Graph removal lemmas

David Conlon[*]   Jacob Fox[†]

Abstract

The graph removal lemma states that any graph on n vertices with $o(n^{v(H)})$ copies of a fixed graph H may be made H-free by removing $o(n^2)$ edges. Despite its innocent appearance, this lemma and its extensions have several important consequences in number theory, discrete geometry, graph theory and computer science. In this survey we discuss these lemmas, focusing in particular on recent improvements to their quantitative aspects.

1   Introduction

The triangle removal lemma states that for all $\epsilon > 0$ there exists $\delta > 0$ such that any graph on n vertices with at most δn^3 triangles may be made triangle-free by removing at most ϵn^2 edges. This result, proved by Ruzsa and Szemerédi [93] in 1976, was originally stated in rather different language.

The original formulation was in terms of the $(6,3)$-problem.[1] This asks for the maximum number of edges $f^{(3)}(n,6,3)$ in a 3-uniform hypergraph on n vertices such that no 6 vertices contain 3 edges. Answering a question of Brown, Erdős and Sós [19], Ruzsa and Szemerédi showed that $f^{(3)}(n,6,3) = o(n^2)$. Their proof used several iterations of an early version of Szemerédi's regularity lemma [110].

This result, developed by Szemerédi in his proof of the Erdős-Turán conjecture on arithmetic progressions in dense sets [109], states that every graph may be partitioned into a small number of vertex sets so that the graph between almost every pair of vertex sets is random-like. Though this result now occupies a central position in graph theory, its importance only emerged over time. The resolution of the $(6,3)$-problem was one of the first indications of its strength.

The Ruzsa-Szemerédi theorem was generalized by Erdős, Frankl and Rödl [32], who showed that $f^{(r)}(n, 3r - 3, 3) = o(n^2)$, where $f^{(r)}(n, 3r - 3, 3)$ is the maximum number of edges in an r-uniform hypergraph such that no $3r - 3$ vertices contain 3 edges. One of the tools used by Erdős, Frankl and Rödl in their proof was a striking result stating that if a graph on n vertices contains no copy of a graph H then it may be made K_r-free, where $r = \chi(H)$ is the chromatic number of H, by removing $o(n^2)$ edges. The proof of this result used the modern formulation of Szemerédi's regularity lemma and is already very close, both in proof and statement, to the following generalization of the triangle

---

[1]The two results are not exactly equivalent, though the triangle removal lemma may be proved by their method. A weak form of the triangle removal lemma, already sufficient for proving Roth's theorem, is equivalent to the Ruzsa-Szemerédi theorem. This weaker form states that any graph on n vertices in which every edge is contained in exactly one triangle has $o(n^2)$ edges. This is also equivalent to another attractive formulation, known as the induced matching theorem. This states that any graph on n vertices which is the union of at most n induced matchings has $o(n^2)$ edges.

removal lemma, known as the graph removal lemma.[2] This was first stated explicitly in the literature by Alon, Duke, Lefmann, Rödl and Yuster [4] and by Füredi [47] in 1994.[3] Note that we use $v(H)$ to denote the number of vertices in a graph (or hypergraph) H.

Theorem 1.1 *For any graph H and any $\epsilon > 0$, there exists $\delta > 0$ such that any graph on n vertices which contains at most $\delta n^{v(H)}$ copies of H may be made H-free by removing at most ϵn^2 edges.*

It was already observed by Ruzsa and Szemerédi that the $(6, 3)$-problem (and, thereby, the triangle removal lemma) is related to Roth's theorem on arithmetic progressions [91]. This theorem states that for any $\delta > 0$ there exists an n_0 such that if $n \geq n_0$ then any subset of the set $[n] := \{1, 2, \ldots, n\}$ of size at least δn contains an arithmetic progression of length 3. Letting $r_3(n)$ be the largest integer such that there exists a subset of the set $\{1, 2, \ldots, n\}$ of size $r_3(n)$ containing no arithmetic progression of length 3, this is equivalent to saying that $r_3(n) = o(n)$. Ruzsa and Szemerédi observed that $f^{(3)}(n, 6, 3) = \Omega(r_3(n)n)$. In particular, since $f^{(3)}(n, 6, 3) = o(n^2)$, this implies that $r_3(n) = o(n)$, yielding a proof of Roth's theorem.

It was further noted by Solymosi [104] that the Ruzsa-Szemerédi theorem yields a stronger result of Ajtai and Szemerédi [1]. This result states that for any $\delta > 0$ there exists an n_0 such that if $n \geq n_0$ then any subset of the set $[n] \times [n]$ of size at least δn^2 contains a set of the form $\{(a, b), (a + d, b), (a, b + d)\}$. That is, dense subsets of the 2-dimensional grid contain axis-parallel isosceles triangles. Roth's theorem is a simple corollary of this statement.

Roth's theorem is the first case of a famous result known as Szemerédi's theorem. This result, to which we alluded earlier, states that for any natural number $k \geq 3$ and any $\delta > 0$ there exists n_0 such that if $n \geq n_0$ then any subset of the set $[n]$ of size at least δn contains an arithmetic progression of length k. This was first proved by Szemerédi [109] in the early seventies using combinatorial techniques and since then several further proofs have emerged. The most important of these are that by Furstenberg [48, 50] using ergodic theory and that by Gowers [53, 54], who found a way to extend Roth's original Fourier analytic argument to general k. Both of these methods have been highly influential.

Yet another proof technique was suggested by Frankl and Rödl [42]. They showed that Szemerédi's theorem would follow from the following generalization of Theorem 1.1, referred to as the hypergraph removal lemma. They proved this theorem for the specific case of $K_4^{(3)}$, the complete 3-uniform hypergraph with 4 vertices. This was then extended to all 3-uniform hypergraphs in [77] and to $K_5^{(4)}$ in [89]. Finally, it was proved for all hypergraphs by Gowers [55, 56] and, independently, by Nagle, Rödl, Schacht and Skokan [78, 88]. Both proofs rely on extending Szemerédi's regularity lemma to hypergraphs in an appropriate fashion.

Theorem 1.2 *For any k-uniform hypergraph $\mathcal{H}$ and any $\epsilon > 0$, there exists $\delta > 0$ such that any k-uniform hypergraph on n vertices which contains at most $\delta n^{v(\mathcal{H})}$ copies of $\mathcal{H}$ may be made $\mathcal{H}$-free by removing at most ϵn^k edges.*

---

[2]The phrase 'removal lemma' is a comparatively recent coinage. It seems to have come into vogue in about 2005 when the hypergraph removal lemma was first proved (see, for example, [67, 78, 106, 112]).

[3]This was also the first time that the triangle removal lemma was stated explicitly, though the weaker version concerning graphs where every edge is contained in exactly one triangle had already appeared in the literature. The Ruzsa-Szemerédi theorem was usually [40, 41, 46] phrased in the following suggestive form: if a 3-uniform hypergraph is linear, that is, no two edges intersect on more than a single vertex, and triangle-free, then it has $o(n^2)$ edges. A more explicit formulation may be found in [23].

As well as reproving Szemerédi's theorem, the hypergraph removal lemma allows one to reprove the multidimensional Szemerédi theorem. This theorem, originally proved by Furstenberg and Katznelson [49], states that for any natural number r, any finite subset S of $\mathbb{Z}^r$ and any $\delta > 0$ there exists n_0 such that if $n \geq n_0$ then any subset of $[n]^r$ of size at least δn^r contains a subset of the form $a \cdot S + d$, that is, a dilated and translated copy of S. That it follows from the hypergraph removal lemma was first observed by Solymosi [105]. This was the first non-ergodic proof of this theorem. A new proof of the special case $S = \{(0,0), (1,0), (0,1)\}$, corresponding to the Ajtai-Szemerédi theorem, was given by Shkredov [102] using a Fourier analytic argument. Recently, a combinatorial proof of the density Hales-Jewett theorem, which is an extension of the multidimensional Szemerédi theorem, was discovered as part of the polymath project [81].

As well as its implications in number theory, the removal lemma and its extensions are central to the area of computer science known as property testing. In this area, one would like to find fast algorithms to distinguish between objects which satisfy a certain property and objects which are far from satisfying that property. This field of study was initiated by Rubinfield and Sudan [92] and, subsequently, Goldreich, Goldwasser and Ron [51] started the investigation of such property testers for combinatorial objects. Graph property testing has attracted a particular degree of interest.

A classic example of property testing is to decide whether a given graph G is ϵ-far from being triangle-free, that is, whether at least ϵn^2 edges will have to removed in order to make it triangle-free. The triangle removal lemma tells us that if G is ϵ-far from being triangle free then it must contain at least δn^3 triangles for some $\delta > 0$ depending only on ϵ. This furnishes a simple probabilistic algorithm for deciding whether G is ϵ-far from being triangle-free. We choose $t = 2\delta^{-1}$ triples of points from the vertices of G uniformly at random. If G is ϵ-far from being triangle-free then the probability that none of these randomly chosen triples is a triangle is $(1-\delta)^t < e^{-t\delta} < \frac{1}{3}$. That is, if G is ϵ-far from being triangle-free we will find a triangle with probability at least $\frac{2}{3}$, whereas if G is triangle-free we will clearly find no triangles. The graph removal lemma may be used to derive a similar test for deciding whether G is ϵ-far from being H-free for any fixed graph H.

