applied to Cayley graphs.

13.5 The speed of properties

Suppose $\mathcal{P}$ is a graph property, i.e. a set of graphs that is closed under isomorphism. We consider properties $\mathcal{P}$ that are *hereditary*, meaning that they are closed under taking induced subgraphs, or even *monotone*, meaning that they are closed under taking arbitrary subgraphs. A monotone property can be characterised as the set of $\mathcal{F}$-free graphs, for some (possibly infinite) family $\mathcal{F}$. Similarly, a hereditary property can be characterised as the set of *induced-$\mathcal{F}$-free* graphs, for some $\mathcal{F}$, i.e. graphs with no copy of any F in $\mathcal{F}$ as an induced subgraph. The *speed* $s(n)$ of $\mathcal{P}$ is the number of labelled graphs in $\mathcal{P}$ on $[n]$. There is a large literature on the speed of properties, too large to adequately cite here, so we refer the reader to [3] as a recent paper with many references.

Consider the problem of counting F-free graphs on $[n]$, for some fixed graph F. By taking all subgraphs of any fixed F-free graph of maximum size $\mathrm{ex}(n, F)$ we obtain at least $2^{ex(n,F)}$ distinct F-free graphs. In fact, this is essentially tight for non-bipartite graphs F, as Erdős, Frankl and Rödl [56] showed an upper bound of $2^{(1+o(1))ex(n,F)}$. (The case when F is bipartite is another story, see [15] for some recent results.) The corresponding generalisation to hereditary properties was proved by Alekseev [1] and by Bollobás and Thomason [22]. They showed that the speed of $\mathcal{P}$ is $2^{(1-1/r+o(1)n^2/2}$, where r is a certain parameter of $\mathcal{P}$ known as the 'colouring number' (informally, it is the maximum number of parts in a partite construction for graphs in $\mathcal{P}$, where each part is complete or empty, and the graph is otherwise arbitrary). These results have been refined to give more precise error terms and even a description of the structure of almost all graphs in a hereditary property. For monotone properties the results are due to Balogh, Bollobás and Simonovits [11, 12], and for hereditary properties to Alon, Balogh, Bollobás and Morris [3]. Bollobás and Thomason [23] studied a generalisation in which a property is measured by its probability of occurring in the random graph $G(n, p)$ (thus the speed corresponds to $p = 1/2$). This generalised problem exhibits extra complexities, analysed by Marchant and Thomason [130] (see Section 13.4).

It is natural to pose the same questions for hypergraph properties. Dotson and Nagle [45] and Ishigami [95] showed that the speed of a hereditary r-graph property $\mathcal{P}$ is $2^{ex(n,\mathcal{P})+o(n^r)}$. Here $\mathrm{ex}(n, \mathcal{P})$ is the maximum size of an r-graph G on $[n]$ on n vertices such that there exists an r-graph H on $[n]$ that is edge-disjoint from G such that $H \cup G' \in \mathcal{P}$ for every subgraph G' of G. (In the case when $\mathcal{P}$ is monotone this is the usual Turán number, i.e. the maximum size of an r-graph in $\mathcal{P}$.) In principle this is analogous to the Alexeev-Bollobás-Thomason result, but we lack a concrete description of $\mathrm{ex}(n, \mathcal{P})$ analogous to the colouring number (even in the monotone case, which is the point of this survey!) In the case of the Fano plane a refined result was obtained by Person and Schacht [151], who showed that almost every Fano-free 3-graph on n vertices is bipartite. One might expect similar results to hold for other Turán problems where we know uniqueness and stability of the extremal construction. This has been established by Balogh and Mubayi [13, 14] for cancellative 3-graphs and 3-graphs with independent neighbourhoods.

13.6 Local sparsity

Brown, Erdős and Sós [30] generalised the hypergraph Turán problem by asking for the maximum number of edges in an r-graph satisfying a 'local sparsity' condition that bounds the number of edges in any set of a given size. Write $\mathrm{ex}^r(n,v,e)$ for the maximum number of edges in an r-graph on n vertices such that no set of v vertices spans at least e edges. For example $\mathrm{ex}^3(n,4,4) = \mathrm{ex}(n,K_4^3)$. A result of [30] is $\mathrm{ex}^r(n,e(r-k)+k,e) = \Theta(n^k)$ for any $1 \le k \le r$; the upper bound follows by noting that any k-set belongs to at most $e-1$ edges, and the lower bound by taking a random r-graph of small constant density and deleting all edges in $e(r-k)+k$-sets with at least e edges. They described the case $r=3$, $v=6$, $e=3$ as 'the most interesting question we were unable to answer'. This was addressed by the celebrated '(6,3)-theorem' of Ruzsa and Szémeredi [165] that $n^{2-o(1)} < \mathrm{ex}^3(n,6,3) < o(n^2)$. It would be very interesting to tighten these bounds: this is connected with regularity theory (see Section 10) and bounds for Roth's theorem (see [90, 166].) Further results on the general problem are $n^{2-o(1)} < \mathrm{ex}^r(n,3(r-1),3) < o(n^2)$ in [56], $\mathrm{ex}^r(n,e(r-k)+k+\lfloor\log_2(e)\rfloor,e) < o(n^k)$ in [167], $\mathrm{ex}^r(n,4(r-k)+k+1,4) < o(n^k)$ for $k \ge 3$ in [168], and $n^{k-o(1)} < \mathrm{ex}^r(n,3(r-k)+k+1,3) < o(n^k)$ in [6]. An interesting open problem is to determine whether $\mathrm{ex}^3(n,7,4)$ is $o(n^2)$.