In property testing, it is often of interest to decide not only whether a graph is far from being H-free but also whether it is far from being induced H-free. A subgraph H' of a graph G is said to be an induced copy of H if there is a one-to-one map $f : V(H) \to V(H')$ such that $(f(u), f(v))$ is an edge of H' if and only if (u, v) is an edge of H. A graph G is said to be induced H-free if it contains no induced copies of H and ϵ-far from being induced H-free if we have to add and/or delete at least ϵn^2 edges to make it induced H-free. Note that it is not enough to delete edges since, for example, if H is the empty graph on two vertices and G is the complete graph minus an edge, then G contains only one induced copy of H, but one cannot simply delete edges from G to make it induced H-free.

By proving an appropriate strengthening of the regularity lemma, Alon, Fischer, Krivelevich and Szegedy [6] showed how to modify the graph removal lemma to this setting. This result, which allows one to test for induced H-freeness, is known as the induced removal lemma.

Theorem 1.3 *For any graph H and any $\epsilon > 0$, there exists a $\delta > 0$ such that any graph on n vertices which contains at most $\delta n^{v(H)}$ induced copies of H may be made induced H-free by adding and/or deleting at most ϵn^2 edges.*

A substantial generalization of this result, known as the infinite removal lemma, was proved by Alon and Shapira [12] (see also [75]). They showed that for each (possibly infinite) family $\mathcal{H}$ of graphs and $\epsilon > 0$ there is $\delta = \delta_{\mathcal{H}}(\epsilon) > 0$ such that if a graph G on n vertices contains at most $\delta n^{v(H)}$ induced

copies of H for every graph H in $\mathcal{H}$, then G may be made induced H-free, for every $H \in \mathcal{H}$, by adding and/or deleting at most ϵn^2 edges. They then used this result to show that every hereditary graph property is testable, where a graph property is hereditary if it is closed under removal of vertices. These results were extended to 3-uniform hypergraphs by Avart, Rödl and Schacht [14] and to k-uniform hypergraphs by Rödl and Schacht [86].

In this survey we will focus on recent developments, particularly with regard to the quantitative aspects of the removal lemma. In particular, we will discuss recent improvements on the bounds for the graph removal lemma, Theorem 1.1, and the induced graph removal lemma, Theorem 1.3, each of which bypasses a natural impediment.

The usual proof of the graph removal lemma makes use of the regularity lemma and gives bounds for the removal lemma which are of tower-type in ϵ. To be more specific, let $T(1) = 2$ and, for each $i \geq 1$, $T(i+1) = 2^{T(i)}$. The bounds that come out of applying the regularity lemma to removal then say that if $\delta^{-1} = T(\epsilon^{-c_H})$ then any graph with at most $\delta n^{v(H)}$ copies of H may be made H-free by removing at most ϵn^2 edges. Moreover, this tower-type dependency is inherent in any proof employing regularity. This follows from an important result of Gowers [52] (see also [24]) which states that the bounds that arise in the regularity lemma are necessarily of tower type. We will discuss this in more detail in Section 2.1 below.

Despite this obstacle, the following improvement was made by Fox [38].

Theorem 1.4 *For any graph H, there exists a constant a_H such that if $\delta^{-1} = T(a_H \log \epsilon^{-1})$ then any graph on n vertices which contains at most $\delta n^{v(H)}$ copies of H may be made H-free by removing at most ϵn^2 edges.*

As is implicit in the bounds, the proof of this theorem does not make an explicit appeal to Szemerédi's regularity lemma. However, many of the ideas used are similar to ideas used in the proof of the regularity lemma. The chief difference lies in the fact that the conditions of the removal lemma (containing few copies of a given graph H) allow us to say more about the structure of these partitions. A simplified proof of this theorem will be the main topic of Section 2.2.

Though still of tower-type, Theorem 1.4 improves substantially on the previous bound. However, it remains very far from the best known lower bound on δ^{-1}. The observation of Ruzsa and Szemerédi [93] that $f^{(3)}(n, 6, 3) = \Omega(r_3(n)n)$ allows one to transfer lower bounds for $r_3(n)$ to a corresponding lower bound for the triangle removal lemma. The best construction of a set containing no arithmetic progression of length 3 is due to Behrend [16] and gives a subset of $[n]$ with density $e^{-c\sqrt{\log n}}$. Transferring this to the graph setting yields a graph containing $\epsilon^{c \log \epsilon^{-1}} n^3$ triangles which cannot be made triangle-free by removing fewer than ϵn^2 edges. This quasi-polynomial lower bound, $\delta^{-1} \geq \epsilon^{-c \log \epsilon^{-1}}$, remains the best known.[4]

The standard proof of the induced removal lemma uses the strong regularity lemma of Alon, Fischer, Krivelevich and Szegedy [6]. We will speak at length about this result in Section 3.1. Here it will suffice to say that, like the ordinary regularity lemma, the bounds which an application of this theorem

---

[4]It is worth noting that the best known upper bound for Roth's theorem, due to Sanders [95], is considerably better than the best upper bound for $r_3(n)$ that follows from triangle removal. This upper bound is $r_3(n) = O\left(\frac{(\log \log n)^5}{\log n} n\right)$. A recent result of Schoen and Shkredov [99], building on further work of Sanders [96], shows that any subset of $[n]$ of density $e^{-c(\frac{\log n}{\log \log n})^{1/6}}$ contains a solution to the equation $x_1 + \cdots + x_5 = 5x_6$. Since arithmetic progressions correspond to solutions of $x_1 + x_2 = 2x_3$, this suggests that the answer should be closer to the Behrend bound. The bounds for triangle removal are unlikely to impinge on these upper bounds for some time, if at all.

4

gives for the induced removal lemma are necessarily very large. Let $W(1) = 2$ and, for $i \geq 1$, $W(i+1) = T(W(i))$. This is known as the wowzer function and its values dwarf those of the usual tower function.[5] By using the strong regularity lemma, the standard proof shows that we may take $\delta^{-1} = W(a_H \epsilon^{-c})$ in the induced removal lemma, Theorem 1.3. Moreover, as with the ordinary removal lemma, such a bound is inherent in the application of the strong regularity lemma. This follows from recent results of Conlon and Fox [24] and, independently, Kalyanasundaram and Shapira [61] showing that the bounds arising in strong regularity are necessarily of wowzer type.

In the other direction, Conlon and Fox [24] showed how to bypass this obstacle and prove that the bounds for δ^{-1} are at worst a tower in a power of ϵ^{-1}.

Theorem 1.5 *There exists a constant $c > 0$ such that, for any graph H, there exists a constant a_H such that if $\delta^{-1} = T(a_H \epsilon^{-c})$ then any graph on n vertices which contains at most $\delta n^{v(H)}$ induced copies of H may be made induced H-free by adding and/or deleting at most ϵn^2 edges.*

A discussion of this theorem will form the subject of Section 3.2. The key observation here is that the strong regularity lemma is used to prove an intermediate statement (Lemma 3.2 below) which then implies the induced removal lemma. This intermediate statement may be proved without recourse to the full strength of the strong regularity lemma. There are also some strong parallels with the proof of Theorem 1.4 which we will draw attention to in due course.

In Section 3.3, we present the proof of Alon and Shapira's infinite removal lemma. In another paper, Alon and Shapira [11] showed that the dependence in the infinite removal lemma can depend heavily on the family $\mathcal{H}$. They proved that for every function $\delta : (0,1) \to (0,1)$, there exists a family $\mathcal{H}$ of graphs such that any $\delta_{\mathcal{H}} : (0,1) \to (0,1)$ which satisfies the infinite removal lemma for $\mathcal{H}$ satisfies $\delta_{\mathcal{H}} = o(\delta)$. However, such examples are rather unusual and the proof presented in Section 3.3 of the infinite removal lemma implies that for many commonly studied families $\mathcal{H}$ of graphs the bound on $\delta_{\mathcal{H}}^{-1}$ is only tower-type, improving the wowzer-type bound from the original proof.

Our discussions of the graph removal lemma and the induced removal lemma will occupy the bulk of this survey but we will also talk about some further recent developments in the study of removal lemmas. These include arithmetic removal lemmas (Section 4) and the recently developed sparse removal lemmas which hold for subgraphs of sparse random and pseudorandom graphs (Section 5). We will conclude with some further comments on related topics.

2 The graph removal lemma

In this section we will discuss the two proofs of the removal lemma, Theorem 1.1, at length. In Section 2.1, we will talk about the regularity lemma and the usual proof of the removal proof. Then, in Section 2.2, we will consider a simplified variant of the second author's recent proof [38], showing how it connects to the weak regularity lemma of Frieze and Kannan [44, 45].

2.1 The standard proof

We begin with the proof of the regularity lemma and then deduce the removal lemma. For vertex subsets S, T of a graph G, we let $e_G(S, T)$ denote the number of pairs in $S \times T$ that are edges of G and

---

[5]To give some indication, we note that $W(2) = 4$, $W(3) = 65536$ and $W(4)$ is a tower of 2s of height 65536.

$d_G(S, T) = \frac{e_G(S,T)}{|S||T|}$ denote the fraction of pairs in $S \times T$ that are edges of G. For simplicity of notation, we drop the subscript if the graph G is clear from context. A pair (S, T) of subsets is ϵ-*regular* if, for all subsets $S' \subset S$ and $T' \subset T$ with $|S'| \geq \epsilon|S|$ and $|T'| \geq \epsilon|T|$, we have $|d(S', T') - d(S, T)| \leq \epsilon$. Informally, a pair of subsets is ϵ-regular with a small ϵ if the edges between S and T are uniformly distributed among large subsets.

Let $G = (V, E)$ be a graph and $P : V = V_1 \cup \ldots \cup V_k$ be a vertex partition of G. The partition of P is *equitable* if each pair of parts differ in size by at most 1. The partition P is ϵ-*regular* if all but at most ϵk^2 pairs of parts (V_i, V_j) are ϵ-regular. We next state Szemerédi's regularity lemma.