A weighted generalisation of this problem is to determine the largest total weight $\mathrm{ex}_{\mathbf{Z}}(n,k,r)$ that can be obtained by assigning integer weights to the edges of a graph on n vertices such that any set of k vertices spans a subgraph of weight at most r. (We stick to graphs for simplicity.) Note that negative weights are allowed, but for comparison with multigraph problems one can also consider the analogous quantity $\mathrm{ex}_{\mathbf{N}}(n,k,r)$ in which weights have to be non-negative. We remarked earlier that the example $\mathrm{ex}_{\mathbf{N}}(n,4,20) \sim 3\binom{n}{2}$ was crucial in determining the Turán density of the Fano plane. In general, Füredi and Kündgen [80] have determined $\mathrm{ex}_{\mathbf{Z}}(n,k,r)$ asymptotically for all k and r, but there remain several interesting open problems, such as determining exact values and extremal constructions, and obtaining similar results for $\mathrm{ex}_{\mathbf{N}}(n,k,r)$.

Another generalisation is to specify exactly what numbers of edges are allowed in any set of a given size. In Section 7 we discussed the problem for 3-graphs in which every 4-set spans 0, 2 or 3 edges. In Section 6 we mentioned the lower bound $\pi(F) \ge 2/7$ given by Frankl and Füredi [67] when F is the 3-graph with 4 vertices and 3 edges. The main result of [67] is an exact result for 3-graphs in any 4-set spans 0 or 2 edges. In fact they classify all such 3-graphs: they are either obtained by (i) blowing up the 3-graph on 6 vertices described in Section 6, or (ii) by placing n points on a circle and taking the edges as all triples that form a triangle containing the centre (assume that the centre is not on the line joining any pair). It is easy to check that the blowup construction (i) is larger for $n \ge 6$. A related problem is a conjecture of Erdős and Sós [60] that any 3-graph with bipartite links has density at most 1/4. Construction (ii) is an example that would be tight for this conjecture. Another example is to take a random tournament and take the edges to be all triples that induce cyclic triangles. In Section 9 we mentioned the improvements on $\pi(K_{r+1}^r)$ given by Lu and Zhao [129]. These were based on a structural result for r-graphs in which every $(r+1)$-set contains 0 or r edges, answering a question of de Caen [41]: if $r = 2$ then G is a complete bipartite graph, and if $r \geq 3$ and $n > r(p-1)$, where p is the smallest prime factor of $r-1$, then G is either the empty graph or a star (all r-sets containing some fixed vertex).

13.7 Counting subgraphs

A further generalisation of the Turán problem is to look not only for the threshold at which a particular r-graph F appears, but how many copies of F are guaranteed by a given number of edges. Even the most basic case counting triangles in graphs is a difficult problem that was open for many years. The following asymptotic solution was recently given by Razborov [157] using flag algebras. Among graphs on n vertices with edge density between $1-1/t$ and $1-1/(t+1)$, the asymptotic minimum number of triangles is achieved by a complete $(t+1)$-partite graph in which t parts are of equal size and larger than the remaining part. (One can give an explicit formula in terms of the edge density, but the resulting expression is rather unwieldy.) It is conjectured that the same example minimises the number of copies of K_s for any s. Nikiforov [147] established this for $s = 4$, and also re-proved Razborov's result for $s = 3$ by different means. In general, Bollobás (see [19, Chapter 6]) showed a lower bound that is equal to the conjecture for densities of the form $1 - 1/t$, and a linear function on each interval $[1 - 1/t, 1 - 1/(t + 1)]$. Very recently, Reiher announced an asymptotic solution to the full conjecture.

The problem takes on a different flavour when one considers graphs where the number of edges exceeds the Turán number, but is asymptotically the same. For example, Rademacher (unpublished) extended Mantel's result by showing that a graph on n vertices with $n^2/4 + 1$ edges contains at least $\lfloor n/2 \rfloor$ triangles (which is tight). This was extended by Erdős [49] and then by Lovász and Simonovits [125], who showed that if $q < n/2$ then $n^2/4 + q$ edges guarantee at least $q\lfloor n/2 \rfloor$ triangles. Mubayi [139, 140] has extended these results in several directions. For a critical graph F he showed that there is $\delta > 0$ such that if n is large and $1 \leq q < \delta n$ then any graph on n vertices with $\text{ex}(n, F) + q$ edges contains at least $qc(n, F)$ copies of F. Here $c(n, F)$ is the minimum number of copies of F created by adding a single edge to the Turán graph, which is easy to compute for any particular example, although a general formula is complicated. The bound is sharp up to an error of $O(qc(n, F)/n)$. For hypergraphs he obtains similar results in many cases where uniqueness and stability of the extremal example is known.

For bipartite graphs F there is an old conjecture of Sidorenko [177] that random graphs achieve the minimum number of copies of F. A precise formulation may be given in terms of homomorphisms. Recall that the homomorphism density $t_F(G)$ is the probability that a random map from $V(F)$ to $V(G)$ is a homomorphism. Then the conjecture is that $t_F(G) \geq d(G)^{e(F)}$, where $d(G) = t_e(G)$ is the edge density of G. This may be viewed as a correlation inequality for the events that edges of F are embedded as edges of G. It also has an equivalent analytic formulation as $t_F(W) \geq t_e(W)^{e(F)}$ for any graphon W, which is worth noting as integrals similar to $t_F(W)$ appear in other contexts (e.g. Feynmann integrals in quantum field theory). Sidorenko was a pioneer of the analytic approach, and surveyed many of his results in [179]. Recent partial results on the Sidorenko conjecture include a local form by