Lemma 2.1 *For every $\epsilon > 0$, there is $K = K(\epsilon)$ such that every graph $G = (V, E)$ has an equitable, ϵ-regular vertex partition into at most K parts. Moreover, we may take K to be a tower of height $O(\epsilon^{-5})$.*

Let $q : [0, 1] \to \mathbb{R}$ be a convex function. For vertex subsets $S, T \subset V$ of a graph G, let $q(S, T) = q(d(S, T))\frac{|S||T|}{|V|^2}$. For partitions $\mathcal{S} : S = S_1 \cup \ldots \cup S_a$ and $\mathcal{T} : T = T_1 \cup \ldots \cup T_b$, let $q(\mathcal{S}, \mathcal{T}) = \sum_{1 \leq i \leq a, 1 \leq j \leq b} q(S_i, T_j)$. For a vertex partition $P : V = V_1 \cup \ldots \cup V_k$ of G, define the mean-q density to be

$$q(P) = q(P, P) = \sum_{1 \leq i, j \leq k} q(V_i, V_j).$$

We next state some simple properties which follow from Jensen's inequality using the convexity of q. A *refinement* of a partition P of a vertex set V is another partition Q of V such that every part of Q is a subset of a part of P.

Proposition 2.1   *1. For partitions $\mathcal{S}$ and $\mathcal{T}$ of vertex subsets S and T, we have $q(\mathcal{S}, \mathcal{T}) \geq q(S, T)$.*

   2. If Q is a refinement of P, then $q(Q) \geq q(P)$.

   3. If $d = d(G) = d(V, V)$ is the edge density of G, then, for any vertex partition P,

$$q(d) \leq q(P) \leq dq(1) + (1 - d)q(0).$$

The first and second part of Proposition 2.1 show that by refining a vertex partition the mean-q density cannot decrease, while the last part gives the range of possible values for $q(P)$ if we only know the edge density d of G.

The convex function $q(x) = x^2$ for $x \in [0, 1]$ is chosen in the standard proof of the graph regularity lemma and we will do the same for the rest of this subsection. The following lemma is the key claim for the proof of the regularity lemma. The set-up is that we have a partition P which is not ϵ-regular. For each pair (V_i, V_j) of parts of P which is not ϵ-regular, there are a pair of witness subsets V_{ij}, V_{ji} to the fact that the pair of parts is not ϵ-regular. We consider the coarsest refinement Q of P so that each witness subset is the union of parts of Q. The lemma concludes that the number of parts of Q is at most exponential in the number of parts of P and, using a Cauchy-Schwarz defect inequality, that the mean-q density of the partition Q is substantially larger than the mean-q density of P. Because it simplifies our calculations a little, we will assume, when we say a partition is equitable, that it is exactly equitable, that is, that all parts have precisely the same size. This does not affect our results substantially but simplifies the presentation.

Lemma 2.2 *If an equitable partition $P : V = V_1 \cup \ldots \cup V_k$ is not ϵ-regular then there is a refinement Q of P into at most $k2^k$ parts for which $q(Q) \geq q(P) + \epsilon^5$.*

Proof: For each pair (V_i, V_j) which is not ϵ-regular, there are subsets $V_{ij} \subset V_i$ and $V_{ji} \subset V_j$ with $|V_{ij}| \geq \epsilon|V_i|$ and $|V_{ji}| \geq \epsilon|V_j|$ such that $|d(V_{ij}, V_{ji}) - d(V_i, V_j)| \geq \epsilon$. For each part V_j such that (V_i, V_j) is not ϵ-regular, we have a partiton P_{ij} of V_i into two parts V_{ij} and $V_i \setminus V_{ij}$. Let P_i be the partition of V_i which is the common refinement of these at most $k - 1$ partitions of V_i, so P_i has at most 2^{k-1} parts. We let Q be the partition of V which is the union of the k partitions of the form P_i, so Q has at most $k2^{k-1}$ parts. We have

$$
\begin{aligned}
q(Q) - q(P) &= \sum_{i,j} \left(q(P_i, P_j) - q(V_i, V_j) \right) \\
&\geq \sum_{(V_i,V_j) \text{ irregular}} \left(q(P_i, P_j) - q(V_i, V_j) \right) \\
&\geq \sum_{(V_i,V_j) \text{ irregular}} \left(q(P_{ij}, P_{ji}) - q(V_i, V_j) \right) \\
&= \sum_{(V_i,V_j) \text{ irregular}} \sum_{U \in P_{ij}, W \in P_{ji}} \frac{|U||W|}{|V|^2} \left(d(U, W) - d(V_i, V_j) \right)^2 \\
&\geq \sum_{(V_i,V_j) \text{ irregular}} \frac{|V_{ij}||V_{ji}|}{|V|^2} \left(d(V_{ij}, V_{ji}) - d(V_i, V_j) \right)^2 \\
&\geq \epsilon k^2 \left(\frac{\epsilon}{k} \right)^2 \epsilon^2 \\
&= \epsilon^5,
\end{aligned}
$$

where the first and third inequalities are by noting that the summands are nonnegative and the second inequality follows from the first part of Proposition 2.1, which shows that the mean-q density cannot decrease when taking a refinement. In the fourth inequality, we used that $|V_{ij}| \geq \epsilon|V_i| \geq \frac{\epsilon}{k}|V|$ and similarly for $|V_{ji}|$. Finally, the equality in the fourth line follows from the identity

$$
\sum_{U \in P_{ij}, W \in P_{ji}} |U||W|d(V_i, V_j) = \sum_{U \in P_{ij}, W \in P_{ji}} |U||W|d(U, W),
$$

which counts $e(V_i, V_j)$ in two different ways. This completes the proof. $\square$

The next lemma, which is rather standard, shows that for any vertex partition Q, there is a vertex equipartition P' with a similar number of parts to Q and mean-square density not much smaller than the mean-square density of Q. It is useful in density increment arguments where at each stage one would like to work with an equipartition. It is proved by first arbitrarily partitioning each part of Q into parts of order $|V|/t$, except possibly one additional remaining smaller part, and then arbitrarily partitioning the union of the smaller remaining parts into parts of order $|V|/t$.

Lemma 2.3 *Let $G = (V, E)$ be a graph and $Q : V = V_1 \cup \ldots \cup V_\ell$ be a vertex partition into k parts. Then, for $q(x) = x^2$, there is an equitable partition P' of V into t parts such that $q(P') \geq q(Q) - 2\frac{\ell}{t}$.*

Combining Lemmas 2.2 and 2.3 with $t = 4\epsilon^{-5}|Q| \leq \epsilon^{-5}k2^{k+2}$, we obtain the following corollary.

Corollary 2.1 *If an equitable partition $P : V = V_1 \cup \ldots \cup V_k$ is not ϵ-regular then there is an equitable refinement P' of P into at most $\epsilon^{-5} k 2^{k+2}$ parts for which $q(P') \geq q(P) + \epsilon^5/2$.*

We next show how Szemerédi's regularity lemma, Lemma 2.1, can be quickly deduced from this result.

Proof: To prove the regularity lemma, we start with the trivial partition P_0 into one part, and iterate the above corollary to obtain a sequence $P_0, P_1, \ldots, P_s$ of equitable partitions with $q(P_{i+1}) \geq q(P_i) + \epsilon^5/2$ until we arrive at an equitable ϵ-regular partition P_s. As the mean-square density of each partition has to lie between 0 and 1, after at most $2\epsilon^{-5}$ iterations we arrive at the equitable ϵ-regular partition P_s with $s \leq 2\epsilon^{-5}$. The number of parts increases by one exponential in each iteration, giving the desired number of parts in the regularity partition. This completes the proof of Szemerédi's regularity lemma. □

The constructions of Gowers [52] and the authors [24] show that the tower-type bound on the number of parts in Szemerédi's regularity lemma is indeed necessary. In particular, the construction in [24] shows that $K(\epsilon)$ in Lemma 2.1 is at least a tower of twos of height $\Omega(\epsilon^{-1})$. The constructions are formed by reverse engineering the upper bound proof. We construct a sequence $P_0, \ldots, P_s$ of partitions with $s = \Omega(\epsilon^{-1})$. As in the upper bound proof, each partition in the sequence uses exponentially more parts than the previous partition in the sequence. We may choose the edges using these partitions and some randomness so as to guarantee that none of these partitions (except the last) are ϵ-regular. Furthermore, we can guarantee that any partition that is ϵ-regular must be close to being a refinement of the last partition in this sequence. This implies that the number of parts must be at least roughly $|P_s|$.

We next prove the graph removal lemma, Theorem 1.1, from the regularity lemma.

Proof: Let m denote the number of edges of H, so $m \leq \binom{h}{2}$. Let $\gamma = \frac{\epsilon^h}{4h}$ and $\delta = (2h)^{-2h} \epsilon^m K^{-h}$, where $K = K(\gamma)$ is as in the regularity lemma. We apply the regularity lemma to G and obtain an equitable, γ-regular partition into $k \leq K$ parts. If the number n of vertices of G satisfies $n < \delta^{-1/h}$, then the number of copies of H in G is at most $\delta n^h < 1$ and G is H-free, in which case there is nothing to prove. So we may assume $n \geq \delta^{-1/h}$. We obtain a subgraph G' of G by removing edges of G between all pairs of parts which are not γ-regular or which have edge density at most ϵ. As there are at most γk^2 ordered pairs of parts which are not γ-regular and each part has order at most $2n/k$, at most $(\gamma k^2/2)(2n/k)^2 = 2\gamma n^2$ edges are deleted between pairs of parts which are not γ-regular. The number of edges between parts which have edge density at most ϵ is at most $\epsilon n^2/2$. Hence, the number of edges of G deleted to obtain G' is at most $2\gamma n^2 + \epsilon n^2/2 < \epsilon n^2$. If G' is H-free, then we are done.

Assume for contradiction that G' is not H-free. A copy of H in G' must have its edges going between pairs of parts which are both γ-regular and have density at least ϵ. Hence, there is a mapping from $V(H)$ to the partition of $V(G)$ so that each edge of H maps to a pair of parts which is both γ-regular and have edge density at least ϵ. But the following standard counting lemma (see, e.g., Lemma 3.2 in Alon, Fischer, Krivelevich and Szegedy [6] for a minor variant) shows that the number of labeled copies of H in G' (and hence in G) is at least $2^{-h} \epsilon^m (n/2k)^h > h! \delta n^h$. This contradicts that G has at most δn^h copies of H, completing the proof. □

Lemma 2.4 *If H is a graph with vertices $1, \ldots, h$ and m edges and G is a graph with not necessarily disjoint vertex subsets $W_1, \ldots, W_h$ such that $|W_i| \geq \gamma^{-1}$ for $1 \leq i \leq h$ and, for every edge (i, j) of H, the pair (W_i, W_j) is γ-regular with density $d(W_i, W_j) > \epsilon$ and $\gamma \leq \frac{\epsilon^h}{4h}$, then G contains at least $2^{-h} \epsilon^m |W_1| \times \cdots \times |W_h|$ labeled copies of H with the copy of vertex i in W_i.*

8

The standard proof of this counting lemma uses a greedy embedding strategy. One considers embedding the vertices one at a time, using the regularity condition to maintain the property that at each step where vertex i of H is not yet embedded the set of vertices of G which could potentially be used to embed vertex i is large.

2.2 An improved bound

A partition $P : V = V_1 \cup \ldots \cup V_k$ of the vertex set of a graph $G = (V, E)$ is *weak ϵ-regular* if, for all subsets $S, T \subset V$, we have

$$\left| e(S, T) - \sum_{1 \leq i,j \leq k} |S \cap V_i||T \cap V_j|d(V_i, V_j) \right| \leq \epsilon |V|^2.$$

That is, the density between two sets may be approximated by taking a weighted average over the densities between the sets which they intersect.

The Frieze-Kannan weak regularity lemma [44, 45] states that any graph has such a weak regular partition.

Lemma 2.5 *Let $R(\epsilon) = 2^{c\epsilon^{-2}}$, where c is an absolute constant. For every graph $G = (V, E)$ and every equitable partition P of G into k parts, there is an equitable partition P' which is a refinement of P into at most $kR(\epsilon)$ parts which is weak ϵ-regular.*

Unlike the usual regularity lemma, the bounds in Lemma 2.5 are quite reasonable.[6] It is therefore natural to try to apply it to prove the removal lemma. However, it seems unlikely that this lemma is itself sufficient to prove the removal lemma, since it only gives control over edge densities of a global nature.

However, as noted by Tao [111] (see also [87]), one can prove a stronger theorem by simply iterating the Frieze-Kannan weak regularity lemma.[7] Tao developed this lemma to give an alternative proof of the regularity lemma[8] which extended more easily to hypergraphs [112]. Here we use it to improve the bounds for removal.

Lemma 2.6 *Let $q : [0, 1] \to \mathbb{R}$ be a convex function, G be a graph with $d = d(G)$, $f : \mathbb{N} \to [0, 1]$ be a decreasing function and $r = (dq(1) + (1 - d)q(0) - q(d))/\gamma$. Then there are equitable partitions P and Q with Q a refinement of P satisfying $q(Q) \leq q(P) + \gamma$, Q is weak $f(|P|)$-regular and $|Q| \leq t_r$, where $t_0 = 1$, $t_i = t_{i-1}R(f(t_{i-1}))$ for $1 \leq i \leq r$ and $R(x) = 2^{cx^{-2}}$ as in the Frieze-Kannan weak regularity lemma.*

The proof of Lemma 2.6 is quite similar to the proof of Szemerédi's regularity lemma discussed in the previous subsection. One starts with the trivial partition P_0 of V into one part. We then apply Lemma

---

[6]They are also sharp, that is, there are graphs for which the minimum number of parts in any weak ϵ-regular partition is $2^{\Omega(\epsilon^{-2})}$. This was proved in [24] (see also [5]).

[7]We will say more about this sort of iteration in Section 3.1 below.

[8]More recently, Conlon and Fox [24] showed that it is also closely related to the regular approximation lemma. This lemma, which arose in the study of graph limits by Lovász and Szegedy [74] and also in work on the hypergraph generalization of the regularity lemma by Rödl and Schacht [85], says that by adding and/or deleting a small number of edges in a graph G, we may find another graph G' which admits very fine regular partitions. We refer the reader to [24] and [87] for further details.

2.5 repeatedly to construct a sequence of partitions $P_0, P_1, \ldots$ so that P_{i+1} is weak $f(|P_i|)$-regular. If $q(P_{i+1}) > q(P_i) + \gamma$, then we continue with this process. Otherwise, $q(P_{i+1}) \leq q(P_i) + \gamma$, so we set $Q = P_{i+1}$ and $P = P_i$ and stop the process. This process must stop within r iterations as the third part of Proposition 2.1 shows that the mean-q density lies in an interval of length $r\gamma$.

Rather than using the usual $q(x) = x^2$, we will use the convex function q on $[0, 1]$ defined by $q(0) = 0$ and $q(x) = x \log x$ for $x \in (0, 1]$. This entropy function is central to the proof since it captures the extra structural information coming from Lemma 2.8 below in a concise fashion. Note that the last part of Proposition 2.1 implies that $d \log d \leq q(P) \leq 0$ for every partition P.

The next lemma is a counting lemma that complements the Frieze-Kannan weak regularity lemma. As one might expect, this lemma gives a global count for the number of copies of H, whereas the counting lemma associated with the usual regularity lemma gives a means of counting copies of H between any $v(H)$ parts of the partition which are pairwise regular. Its proof, which we omit, is by a simple telescoping sum argument.

Lemma 2.7 *([18], Theorem 2.7 on page 1809) Let H be a graph on $\{1, \ldots, h\}$ with m edges. Let $G = (V, E)$ be a graph on n vertices and $Q : V = V_1 \cup \ldots \cup V_t$ be a vertex partition which is weak ϵ-regular. The number of homomorphisms from H to G is within $\epsilon m n^h$ of*

$$\sum_{1 \leq i_1, \ldots, i_h \leq t} \prod_{(r,s) \in E(H)} d(V_{i_r}, V_{i_s}) \prod_{a=1}^{h} |V_{i_a}|.$$

Let P and Q be vertex partitions of a graph G with Q a refinement of P. A pair (V_i, V_j) of parts of P is (α, c)-*shattered* by Q if at least a c-fraction of the pairs $(u, v) \in V_i \times V_j$ go between pairs of parts of Q with edge density between them less than α.

One of the key components of the proof is the following lemma, which says that if P and Q are vertex partitions like those given by Lemma 2.6, then there are many pairs of vertex sets in P which are shattered by Q.

Lemma 2.8 *Let H be a graph on $\{1, \ldots, h\}$ with m edges and let $\alpha > 0$. Suppose G is a graph on n vertices for which there are less than δn^h homomorphisms of H into G, where $\delta = \frac{1}{4}\alpha^m(2k)^{-h}$. Suppose P and Q are equitable vertex partitions of G with $|P| = k \leq n$ and Q is a refinement of P which is weak $f(k)$-regular, where $f(k) = \frac{1}{4m}\alpha^m(2k)^{-h}$. For every h-tuple $V_1, \ldots, V_h$ of parts of P, there is an edge (i, j) of H for which the pair (V_i, V_j) is $(\alpha, \frac{1}{2m})$-shattered by Q.*

Proof: As $|P| = k \leq n$, we have $|V_i| \geq \frac{n}{2k}$ for each i. Let Q_i denote the partition of V_i which consists of the parts of Q which are subsets of V_i. Consider an h-tuple $(v_1, \ldots, v_h) \in V_1 \times \cdots \times V_h$ picked uniformly at random. Also consider the event E that, for each edge (i, j) of H, the pair (v_i, v_j) goes between parts of Q_i and Q_j with density at least α. If E occurs with probability at least $1/2$, as Q is weak $f(k)$-regular, Lemma 2.7 implies that the number of homomorphisms of H into G where the copy of vertex i is in V_i for $1 \leq i \leq h$ is at least

$$\frac{1}{2}\alpha^m \prod_{i=1}^{h} |V_i| - mf(k)n^h \geq \left(\frac{1}{2}\alpha^m(2k)^{-h} - mf(k)\right)n^h = \delta n^h,$$

contradicting that there are less than δn^h homomorphisms of H into G. So E occurs with probability less than $1/2$. Hence, for at least $1/2$ of the h-tuples $(v_1, \ldots, v_h) \in V_1 \times \cdots \times V_h$, there is an edge (i, j) of H such that the pair (v_i, v_j) goes between parts of Q_i and Q_j with density less than α. This implies that for at least one edge (i, j) of H, the pair (V_i, V_j) is $(\alpha, \frac{1}{2m})$-shattered by Q. $\quad\square$

We will need the following lemma from [38] which tells us that if a pair of parts from P is shattered by Q then there is an increment in the mean-entropy density. Its proof is by a simple application of Jensen's inequality.