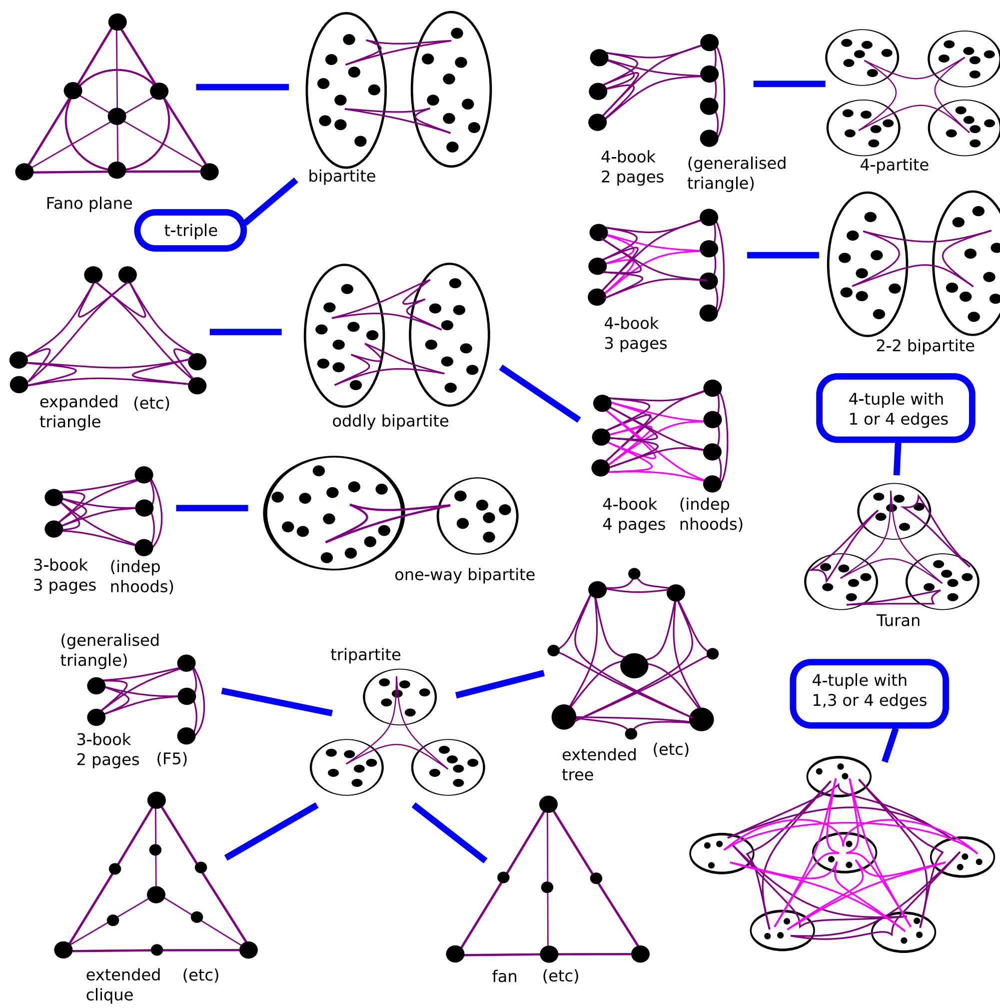


Figure 1: The exact results

Lovász [124] and an approximate form by Conlon, Fox and Sudakov [35]. Note that examples in [177] show that the natural hypergraph generalisation of the conjecture is false.

In the other direction, one may ask to maximise the number of copies of a fixed r-graph F in an r-graph G, given the number of edges and vertices in G. We start with the case when $F = K_t^r$ is a clique. This turns out not to depend on the number of vertices in G. For example, when $e(G) = \binom{m}{r}$ the extremal example is K_m^r, which has $\binom{m}{t}$ copies of K_t^r. In general the extremal example is determined by the Kruskal-Katona theorem [118, 100]. Results for general graphs were obtained by Alon [2] and for hypergraphs by Friedgut and Kahn [76]. Here we do not expect precise answers, but just seek the order of magnitude. The result of [76] is that the maximum number of copies of an r-graph F in an r-graph G with e edges has order $e^{\alpha^*(F)}$, where $\alpha^*(F)$ is the fractional independence number of F.

Going back to cliques in graphs, Sós and Straus [184] proved the following (generalisation of a) conjecture of Erdős [48]. Suppose G is a graph and let N_t denote the number of K_t's in G. If $N_{k+1} = 0$ (i.e. G is K_{k+1}-free) and $t \geq 0$ then

$N_{t+1} \le \binom{k}{t+1}\binom{k}{t}^{-(t+1)/t} N_t^{(t+1)/t}$. Note that equality holds if G is a blowup of K_k. Repeated application gives a bound for the number of K_t's in a K_{k+1}-free graph in terms of the number of edges N_2: we have $N_t \le \binom{k}{t}\binom{k}{2}^{-t/2} N_2^{t/2}$. The proof uses a far-reaching generalisation of the Lagrangian method considered in Section 3. In fact, it is hard to appreciate the scope of the method in the generality presented in [184], and it may well have applications in other contexts yet to be discovered. The idea is to assign a variable x_T to each K_t in G and consider the polynomial $f_G(x) = \sum_S \prod_{T \subseteq S} x_T$ in the variables $x = (x_T)$, where the sum is over all K_{t+1}'s S in G. Let λ be the maximum value of $f_G(x)$ over all x with every $x_T \ge 0$ and $\sum_T x_T^t = 1$ (note the power). A general transfer lemma in [184] implies that a maximising x can be chosen with the property that the vertices incident to variables of positive weight induce a complete subgraph. This implies that x is supported on the K_t's contained in some clique, which has size at most k, since $N_{k+1} = 0$. The maximum is achieved with equal weights $\binom{k}{t}^{-1/t}$, which gives $\lambda = \binom{k}{t+1}\binom{k}{t}^{-(t+1)/t}$. On the other hand, setting every x_T equal to $N_t^{-1/t}$ is a valid assignment, and gives a lower bound $\lambda \ge N_{t+1} N_t^{-(t+1)/t}$, so the result follows.