Lemma 2.9 *([38], Lemma 7 on page 570) Let $q : [0,1] \to \mathbb{R}$ be the convex function given by $q(0) = 0$ and $q(x) = x \log x$ for $x > 0$. Let $\epsilon_1, \ldots, \epsilon_r$ and $d_1, \ldots, d_r$ be nonnegative real numbers with $\sum_{i=1}^r \epsilon_i = 1$ and $d = \sum_{i=1}^s \epsilon_i d_i$. Suppose $\beta < 1$ and $I \subset [r]$ is such that $d_i \leq \beta d$ for $i \in I$ and let $s = \sum_{i \in I} \epsilon_i$. Then*

$$\sum_{i=1}^r \epsilon_i q(d_i) \geq q(d) + (1 - \beta + q(\beta))sd.$$

We are now ready to prove Theorem 1.4 in the following precise form.

Theorem 2.1 *Let H be a graph on $\{1, \ldots, h\}$ with m edges. Let $\epsilon > 0$ and δ^{-1} be a tower of twos of height $8h^4 \log \epsilon^{-1}$. If G is a graph on n vertices in which at least ϵn^2 edges need to be removed to make it H-free, then G contains at least δn^h copies of H.*

Proof: Suppose for contradiction that there is a graph G on n vertices in which at least ϵn^2 edges need to be removed from G to delete all copies of H, but G contains fewer than δn^h copies of H. If $n \leq \delta^{-1/h}$, then the number of copies of H in G is less than $\delta n^h \leq 1$, so G is H-free, contradicting that at least ϵn^2 edges need to be removed to make the graph H-free. Hence, $n > \delta^{-1/h}$. Note that the number of mappings from $V(H)$ to $V(G)$ which are not one-to-one is $n^h - h!\binom{n}{h} \leq h^2 n^{h-1} < h^2 \delta^{1/h} n^h$. Let $\delta' = 2h^2 \delta^{1/h}$, so the number of homomorphisms from H to G is at most $\delta' n^h$.

The graph G contains at least $\epsilon n^2/m$ edge-disjoint copies of H. Let G' be the graph on the same vertex set which consists entirely of the at least $\epsilon n^2/m$ edge-disjoint copies of H. Then $d(G') \geq m \cdot \epsilon/m = \epsilon$ and G' consists of $\frac{d(G')}{m} n^2$ edge-disjoint copies of H. We will show that there are at least $\delta' n^h$ homomorphisms from H to G' (and hence to G as well). For the rest of the argument, we will assume the underlying graph is G'.

Let $\alpha = \frac{\epsilon}{8m}$. Apply Lemma 2.6 to G' with $f(k) = \frac{1}{4m}\alpha^m(2k)^{-h}$ and $\gamma = \frac{d(G')}{2h^4}$. Note that r as in Lemma 2.6 is

$$r = d(G') \log(1/d(G'))/\gamma = 2h^4 \log(1/d(G')) \leq 2h^4 \log \epsilon^{-1}.$$

Hence, we get a pair of equitable vertex partitions P and Q, with Q a refinement of P, $q(Q) \leq q(P) + \gamma$, Q is weak $f(|P|)$-regular and $|Q|$ is at most a tower of twos of height $3r \leq 6h^4 \log \epsilon^{-1}$. Let $V_1, \ldots, V_k$ denote the parts of P and Q_i denote the partition of V_i consisting of the parts of Q which are subsets of V_i.

Suppose that (V_a, V_b) is a pair of parts of P with edge density $d = d(V_a, V_b) \geq \epsilon/m$ which is $(\alpha, \frac{1}{2m})$-shattered by Q. Note that $\alpha \leq d/8$. Arbitrarily order the pairs $U_i \times W_i \in Q_a \times Q_b$, letting $d_i = d(U_i, W_i)$ and $\epsilon_i = \frac{|U_i||W_i|}{|V_a||V_b|}$, so that the conditions of Lemma 2.9 with $\beta = 1/8$ are satisfied. Applying Lemma

2.9, we get, since $q(\beta) = -\frac{1}{8}\log 8 = -\frac{3}{8}$, that

$$q(Q_a, Q_b) - q(V_a, V_b) \geq (1 - \beta + q(\beta))\frac{1}{2m}d(V_a, V_b)|V_a||V_b|/n^2 \geq \frac{1}{4m}e(V_a, V_b)/n^2.$$

Note that

$$q(Q) - q(P) = \sum_{1 \leq a,b \leq k} (q(Q_a, Q_b) - q(V_a, V_b)),$$

which shows that $q(Q) - q(P)$ is the sum of nonnegative summands.

There are at most $\frac{\epsilon}{m}n^2/2$ edges of G' going between pairs of parts of P with density at most $\frac{\epsilon}{m}$. Hence, at least $1/2$ of the edge-disjoint copies of H making up G' have all its edges going between pairs of parts of P of density at least $\frac{\epsilon}{m}$. By Lemma 2.8, for each copy of H, at least one of its edges goes between a pair of parts of P which is $(\alpha, \frac{1}{2m})$-shattered by Q. Thus,

$$q(Q) - q(P) \geq \sum \frac{1}{4m}e(V_a, V_b)/n^2 \geq \frac{1}{4m} \cdot \frac{d(G')}{2m} = \frac{d(G')}{8m^2} > \gamma,$$

where the sum is over all ordered pairs (V_a, V_b) of parts of P which are $(\alpha, \frac{1}{2m})$-shattered by Q and with $d(V_a, V_b) \geq \frac{\epsilon}{m}$. This contradicts $q(Q) \leq q(P) + \gamma$ and completes the proof. $\qquad\square$

3 The induced removal lemma

As in the last section, we will again discuss two different proofs of the induced removal lemma, Theorem 1.3. In Section 3.1, we will discuss the proof of Alon, Fischer, Krivelevich and Szegedy [6], which uses their strong regularity lemma and gives a wowzer-type bound. In Section 3.2, we will examine the authors' recent proof [24] of a tower-type bound. We will discuss Alon and Shapira's generalization of the induced removal lemma, which applies to infinite families of graphs, in Section 3.3.

3.1 The usual proof

For an equitable partition $P = \{V_i | 1 \leq i \leq k\}$ of $V(G)$ and an equitable refinement $Q = \{V_{i,j} | 1 \leq i \leq k, 1 \leq j \leq \ell\}$ of P, we say that Q is ϵ-*close* to P if the following is satisfied. All $1 \leq i \leq i' \leq k$ but at most ϵk^2 of them are such that, for all $1 \leq j, j' \leq \ell$ but at most $\epsilon\ell^2$ of them, $|d(V_i, V_{i'}) - d(V_{i,j}, V_{i',j'})| < \epsilon$ holds. This notion roughly says that Q is an approximation of P. The strong regularity lemma of Alon, Fischer, Krivelevich and Szegedy [6] is now as follows.

Lemma 3.1 (Strong regularity lemma) *For every function $f : \mathbb{N} \to (0,1)$ there exists a number $S = S(f)$ with the following property. For every graph $G = (V, E)$, there is an equitable partition P of the vertex set V and an equitable refinement Q of P with $|Q| \leq S$ such that the partition P is $f(1)$-regular, the partition Q is $f(|P|)$-regular and Q is $f(1)$-close to P.*

That is, there is a regular partition P and a refinement Q such that Q is very regular and yet the densities between parts $(V_{i,j}, V_{i',j'})$ of Q are usually close to the densities between the parts $(V_i, V_{i'})$ of P containing them.

Let $f(1) = \epsilon$. Here, and throughout this section, we let $q(x) = x^2$ be the square function as in the proof of Szemerédi's regularity lemma in the previous section. The condition that Q be ϵ-close to P

is equivalent, up to a polynomial change in ϵ, to $q(Q) \leq q(P) + \epsilon$. Indeed, if Q is ϵ-close to P, then $q(Q) \leq q(P) + O(\epsilon)$, while if $q(Q) \leq q(P) + \epsilon$, then Q is $O(\epsilon^{1/4})$-close to P. A version of this statement is present in Lemma 3.7 of [6]. As it is sufficient and more convenient to work with mean-square density instead of ϵ-closeness, we do so from now on. That is, we replace the third condition in the regularity lemma with the condition that $q(Q) \leq q(P) + \epsilon$.

With this observation, the proof of the strong removal lemma becomes quite straightforward. Note that we may assume that f is a decreasing function by replacing it, if necessary, with the function given by $f'(i) = \min_{1 \leq j \leq i} f(j)$. We consider a series of partitions $P_1, P_2, \ldots$, where P_1 is an $f(1)$-regular partition and P_{i+1} is an $f(|P_i|)$-regular refinement of the partition P_i. Since P_{i+1} is a refinement of P_i we know that the mean-square density must have increased, that is, $q(P_{i+1}) \geq q(P_i)$. If also $q(P_{i+1}) \leq q(P_i) + \epsilon$ then, since f is decreasing and P_{i+1} is $f(|P_i|)$-regular, we see that all three conditions of the theorem are satisfied with $P = P_i$ and $Q = P_{i+1}$ as the required partitions. Otherwise, we have $q(P_{i+1}) > q(P_i) + \epsilon$. However, since the mean-square density is bounded above by 1, this can happen at most ϵ^{-1} times, concluding the proof.

It is not hard to see why this proof results in wowzer-type bounds. At each step, we are applying the regularity lemma to find a partition P_{i+1} which is regular in the number of parts in the previous partition P_i. The bounds coming from the regularity lemma then imply that $|P_{i+1}| = T(f(|P_i|)^{-O(1)})$. But this iterated tower-type bound is essentially how we define the wowzer function.

That this is the correct behaviour for the bounds in the strong regularity lemma was proved independently by Conlon and Fox [24] and by Kalyanasundaram and Shapira [61], though both proofs use slightly different ideas and result in slightly different bounds. With the function $f : \mathbb{N} \to (0,1)$ taken to be $f(n) = \epsilon/n$, the proof given in [24] shows that the number of parts in the smaller partition P may need to be as large as wowzer in a power of ϵ^{-1}, while that given in [61] proves that it must be at least wowzer in $\sqrt{\log \epsilon^{-1}}$.

The following easy corollary of the strong regularity lemma [6] is the key to proving the induced graph removal lemma.

Lemma 3.2 *For each $0 < \epsilon < 1/3$ and decreasing function $f : \mathbb{N} \to (0, 1/3)$, there is $\delta' = \delta'(\epsilon, f)$ such that every graph $G = (V, E)$ with $|V| \geq \delta'^{-1}$ has an equitable partition $V = V_1 \cup \ldots \cup V_k$ and vertex subsets $W_i \subset V_i$ such that $|W_i| \geq \delta'|V|$, each pair (W_i, W_j) with $1 \leq i \leq j \leq k$ is $f(k)$-regular and all but at most ϵk^2 pairs $1 \leq i \leq j \leq k$ satisfy $|d(V_i, V_j) - d(W_i, W_j)| \leq \epsilon$.*

In fact, Lemma 3.2 is a little bit stronger than the original version in [6] in that each set W_i is $f(k)$-regular with itself.[9] The original version follows from the strong regularity lemma, applied with

$$f'(k) = \min\left(f(k), \frac{\epsilon}{4}, \frac{1}{2}\binom{k+2}{2}^{-1} \right),$$

by taking the partition $V = V_1 \cup \ldots \cup V_k$ to be the partition P in the strong regularity lemma and the subset W_i to be a random part $V_{i,p} \subset V_i$ of the refinement Q of P in the strong regularity lemma. Since $f'(k) \leq \frac{1}{2}\binom{k+2}{2}^{-1}$, it is straightforward to check that all pairs (W_i, W_j) are $f(k)$-regular with probability greater than $\frac{1}{2}$. Moreover, the expected number of pairs with $1 \leq i < j \leq k$ for which

---

[9]This stronger version may be derived from an extra application of the regularity lemma within each of the pieces W_i, together with a suitable application of Ramsey's theorem. This is essentially the process carried out in [6], though they do not state their final result in the same form as Lemma 3.2.

$|d(V_i, V_j) - d(W_i, W_j)| > \epsilon$ is at most

$$\frac{\epsilon}{4}\binom{k}{2} + \frac{\epsilon}{4}\binom{k}{2} = \frac{\epsilon}{2}\binom{k}{2}.$$

Here, the two $\frac{\epsilon}{4}$ factors come from the definition of $f(1)$-closeness. The first factor comes from the fact that at most an $\frac{\epsilon}{4}$-fraction of the pairs (V_i, V_j) do not have good approximations while the second factor comes from the fact that for all other pairs there are at most an $\frac{\epsilon}{4}$ fraction of pairs (W_i, W_j) which do not satisfy $|d(V_i, V_j) - d(W_i, W_j)| \leq \epsilon$. Therefore, by Markov's inequality, the probability that the number of bad pairs is greater than $\epsilon\binom{k}{2}$ is less than $\frac{1}{2}$. We therefore see that with positive probability there is a choice of W_i satisfying the required weaker version of Lemma 3.2.

If we assume the full strength of Lemma 3.2 as stated, that is, that each W_i is also $f(k)$-regular with itself, it is easy to deduce the induced removal lemma. Let $h = |V(H)|$ and take $f(k) = \frac{\epsilon^h}{4h}$. If there is a mapping $\phi : V(H) \to \{1, \ldots, k\}$ such that for all adjacent vertices v, w of H, the edge density between $W_{\phi(v)}$ and $W_{\phi(w)}$ is at least ϵ and for all distinct nonadjacent vertices v, w of H, the edge density between $W_{\phi(v)}$ and $W_{\phi(w)}$ is at most $1 - \epsilon$, then the following standard counting lemma (see, e.g., Lemma 3.2 in Alon, Fischer, Krivelevich and Szegedy [6] for a minor variant) shows that G contains at least δn^h induced copies of H, where $\delta = \frac{1}{h!}(\epsilon/4)^{\binom{h}{2}}\delta'^h$. As with Lemma 2.4, the standard proof of this counting lemma uses a greedy embedding strategy.

Lemma 3.3 *If H is a graph with vertices $1, \ldots, h$ and G is a graph with not necessarily disjoint vertex subsets $W_1, \ldots, W_h$ such that every pair (W_i, W_j) with $1 \leq i < j \leq h$ is γ-regular with $\gamma \leq \frac{\eta^h}{4h}$, $|W_i| \geq \gamma^{-1}$ for $1 \leq i \leq h$ and, for $1 \leq i < j \leq k$, $d(W_i, W_j) > \eta$ if (i, j) is an edge of H and $d(W_i, W_j) < 1 - \eta$ otherwise, then G contains at least $\left(\frac{\eta}{4}\right)^{\binom{h}{2}} |W_1| \times \cdots \times |W_h|$ induced copies of H with the copy of vertex i in W_i.*

Hence, we may assume that there is no such mapping ϕ. We then delete the edges between V_i and V_j if the edge density between W_i and W_j is less than ϵ and add the edges between V_i and V_j if the density between W_i and W_j is more than $1 - \epsilon$. The total number of edges added or removed is at most $5\epsilon n^2$ and no induced copy of H remains. Replacing ϵ by $\epsilon/8$ in the above argument gives the induced removal lemma.

3.2 An improved bound

The main goal of this section is to prove Theorem 1.5, which gives a bound on δ^{-1} which is a tower in h of height polynomial in ϵ^{-1}. We in fact prove the key corollary of the strong regularity lemma, Lemma 3.2, with a tower-type bound. This is sufficient to prove the desired tower-type bound for the induced graph removal lemma.

As in Section 2.2, the key idea will be to take a weak variant of Szemerédi's regularity lemma and iterate it. The particular variant we will use, due to Duke, Lefmann and Rödl [28], was originally used by them to derive a fast approximation algorithm for the number of copies of a fixed graph in a large graph.

A *k-cylinder* (or cylinder for short) in a graph G is a product of k vertex subsets. Given a k-partite graph $G = (V, E)$ with k-partition $V = V_1 \cup \ldots \cup V_k$, we will consider a partition $\mathcal{K}$ of the cylinder $V_1 \times \cdots \times V_k$ into cylinders $K = W_1 \times \cdots \times W_k$, $W_i \subset V_i$ for $i = 1, \ldots, k$ and we let $V_i(K) = W_i$.

We say that a cylinder is ϵ-*regular* if all $\binom{k}{2}$ pairs of subsets (W_i, W_j), $1 \leq i < j \leq k$, are ϵ-regular. The partition $\mathcal{K}$ is ϵ-*regular* if all but an ϵ-fraction of the k-tuples $(v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$ are in ϵ-regular cylinders in the partition $\mathcal{K}$.

The weak regularity lemma of Duke, Lefmann and Rödl [28] is now as follows. Note that, like the Frieze-Kannan weak regularity lemma, it has only a single-exponential bound on the number of parts. We will sometimes refer to this lemma as the cylinder regularity lemma.

Lemma 3.4 *Let $0 < \epsilon < 1/2$ and $\beta = \beta(\epsilon) = \epsilon^{k^2 \epsilon^{-5}}$. Suppose $G = (V, E)$ is a k-partite graph with k-partition $V = V_1 \cup \ldots \cup V_k$. Then there exists an ϵ-regular partition $\mathcal{K}$ of $V_1 \times \cdots \times V_k$ into at most β^{-1} parts such that, for each $K \in \mathcal{K}$ and $1 \leq i \leq k$, $|V_i(K)| \geq \beta|V_i|$.*

We would now like to iterate this lemma to get a stronger version, the strong cylinder regularity lemma. Like Lemmas 2.6 and 3.1, this will yield two closely related cylinder partitions P and Q with P regular and Q regular in a function of $|P|$. To state the lemma, we first strengthen the definition of regular cylinders so that pieces are also regular with themselves.

A k-cylinder $W_1 \times \cdots \times W_k$ is *strongly ϵ-regular* if all pairs (W_i, W_j) with $1 \leq i, j \leq k$ are ϵ-regular. A partition $\mathcal{K}$ of $V_1 \times \cdots \times V_k$ into cylinders is *strongly ϵ-regular* if all but $\epsilon|V_1| \times \cdots \times |V_k|$ of the k-tuples $(v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$ are contained in strongly ϵ-regular cylinders $K \in \mathcal{K}$.

We now state the strong cylinder regularity lemma. Here $t_i(x)$ is a variant of the tower function defined by $t_0(x) = x$ and $t_{i+1}(x) = 2^{t_i(x)}$. Also, given a cylinder partition $\mathcal{K}$, $Q(\mathcal{K})$ is the coarsest vertex partition such that every set $V_i(K)$ with $i \in [k]$ and $K \in \mathcal{K}$ is the union of parts of $Q(\mathcal{K})$.