14 Summary of results

This survey has been organised by methods, so for easy reference we summarise the results here. The exact results are illustrated in Figure 1 (some infinite families are indicated by a representative example). A list follows: F_5 [66] (generalising cancellative 3-graphs [20]), Fano plane [84, 110], expanded triangle [111], generalised 4-graph triangle = 4-book with 2 pages [152] (generalising cancellative 4-graphs [173]), 4-book with 3 pages [83], 3-graphs with independent neighbourhoods [82], 4-graphs with independent neighbourhoods = 4-book with 4 pages [81], extended complete graphs [155] (refining [138]), generalised fans [141], extended trees [174], 3-graph 4-sets with 1, 3 or 4 edges [67] 3-graph 4-sets with 1 or 4 edges [154] (refining [158]), 3-graph t-triples [192]. Besides these, there is an 'almost exact' result for generalised 5-graph and 6-graph triangles [70], and asymptotic results (i.e. exact Turán densities) for expanded cliques [176] and 5 3-graphs related to $F(3,3)$ [143]. Many further asymptotic results follow from Theorem 3.1.

References

[1] V.E. Alekseev, On the entropy values of hereditary classes of graphs, *Discrete Math. Appl.* **3** (1993), 191–199.

[2] N. Alon, On the number of subgraphs of prescribed type of graphs with a given number of edges, *Israel J. Math.* **38** (1981), 116–130.

[3] N. Alon, J. Balogh, B. Bollobás and R. Morris, The structure of almost all graphs in a hereditary property, arXiv:0905.1942, 2009.

[4] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy, Efficient testing of large graphs, *Combinatorica* **20** (2000), 451–476.

[5] N. Alon, A. Krech and T. Szabó, Turán's theorem in the hypercube, *SIAM J. Disc. Math.* **21** (2007), 66-72.

[6] N. Alon and A. Shapira, On an Extremal Hypergraph Problem of Brown, Erdős and Sós, *Combinatorica* **26** (2006), 627–645.

[7] B. Andrásfai, P. Erdős and V. T. Sós, On the connection between chromatic number, maximal clique and minimal degree of a graph. *Disc. Math.* **8** (1974), 205–218.

[8] T. Austin, On exchangeable random variables and the statistics of large graphs and hypergraphs, *Probability Surveys* **5** (2008), 80–145.

[9] R. Baber and J. Talbot, Hypergraphs do jump, to appear in *Combin. Probab. Comput.*

[10] J. Balogh, The Turán density of triple systems is not principal, *J. Combin. Theory Ser. A* **100** (2002), 176–180.

[11] J. Balogh, B. Bollobás and M. Simonovits, On the number of graphs without forbidden subgraph, *J. Combin. Theory Ser. B* **91** (2004), 1–24.

[12] J. Balogh, B. Bollobás and M. Simonovits, The typical structure of graphs without given excluded subgraphs, to appear in *Random Structures Algorithms.*

[13] J. Balogh and D. Mubayi, Almost all triple systems with independent neighborhoods are semi-partite, arXiv:1002.1925, 2010.

[14] J. Balogh and D. Mubayi, Almost all cancellative triple systems are tripartite, arXiv:0910.2941, 2009.

[15] J. Balogh and W. Samotij, The number of K_{st}-free graphs, to appear in *J. London Math. Soc.*

[16] A. Björner and G. Kalai, An extended Euler-Poincaré theorem, *Acta Math.* **161** (1988), 279–303.

[17] T. Bohman, A. Frieze, D. Mubayi and O. Pikhurko, Hypergraphs with independent neighborhoods, to appear in *Combinatorica.*

[18] T. Bohman and P. Keevash, The early evolution of the H-free process, *Invent. Math.* **181** (2010), 291–336.

[19] B. Bollobás, *Extremal graph theory*, Reprint of the 1978 original, Dover Publications, Inc., Mineola, NY, 2004.

[20] B. Bollobás, Three-graphs without two triples whose symmetric difference is contained in a third, *Disc. Math.* **8** (1974), 21–24.

[21] B. Bollobás, P. Keevash and B. Sudakov, Multicoloured extremal problems, *J. Combin. Theory Ser. A* **107** (2004), 295–312.

[22] B. Bollobás and A. Thomason, Projections of bodies and hereditary properties of hypergraphs, *Bull. London Math. Soc.* **27** (1995) 417–424.

[23] B. Bollobás and A. Thomason, The structure of hereditary properties and colourings of random graphs *Combinatorica* **20** (2000), 173–202.

[24] C. Borgs, J. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi, Convergent sequences of dense graphs I: Subgraph frequences, metric properties and testing, *Adv. Math.* **219** (2008), 1801—1851.

[25] D. E. Boyer, D. L. Kreher, S. Radziszowski and A. Sidorenko, On $(n, 5, 3)$-Turán systems, *Ars Combin.* **37** (1994), 13–31.

[26] W. G. Brown, On an open problem of Paul Turán concerning 3-graphs, *Studies in Pure Mathematics*, 91–93, Basel-Boston: Birkhauser, 1983.

[27] W. G. Brown, P. Erdős and M. Simonovits, Extremal problems for directed graphs, *J. Combin. Theory Ser. B* **15** (1973), 77–93.

[28] W. G. Brown, P. Erdős and M. Simonovits, Inverse extremal digraph problems, Finite and infinite sets, Vol. I, II (Eger, 1981), 119–156, *Colloq. Math. Soc. János Bolyai* **37**, North-Holland, Amsterdam, 1984.

[29] W. G. Brown, P. Erdős and M. Simonovits, Algorithmic solution of extremal digraph problems, *Trans. Amer. Math. Soc.* **292** (1985), 421–449.