Lemma 3.5 *For $0 < \epsilon < 1/3$, positive integer s, and decreasing function $f : \mathbb{N} \rightarrow (0, \epsilon]$, there is $S = S(\epsilon, s, f)$ such that the following holds. For every graph G, there is an integer $s \leq k \leq S$, an equitable partition $P : V = V_1 \cup \ldots \cup V_k$ and a strongly $f(k)$-regular partition $\mathcal{K}$ of the cylinder $V_1 \times \cdots \times V_k$ into cylinders satisfying that the partition $Q = Q(\mathcal{K})$ of V has at most S parts and $q(Q) \leq q(P) + \epsilon$. Furthermore, there is an absolute constant c such that letting $s_1 = s$ and $s_{i+1} = t_4\left(\left(s_i/f(s_i)\right)^c\right)$, we may take $S = s_\ell$ with $\ell = 2\epsilon^{-1} + 1$.*

To prove this lemma, we need to find a way to guarantee that the parts of the cylinder partition are regular with themselves as required in the definition of strong cylinder regularity. For a graph $G = (V, E)$, a vertex subset $U \subset V$ is ϵ-*regular* if the pair (U, U) is ϵ-regular. The following lemma, which demonstrates that any graph contains a large vertex subset which is ϵ-regular, is the first step.

Lemma 3.6 *For each $0 < \epsilon < 1/2$, let $\delta = \delta(\epsilon) = 2^{-\epsilon^{-(10/\epsilon)^4}}$. Every graph $G = (V, E)$ contains an ϵ-regular vertex subset U with $|U| \geq \delta|V|$.*

One way to prove this lemma is to first find a large collection C of disjoint subsets of equal order which are pairwise α-regular with $\alpha = (\epsilon/3)^2$. This can be done by an application of Szemerédi's regularity lemma and Turán's theorem, but then the bounds are quite weak. Instead, one can easily deduce this from Lemma 3.4. A further application of Ramsey's theorem allows one to get a subcollection C' of size $s \geq 2\alpha^{-1}$ such that the edge density between each pair of distinct subsets in C' lies in an interval of length at most α. The union of the sets in C' is then an ϵ-regular subset of the desired order.

It is crucial in this lemma that δ^{-1} be of bounded tower height in ϵ^{-1}. While our bound gives a double exponential dependence, we suspect that the truth is more likely to be a single exponential. We leave this as an open problem.

15

Repeated applications of Lemma 3.6 allow us to pull out large, regular subsets until a small fraction of vertices remain. By distributing the remaining vertices amongst these subsets, we only slightly weaken their regularity, while giving a partition of any graph into large parts each of which is ϵ-regular with itself. This will be sufficient for our purposes.

Lemma 3.7 *For each $0 < \epsilon < 1/2$, let $\delta = \delta(\epsilon) = 2^{-\epsilon^{-(20/\epsilon)^4}}$. Every graph $G = (V, E)$ has a vertex partition $V = V_1 \cup \ldots \cup V_k$ such that for each i, $1 \leq i \leq k$, $|V_i| \geq \delta |V|$ and V_i is an ϵ-regular set.*

We are now ready to prove the strong cylinder regularity lemma.

Proof of Lemma 3.5: We may assume $|V| \geq S$, as otherwise we can let P and Q be the trivial partitions into singletons, and it is easy to see the lemma holds. We will define a sequence of partitions $P_1, P_2, \ldots$ of equitable partitions, with P_{j+1} a refinement of P_j and $q(P_{j+1}) > q(P_j) + \epsilon/2$. Let P_1 be an arbitrary equitable partition of V consisting of $s_1 = s$ parts. Suppose we have already found an equitable partition $P_j : V = V_1 \cup \ldots \cup V_k$ with $k \leq s_j$.

Let $\beta(x, \ell) = x^{\ell^2 x^{-5}}$ as in Lemma 3.4 and $\delta(x) = 2^{-x^{-(20/x)^4}}$ as in Lemma 3.7. We apply Lemma 3.7 to each part V_i of the partition P_j to get a partition of each part $V_i = V_{i1} \cup \ldots \cup V_{ih_i}$ of P_i into parts each of cardinality at least $\delta |V_i|$, where $\delta = \delta(\gamma)$ and $\gamma = f(k) \cdot \beta$ with $\beta = \beta(f(k), k)$, such that each part V_{ih} is γ-regular. Note that δ^{-1} is at most triple-exponential in a polynomial in $k/f(k)$. For each k-tuple $\ell = (\ell_1, \ldots, \ell_k) \in [h_1] \times \cdots \times [h_k]$, by Lemma 3.4 there is an $f(k)$-regular partition $\mathcal{K}_\ell$ of the cylinder $V_{1\ell_1} \times \cdots \times V_{k\ell_k}$ into at most β^{-1} cylinders such that, for each $K \in \mathcal{K}_\ell$, $|V_{i\ell_i}(K)| \geq \beta |V_{i\ell_i}|$. The union of the $\mathcal{K}_\ell$ forms a partition $\mathcal{K}$ of $V_1 \times \cdots \times V_k$ which is strongly $f(k)$-regular.

Recall that $Q = Q(\mathcal{K})$ is the partition of V which is the common refinement of all parts $V_i(K)$ with $i \in [k]$ and $K \in \mathcal{K}$. The number of parts of $\mathcal{K}$ is at most $\delta^{-k}\beta^{-1}$ and hence the number of parts of Q is at most $k2^{1/(\delta^k \beta)}$. Thus, the number of parts of Q is at most quadruple-exponential in a polynomial in $k/f(k)$. Let P_{j+1} be an equitable partition into $4\epsilon^{-1}|Q|$ parts with $q(P_{j+1}) \geq q(Q) - \frac{\epsilon}{2}$, which exists by Lemma 2.3. Hence, there is an absolute constant c such that

$$|P_{j+1}| \leq t_4 \left((k/f(k))^c \right) \leq s_{j+1}.$$

If $q(Q) \leq q(P_j) + \epsilon$, then we may take $P = P_j$ and $Q = Q(\mathcal{K})$, and these partitions satisfy the desired properties. Otherwise, $q(P_{j+1}) \geq q(Q) - \frac{\epsilon}{2} > q(P_j) + \frac{\epsilon}{2}$, and we continue the sequence of partitions. Since $q(P_1) \geq 0$ and the mean-square density goes up by more than $\epsilon/2$ at each step and is always at most 1, this process must stop within $2/\epsilon$ steps, and we obtain the desired partitions. $\square$

Let $G = (V, E)$, $P : V = V_1 \cup \ldots \cup V_k$ be an equipartition and $\mathcal{K}$ be a partition of the cylinder $V_1 \times \cdots \times V_k$ into cylinders. For $K = W_1 \times \cdots \times W_k \in \mathcal{K}$, define the density $d(K) = \frac{|W_1| \times \cdots \times |W_k|}{|V_1| \times \cdots \times |V_k|}$. The cylinder K is ϵ-*close* to P if $|d(W_i, W_j) - d(V_i, V_j)| \leq \epsilon$ for all but at most ϵk^2 pairs $1 \leq i \neq j \leq k$. The cylinder partition $\mathcal{K}$ is ϵ-*close* to P if $\sum d(K) \leq \epsilon$, where the sum is over all $K \in \mathcal{K}$ that are not ϵ-close to P. As with the definition of closeness used in the strong regularity lemma, this definition is closely related to the condition that $q(Q) \leq q(P) + \epsilon$, where here $Q = Q(\mathcal{K})$.

The connection we shall need to prove Lemma 3.2 is contained in the following statement.

Lemma 3.8 *Let $G = (V, E)$ and $P : V = V_1 \cup \ldots \cup V_k$ be an equipartition with $k \geq 2\epsilon^{-1}$ and $|V| \geq 4k\epsilon^{-1}$. Let $\mathcal{K}$ be a partition of the cylinder $V_1 \times \cdots \times V_k$ into cylinders. If $Q = Q(\mathcal{K})$ satisfies $q(Q) \leq q(P) + \epsilon$, then $\mathcal{K}$ is $(2\epsilon)^{1/4}$-close to P.*

Proof: It will be helpful to assume that all parts of the equipartition P have equal size - this affects the calculations only slightly. It will also be helpful to introduce a slight variant of the mean-square density as follows. Let $q'(P) = \sum_{i<j} d^2(V_i, V_j) p_{ij}$, where $p_{ij} = |V_i||V_j|/\sum_{i<j}|V_i||V_j|$. Thus, $q'(P)$ is the mean of the square densities between the pairs of distinct parts. It is easy to check that $q'(P)$ is close to $q(P)$. Indeed, we have $q'(P) - q(P) = \frac{1}{k}(q'(P) - \bar{q})$, where $\bar{q} = \sum_{i=1}^{k} d^2(V_i)/k$ is the average of the square densities inside the parts. Hence, $|q'(P) - q(P)| \leq \frac{1}{k}$. We similarly have $|q'(Q) - q(Q)| \leq \frac{1}{k}$. Let

$$q(\mathcal{K}) = \binom{k}{2}^{-1} \sum_{i<j} \sum_{K \in \mathcal{K}} d^2(V_i(K), V_j(K)) d(K).$$

We have the following equalities

$$
\begin{aligned}
q(\mathcal{K}) - q'(P) &= \binom{k}{2}^{-1} \sum_{i<j} \sum_{K \in \mathcal{K}} \left(d^2(V_i(K), V_j(K)) - d^2(V_i, V_j)\right) d(K) \\
&= \binom{k}{2}^{-1} \sum_{i<j} \sum_{K \in \mathcal{K}} \left(d(V_i(K), V_j(K)) - d(V_i, V_j)\right)^2 d(K),
\end{aligned}
$$

where the last equality uses the identity $d(V_i, V_j) = \sum_{K \in \mathcal{K}} d(V_i(K), V_j(K)) d(K)$. This equality shows that $q(\mathcal{K}) \geq q'(P)$ as it expresses their difference as a sum of nonnegative terms. Furthermore, it shows that if $\mathcal{K}$ is not β-close to P, then $q(\mathcal{K}) \geq q(P) + \binom{k}{2}^{-1} \cdot \frac{\beta k^2}{2} \cdot \beta^2 \cdot \beta \geq q(P) + \beta^4$. In particular, if $q(\mathcal{K}) \leq q'(P) + 2\epsilon$, then $\mathcal{K}$ is $(2\epsilon)^{1/4}$-close to P. So assume for contradiction that $q(\mathcal{K}) > q'(P) + 2\epsilon$. A similar equality implies $q'(Q) \geq q(\mathcal{K})$. We therefore have

$$
\begin{aligned}
q(Q) - q(P) &= \left(q(Q) - q'(Q)\right) + \left(q'(Q) - q(\mathcal{K})\right) + \left(q(\mathcal{K}) - q'(P)\right) + \left(q'(P) - q(P)\right) \\
&\geq -\frac{1}{k} + 0 + \left(q(\mathcal{K}) - q'(P)\right) - \frac{1}{k} \\
&> \epsilon,
\end{aligned}
$$

contradicting the assumption of Lemma 3.8 and completing the proof. $\qquad \square$

With this in hand, we can readily deduce a tower-type bound for Lemma 3.2.