[30] W. G. Brown, P. Erdős and V. T. Sós, Some extremal problems on r-graphs, *New directions in the theory of graphs* (Proc. Third Ann Arbor Conf., Univ. Michigan, Ann Arbor, Mich., 1971), 53–63, Academic Press, New York, 1973.

[31] S. R. Buló and M. Pelillo, A generalization of the Motzkin-Straus theorem to hypergraphs, *Optim Lett* **3** (2009), 287–295.

[32] 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.

[33] V. Chvátal, An extremal set-intersection theorem, *J. London Math. Soc.* **9** (1974/75), 355–359.

[34] D. Conlon, An extremal theorem in the hypercube, arXiv:1005.0582, 2010.

[35] D. Conlon, J. Fox and B. Sudakov, An approximate version of Sidorenko's conjecture, to appear in *Geom. Funct. Anal.*

[36] D. Conlon and W. T. Gowers, Combinatorial theorems in sparse random sets, arXiv:1011.4310, 2011.

[37] R. Csákány and J. Kahn, A homological approach to two problems on finite sets, *J. Algebraic Combin.* **9** (1999), 141–149.

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

[39] D. de Caen, Linear constraints related to Turán's problem, Proceedings of the fourteenth Southeastern conference on combinatorics, graph theory and computing (Boca Raton, Fla., 1983), *Congr. Numer.* **39** (1983), 291–303.

[40] D. de Caen, On upper bounds for 3-graphs without tetrahedra, *Congressus Numer.* **62** (1988), 193–202.

[41] 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, pp. 187–197, János Bolyai Math. Soc., Budapest, 1994.

[42] D. de Caen and Z. Füredi, The maximum size of 3-uniform hypergraphs not containing a Fano plane, *J. Combin. Theory Ser. B* **78** (2000), 274-276.

[43] D. de Caen, D.L. Kreher and J. Wiseman, On constructive upper bounds for the Turán numbers T(n,2r+1,2r), Nineteenth Southeastern Conference on Combinatorics, Graph Theory, and Computing (Baton Rouge, LA, 1988), Congr. Numer. 65 (1988) 277–280.

[44] G.A. Dirac, Some theorems on abstract graphs, *Proc. Lond. Math. Soc.* **2** (1952), 69–81.

[45] R. Dotson and B. Nagle, Hereditary properties of hypergraphs, *J. Combin. Theory Ser. B* **99** (2009), 460–473.

[46] G. Elek and B. Szegedy, Limits of Hypergraphs, Removal and Regularity Lemmas. A Non-standard Approach, arXiv:0705.2179, 2007.

[47] G. Elek and B. Szegedy, A measure-theoretic approach to the theory of dense hypergraphs, arXiv:0810.4062, 2008.

[48] P. Erdős, On the number of complete subgraphs contained in certain graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **7** (1962), 459–464.

[49] P. Erdős, On a theorem of Rademacher-Turán, *Illinois J. Math* **6** (1962), 122–127.

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

[51] P. Erdős, Extremal problems in graph theory, Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963), pp. 29–36, Publ. House Czech. Acad. Sci., Prague, 1964.

[52] P. Erdős, Topics in combinatorial analysis, *Proc. Second Louisiana Conf. on Comb., Graph Theory and Computing*, 2–20, Louisiana State Univ., Baton Rouge 1971.

[53] P. Erdős, On some extremal problems on r-graphs, *Disc. Math.* **1** (1971), 1–6.

[54] P. Erdős, On the combinatorial problems which I would most like to see solved, *Combinatorica* **1** (1981), 25–42.

[55] P. Erdős, Some problems in graph theory, combinatorial analysis and combinatorial number theory, Graph Theory and Combinatorics, Academic Press (1984), 1-17.

[56] P. Erdős, P. Frankl and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs Combin.* **2** (1986), 113–121.

[57] P. Erdős, H. Ko and R. Rado, Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford Ser.* **12** (1961) 313–320.

[58] P. Erdős and M. Simonovits, A limit theorem in graph theory, *Studia Sci. Math. Hung. Acad.* **1**, 51–57, (1966).

[59] P. Erdős and M. Simonovits, Supersaturated graphs and hypergraphs, *Combinatorica* **3** (1982), 181–192.

[60] P. Erdős and V. T. Sós, On Ramsey-Turán type theorems for hypergraphs, *Combinatorica* **2** (1982), 289–295.

[61] P. Erdős and J. H. Spencer, *Probabilistic Methods in Combinatorics*, Academic Press, 1974.

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

[63] J. Faudree, R. Faudree and J. R. Schmitt, A survey of minimum saturated graphs and hypergraphs, preprint.

[64] D. G. Fon-Der-Flaass, A method for constructing (3, 4)-graphs, *Math. Zeitschrift* **44** (1988), 546–550.

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

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

[67] P. Frankl and Z. Füredi, An exact result for 3-graphs, *Disc. Math.* **50** (1984), 323–328.

[68] P. Frankl and Z. Füredi, Exact solution of some Turán-type problems, *J. Combin. Theory Ser. A.* **45** (1987), 226–262.

[69] P. Frankl and Z. Füredi, Extremal problems and the Lagrange function of hypergraphs, *Bulletin Institute Math. Academia Sinica* **16** (1988), 305–313.

[70] P. Frankl and Z. Füredi, Extremal problems whose solutions are the blowups of the small Witt-designs, *J. Combin. Theory Ser. A* **52** (1989), 129–147.

[71] P. Frankl, Y. Peng, V. Rödl and J. Talbot, A note on the jumping constant conjecture of Erdős, *J. Combin. Theory Ser. B* **97** (2007), 204–216.

[72] P. Frankl and V. Rödl, Hypergraphs do not jump, *Combinatorica* **4** (1984), 149–159.

[73] P. Frankl and V. Rödl, Lower bounds for Turán's problem, *Graphs Combin.* **1** (1985), 213–216.

[74] P. Frankl and V. Rödl, Large triangle-free subgraphs in graphs without K_4, *Graphs Combin.* **2** (1986), 135–144.

[75] P. Frankl and R. M. Wilson, Intersection theorems with geometric consequences, *Combinatorica* **1** (1981), 357–368.

[76] E. Friedgut and J. Kahn, On the number of copies of one hypergraph in another, *Israel J. Math.* **105** (1998), 251–256.

[77] A. Frieze and R. Kannan, Quick approximation to matrices and applications, *Combinatorica* **19** (1999), 175–220.

[78] A. Frohmader, More constructions for Turán's (3,4)-conjecture, *Electronic J. Combin.* **15** (2008), R137.

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

[80] Z. Füredi and A. Kündgen, Turán problems for integer-weighted graphs, *J. Graph Theory* **40** (2002), 195–225.

[81] Z. Füredi, D. Mubayi and O. Pikhurko, Quadruple systems with independent neighborhoods, *J. Comb. Theory Ser. A* **115** (2008), 1552–1560.

[82] Z. Füredi, O. Pikhurko and M. Simonovits, On triple systems with independent neighborhoods, *Combin. Probab. Comput.* **14** (2005), 795–813.

[83] Z. Füredi, O. Pikhurko and M. Simonovits, 4-books of three pages, *J. Comb. Theory Ser. A* **113** (2006), 882–891.

[84] Z. Füredi and M. Simonovits, Triple systems not containing a Fano configuration, *Combin. Probab. Comput.* **14** (2005), 467–484.

[85] G. R. Giraud, Remarques sur deux problèmes extrémaux, *Disc. Math.* **84** (1990), 319–321.

[86] J. Goldwasser, On the Turán Number of $\{123, 124, 345\}$, manuscript.

[87] W. T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, *Geom. Funct. Anal.* **7** (1997), 322–337.

[88] W. T. Gowers, Quasirandomness, counting and regularity for 3-uniform hypergraphs, *Combin. Probab. Comput.* **15** (2006), 143–184.

[89] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Ann. of Math.* **166** (2007), 897–946.

[90] B. J. Green and J. Wolf, A note on Elkin's improvement of Behrend's construction, arXiv:0810.0732, 2010.

[91] A. Hajnal and E. Szemerédi, Proof of a conjecture of Erdős, *Combinatorial Theory and Its Applications*, Amsterdam, Netherlands: North-Holland, pp. 601-623, 1970.

[92] H. Hatami and Serguei Norine, Undecidability of linear inequalities in graph homomorphism densities, to appear in *J. AMS*, arXiv:1005.2382, 2010.

[93] A. J. W. Hilton, On ordered set systems and some conjectures related to the Erdős-Ko-Rado Theorem and Turán's Theorem, *Mathematika* **28** (1981), 54–66.

[94] Y. Ishigami, A simple regularization of hypergraphs, arXiv:math/0612838, 2006.

[95] Y. Ishigami, The number of hypergraphs and colored hypergraphs with hereditary properties, arXiv:0712.0425, 2007.

[96] J. R. Johnson and J. Talbot, Vertex Turán problems in the hypercube, *J. Combin. Theory Ser. A* **117** (2010), 454–465.

[97] G. Kalai, Hyperconnectivity of graphs, *Graphs Combin.* **1** (1985), 65–79.

[98] G. Kalai, A new approach to Turán's conjecture, *Graphs Combin.* **1** (1985), 107–109.

[99] G. Kalai, Algebraic shifting, Computational commutative algebra and combinatorics (Osaka, 1999), 121–163, *Adv. Stud. Pure Math.* **33**, Math. Soc. Japan, Tokyo, 2002.

[100] G. Katona, A theorem of finite sets, *Theory of graphs* (Proc. Colloq., Tihany, 1966), 187–207, Academic Press, New York, 1968.

[101] G. Katona, T. Nemetz and M. Simonovits, On a problem of Turán in the theory of graphs, *Mat. Lapok* **15** (1964), 228–238.

[102] P. Keevash, The Turán problem for hypergraphs of fixed size, *Electron. J. Combin.* **12** (2005), N11.

[103] P. Keevash, The Turán problem for projective geometries, *J. Combin. Theory Ser. A* **111** (2005), 289–309.

[104] P. Keevash, A hypergraph regularity method for generalised Turán problems, *Random Structures Algorithms* **34** (2009), 123–164.

[105] T. Kövari, V. T. Sós and P. Turán, On a problem of K. Zarankiewicz, *Colloquium Math.* **3** (1954), 50–57.

[106] P. Keevash and D. Mubayi, Stability theorems for cancellative hypergraphs, *J. Combin. Theory Ser. B* **92** (2004), 163–175.

[107] P. Keevash and D. Mubayi, Set systems without a simplex or a cluster, *Combinatorica* **30** (2010), 175–200.

[108] P. Keevash, D. Mubayi, B. Sudakov and J. Verstraëte, Rainbow Turán problems, *Combin. Probab. Comput.* **16** (2007), 109–126.

[109] P. Keevash, M. Saks, B. Sudakov and J. Verstraëte, Multicolour Turan problems, *Adv. in Applied Math.* **33** (2004), 238–262.

[110] P. Keevash and B. Sudakov, The Turán number of the Fano plane, *Combinatorica* **25** (2005), 561-574.

[111] P. Keevash and B. Sudakov, On a hypergraph Turán problem of Frankl, *Combinatorica* **25** (2005), 673-706.

[112] P. Keevash and B. Sudakov, Set systems with restricted cross-intersections and the minimum rank of inclusion matrices, *SIAM J. Discrete Math.* **18** (2005), 713–727.

[113] P. Keevash and Y. Zhao, Codegree problems for projective geometries, *J. Combin. Theory Ser. B* **97** (2007), 919–928.

[114] K. Kim and F. Roush, On a problem of Turán, Studies in pure mathematics, 423-425, Birkhäuser, Basel-Boston, Mass. 1983.

[115] J. Komlós, G. N. Sárközy and E. Szemerédi, Proof of the Seymour conjecture for large graphs, *Ann. Comb.* **2** (1998), 43–60.

[116] J. Komlós and M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), 295–352, *Bolyai Soc. Math. Stud.* **2** János Bolyai Math. Soc., Budapest, 1996.

[117] A. Kostochka, A class of constructions for Turán's (3,4) problem, *Combinatorica* **2** (1982), 187–192.

[118] J. B. Kruskal, The number of simplices in a complex, *Mathematical optimization techniques*, 251–278 Univ. of California Press, Berkeley, Calif, 1963.

[119] D. Kühn and D. Osthus, Embedding large subgraphs into dense graphs, *Surveys in Combinatorics*, Cambridge University Press, 2009, 137–167.

[120] D. Kühn and D. Osthus, The minimum degree threshold for perfect graph packings, *Combinatorica* **29** (2009), 65–107.

[121] R.S. Li and W.W. Li, Independence numbers of graphs and generators of ideals, *Combinatorica* **1** (1981), 55–61.

[122] L. Lovász, Semidefinite programs and combinatorial optimization, *Recent advances in algorithms and combinatorics*, CMS Books Math. vol. 11, 137–194, Springer, New York, 2003.

[123] L. Lovász, Graph homomorphisms: Open problems, manuscript, 2008.

[124] L. Lovász, Subgraph densities in signed graphons and the local Sidorenko conjecture, arXiv:1004.3026, 2010.

[125] L. Lovász and M. Simonovits, On the number of complete subgraphs of a graph. II, Studies in pure mathematics, 459–495, Birkhäuser, Basel, 1983.

[126] L. Lovász and B. Szegedy, Limits of dense graph sequences, *J. Combin. Theory Ser. B* **96** (2006), 933–957.

[127] L. Lovász and B. Szegedy, Szemerédi's lemma for the analyst, *Geom. Funct. Anal.* **17** (2007), 252–270.

[128] L. Lu and L. Székely, Using Lovász Local Lemma in the space of random injections, *Electronic J. Combin.* **14** (2007), R63.

[129] L. Lu and Y. Zhao, An exact result for hypergraphs and upper bounds for the Turán density of K^r_{r+1}, *SIAM J. Disc. Math.* **23** (2009), 1324–1334.

[130] E. Marchant and A. Thomason, Extremal graphs and multigraphs with two weighted colours, *Fete of Combinatorics and Computer Science*, Bolyai Society Mathematical Studies, Vol. 20, 2010.

[131] K. Markström, Extremal hypergraphs and bounds for the Turán density of the 4-uniform K_5 *Disc. Math.* **309** (2009), 5231–5234.

[132] K. Markström and J. Talbot, On the density of 2-colorable 3-graphs in which any four points span at most two edges, *J. Combin. Des.* **18** (2010), 105–114.

[133] K. Markström, A web archive of Turán graphs, URL: http://abel.math.umu.se/klasm/Data/hypergraphs/turanhypergraphs.html.

[134] U. Matthias, Hypergraphen ohne vollstände r-partite Teilgraphen, PhD thesis, Heidelberg, 1994.

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

[136] D. Mubayi, On hypergraphs with every four points spanning at most two triples, *Electronic J. Combin.* **10** (2003), N10.

[137] D. Mubayi, The co-degree density of the Fano plane, *J. Combin. Theory Ser. B* **95** (2005), 333–337.

[138] D. Mubayi, A hypergraph extension of Turán's theorem, *J. Combin. Theory Ser. B* **96** (2006), 122–134.

[139] D. Mubayi, Counting substructures I: color critical graphs, to appear in *Adv. Math.*

[140] D. Mubayi, Counting substructures II: hypergraphs, preprint.

[141] D. Mubayi and O. Pikhurko, A new generalization of Mantel's theorem to k-graphs, *J. Combin Theory Ser. B* **97** (2007), 669–678.

[142] D. Mubayi and O. Pikhurko, Constructions of non-principal families in extremal hypergraph theory, *Disc. Math.* **308** (2008), 4430–4434.

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

[144] D. Mubayi and J. Talbot, Extremal problems for t-partite and t-colorable hypergraphs, *Electronic J. Combin.* **15** (2008), R26.

[145] D. Mubayi and J. Verstraëte, Proof of a conjecture of Erdős on triangles in set-systems, *Combinatorica* **25** (2005), 599–614.

[146] D. Mubayi and Y. Zhao, Co-degree density of hypergraphs, *J. Combin. Theory Ser. A* **114** (2007), 1118–1132.

[147] V. Nikiforov, The number of cliques in graphs of given order and size, to appear in *Trans. AMS*.

[148] B. Nagle, V. Rödl and M. Schacht, The counting lemma for regular k-uniform hypergraphs, *Random Structures Algorithms* **28** (2006), 113–179.

[149] J. Oxley, *Matroid Theory*, Oxford University Press, 1992.

[150] Y. Peng, Non-jumping numbers for 4-uniform hypergraphs, Graphs Combin. **23** (2007), 97–110.

[151] Y. Person and M. Schacht, Almost all hypergraphs without Fano planes are bipartite, *Proc. Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 09)*, 217–226.

[152] O. Pikhurko, An exact Turán result for the generalized triangle, *Combinatorica* **28** (2008) 187–208.

[153] O. Pikhurko, An analytic approach to stability, *Disc. Math.* **310** (2010), 2951–2964.

[154] O. Pikhurko, The minimum size of 3-graphs without a 4-set spanning no or exactly three edges, preprint.

[155] O. Pikhurko, Exact computation of the hypergraph Turán function for expanded complete 2-graphs, accepted by *J. Combin Theory Ser. B*, publication suspended for an indefinite time, see http://www.math.cmu.edu/ pikhurko/Copyright.html

[156] A. A. Razborov, Flag Algebras, J. Symbolic Logic **72** (2007), 1239–1282,

[157] A. A. Razborov, On the Minimal Density of Triangles in Graphs, *Combin. Probab. Comput.* **17** (2008), 603–618.

[158] A. A. Razborov, On 3-hypergraphs with forbidden 4-vertex configurations, *SIAM J. Disc. Math.* **24** (2010), 946–963.

[159] A. A. Razborov, On the Fon-der-Flaass interpretation of extremal examples for Turán's (3,4)-problem, arXiv:1008.4707, 2010.

[160] V. Rödl, A. Ruciński and E. Szemerédi, An approximate Dirac-type theorem for k-uniform hypergraphs, *Combinatorica* **28** (2008), 229–260.

[161] V. Rödl and A. Ruciński, Dirac-type questions for hypergraphs – a survey (or more problems for Endre to solve), An Irregular Mind (Szemerédi is 70), *Bolyai Soc. Math. Studies* **21** (2010).

[162] V. Rödl and M. Schacht, Regular partitions of hypergraphs: regularity lemmas, *Combin. Probab. Comput.* **16** (2007), 833–885.

[163] V. Rödl and J. Skokan, Regularity lemma for uniform hypergraphs, *Random Structures Algorithms* **25** (2004), 1–42.

[164] V. Rödl and A. Sidorenko, On the jumping constant conjecture for multigraphs, *J. Combin. Theory Ser. A* **69** (1995), 347–357.

[165] I. Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, pp. 939–945, *Colloq. Math. Soc. János Bolyai* **18**, North-Holland, Amsterdam-New York, 1978.

[166] T. Sanders, On certain other sets of integers, arXiv:1007.5444, 2010.

[167] G. N. Sárközy and S. M. Selkow, An extension of the Ruzsa-Szemerédi Theorem, *Combinatorica* **25** (2005), 77–84.

[168] G. N. Sárközy and S. M. Selkow, On a Turán-type hypergraph problem of Brown, Erdos and T. S´s, *Disc. Math.* **297** (2005), 190–195.

[169] M. Schacht, Extremal results for random discrete structures, preprint.

[170] J. B. Shearer, A new construction for cancellative families of sets, *Electron. J. Combin.* **3** (1996), R15.

[171] A. F. Sidorenko, Systems of sets that have the T-property, *Moscow University Mathematics Bulletin* **36** (1981), 22–26.

[172] A. F. Sidorenko, Method of quadratic forms in the Turán combinatorial problem, *Moscow University Mathematics Bulletin* **37** (1982), 3–6.

[173] 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*.

[174] A. F. Sidorenko, Asymptotic solution for a new class of forbidden r-graphs. *Combinatorica* **9** (1989), 207–215.

[175] A. F. Sidorenko, Extremal combinatorial problems in spaces with continuous measure (in Russian), *Issledovanie Operatchiy i ASU* **34** (1989), 34–40.

[176] A. Sidorenko, Asymptotic solution of the Turán problem for some hypergraphs, *Graphs Combin.* **8** (1992), 199–201.

[177] A. Sidorenko, A correlation inequality for bipartite graphs, *Graphs Combin.* **9** (1993), 201–204.

[178] A. Sidorenko, Boundedness of optimal matrices in extremal multigraph and digraph problems, *Combinatorica* **13** (1993), 109–120.

[179] A. 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.

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

[181] A. Sidorenko, Upper bounds for Turán numbers, *J. Combin. Theory Ser. A* **77** (1997), 134–147.

[182] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, *Theory of Graphs (Proc. Colloq. Tihany, 1966)*, Academic Press, New York, and Akad. Kiadó, Budapest, 1968, 279–319.

[183] V. Sós, Remarks on the connection of graph theory, finite geometry and block designs, *Teorie Combinatorie*, Tomo II, Accad. Naz. Lincei, Rome, 1976, 223–233.

[184] V. T. Sós and E. G. Straus, Extremals of functions on graphs with applications to graphs and hypergraphs. *J. Combin. Theory Ser. B* **32** (1982), 246–257.

[185] B. Sudakov and V. Vu, Local resilience of graphs, *Random Structures Algorithms* **33** (2008), 409–433.

[186] J. Talbot, Chromatic Turán problems and a new upper bound for the Turán density of K_4^-, *European J. Combin.* **28** (2007), 2125–2142.

[187] T. Tao, A variant of the hypergraph removal lemma, *J. Combin. Theory Ser. A* **113** (2006), 1257–1280.

[188] T. Tao, A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma, *J. Anal. Math.* **103** (2007), 1–45.

[189] A. Thomason and P. Wagner, Bounding the size of square-free subgraphs of the hypercube, *Disc. Math.* **309** (2009), 1730–1735.

[190] P. Turán, On an extremal problem in graph theory (in Hungarian), *Mat. Fiz. Lapok* **48** (1941), 436–452.

[191] P. Turán, Research problem, *Közl MTA Mat. Kutató Int.* **6** (1961), 417–423.

[192] B. Zhou, A Turán-type problem on 3-graphs, *Ars Combin.* **31** (1991), 177–181.

[193] A. A. Zykov, On some properties of linear complexes, *Mat. Sbornik (N. S.)* **24** (66) (1949) 163–188.

School of Mathematical Sciences,
Queen Mary, University of London,
Mile End Road,
London E1 4NS,
UK.
p.keevash@qmul.ac.uk