Lemma 3.9 *For each $0 < \epsilon < 1/3$ and decreasing function $f : \mathbb{N} \to (0, \epsilon]$, there is $\delta' = \delta'(\epsilon, f)$ such that every graph $G = (V, E)$ with $|V| \geq \delta'^{-1}$ has an equitable partition $V = V_1 \cup \ldots \cup V_k$ and vertex subsets $W_i \subset V_i$ such that $|W_i| \geq \delta'|V|$, each pair (W_i, W_j) with $1 \leq i \leq j \leq k$ is $f(k)$-regular and all but at most ϵk^2 pairs $1 \leq i \leq j \leq k$ satisfy $|d(V_i, V_j) - d(W_i, W_j)| \leq \epsilon$. Furthermore, we may take $\delta' = \frac{1}{8S^2}$, where $S = S(\frac{\epsilon^4}{2}, s, f)$ is defined as in Lemma 3.5 and $s = 2\epsilon^{-1}$.*

Proof: Let $\alpha = \frac{\epsilon^4}{2}$, $s = 2\epsilon^{-1}$, and $\delta' = \frac{1}{8S^2}$, where $S = S(\alpha, s, f)$ is as in Lemma 3.5. We apply Lemma 3.5 with α in place of ϵ. We get an equipartition $P : V = V_1 \cup \ldots \cup V_k$ with $s \leq k \leq S$ and a strongly $f(k)$-regular partition $\mathcal{K}$ of $V_1 \times \cdots \times V_k$ into cylinders such that the refinement $Q = Q(\mathcal{K})$ of P has at most $S = S(\alpha, s, f)$ parts and satisfies $q(Q) \leq q(P) + \alpha$. Since $|V| \geq \delta'^{-1} = 8S^2$, and P is an equipartition into $k \leq S$ parts, the cardinality of each part $V_i \in P$ satisfies $|V_i| \geq \frac{|V|}{2S}$. By Lemma 3.8, as $(2\alpha)^{1/4} = \epsilon$, the cylinder partition $\mathcal{K}$ is ϵ-close to P. Hence, at most an ϵ-fraction of the k-tuples $(v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$ belong to parts $K = W_1 \times \cdots \times W_k$ of $\mathcal{K}$ that are not ϵ-close to P. Since $Q(\mathcal{K})$ has at most S parts, the fraction of k-tuples $(v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$ that belong

17

to parts $K = W_1 \times \cdots \times W_k$ of $\mathcal{K}$ with $|W_i| < \frac{1}{4S}|V_i|$ for at least one $i \in [k]$ is at most $\frac{1}{4S} \cdot S = \frac{1}{4}$. Therefore, at least a fraction $1 - f(k) - \epsilon - \frac{1}{4} > 0$ of the k-tuples $(v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$ belong to parts $K = W_1 \times \cdots \times W_k$ of $\mathcal{K}$ satisfying K is strongly $f(k)$-regular, $|W_i| \geq \frac{1}{4S}|V_i| \geq \delta'|V|$ for $i \in [k]$ and K is ϵ-close to P. Since a positive fraction of the k-tuples belong to such K, there is at least one such K. This K has the desired properties. Indeed, the number of pairs $1 \leq i \neq j \leq k$ for which $|d(W_i, W_j) - d(V_i, V_j)| > \epsilon$ is at most ϵk^2 and hence the number of pairs $1 \leq i \leq j \leq k$ for which $|d(W_i, W_j) - d(V_i, V_j)| > \epsilon$ is at most $\epsilon k^2/2 + k \leq \epsilon k^2$. This completes the proof. $\qquad\square$

By using the induced counting lemma, Lemma 3.3, we may now conclude the proof as in Section 3.1 to obtain the following quantitative version of Theorem 1.3.

Theorem 3.1 *There exists a constant c such that, for any graph H on h vertices and $0 < \epsilon < 1/2$, if $\delta^{-1} = t_j(h)$, where $j = c\epsilon^{-4}$, then any graph G on n vertices with at most δn^h induced copies of H may be made induced H-free by adding and/or deleting at most ϵn^2 edges.*

3.3 Infinite removal lemma

In order to characterize the natural graph properties which are testable, the induced removal lemma was extended by Alon and Shapira [12] to the following infinite version. For a family $\mathcal{H}$ of graphs, a graph G is induced $\mathcal{H}$-free if G does not contain any graph H in $\mathcal{H}$.

Theorem 3.2 *For every (possibly infinite) family of graphs $\mathcal{H}$ and $\epsilon > 0$, there are n_0, h_0, and δ such that the following holds. If a graph $G = (V, E)$ on $n \geq n_0$ vertices has at most δn^h induced copies of each graph $H \in \mathcal{H}$ on $h \leq h_0$ vertices, then G can be made induced $\mathcal{H}$-free by adding and/or deleting at most ϵn^2 edges.*

Proof: The proof is a natural extension of the proof of the induced removal lemma and similarly uses the key corollary, Lemma 3.2, of the strong regularity lemma. The main new idea is to pick an appropriate function f to apply Lemma 3.2. The choice of the function f will depend heavily on the family $\mathcal{H}$.

For a graph H and an edge-coloring c of the edges of the complete graph with loops R on $[k]$ with colors white, black and grey, we write $H \rightarrow_c R$ if there is a mapping $\phi : V(H) \rightarrow [k]$ such that for each edge (u, v) of H we have that $c(\phi(u), \phi(v))$ is black or grey and for each pair (u, v) of distinct vertices of H which do not form an edge we have that $c(\phi(u), \phi(v))$ is white or grey. We write $H \not\rightarrow_c R$ if $H \rightarrow_c R$ does not hold.

Let $P : V = V_1 \cup \ldots \cup V_k$ be a vertex partition of G. A key observation is that if we *round* G by the partition P and the coloring c to obtain a graph G' on the same vertex set as G by adding edges to make (V_i, V_j) complete if (i, j) is black, deleting edges to make (V_i, V_j) empty if (i, j) is white and we have that $H \not\rightarrow_c R$, then G' does not contain H as an induced subgraph.

For any (possibly infinite) family of graphs $\mathcal{H}$ and any integer r, let $\mathcal{H}_r$ be the following set of colored complete graphs with loops: a colored complete graph with loops R belongs to $\mathcal{H}_r$ if and only if it has at most r vertices and there is at least one $H \in \mathcal{H}$ such that $H \rightarrow_c R$. For any family $\mathcal{H}$ of graphs and integer r for which $\mathcal{H}_r \neq \emptyset$, let

$$\Psi_{\mathcal{H}}(r) = \max_{R \in \mathcal{H}_r} \min_{H \in \mathcal{H}:H \rightarrow_c R} |V(H)|.$$

18

If $\mathcal{H}_r = \emptyset$, define $\Psi_{\mathcal{H}}(r) = 1$. Note that $\Psi_{\mathcal{H}}(r)$ is a monotonically increasing function of r. Let

$$f(r) = \frac{\epsilon^{\Psi_{\mathcal{H}}(r)}}{4\Psi_{\mathcal{H}}(r)}.$$

Note that the function f only depends on ϵ and $\mathcal{H}$.

Let $\delta' = \delta'(\epsilon, f)$ be as in Lemma 3.2, which only depends on ϵ and $\mathcal{H}$. Also let $k_0 = 2\delta'^{-1}$, $h_0 = \Psi_{\mathcal{H}}(k_0)$, $n_0 = 1/(\delta' f(k_0))$ and $\delta = \frac{1}{h_0!}(\epsilon/4)^{h_0^2}\delta'^{h_0}$. We have that k_0, h_0, n_0 and $\delta > 0$ only depend on ϵ and $\mathcal{H}$. By assumption, G has $n \geq n_0$ vertices.

We apply Lemma 3.2 to G. We get an equitable vertex partition $P : V = V_1 \cup \ldots \cup V_k$ of G and subsets $W_i \subset V_i$ with $|W_i| \geq \delta'|V|$ such that, for $1 \leq i \leq j \leq k$, the pair (W_i, W_j) is $f(k)$-regular and all but at most ϵk^2 pairs $1 \leq i \leq j \leq k$ satisfy $|d(V_i, V_j) - d(W_i, W_j)| \leq \epsilon$. As $\delta'|V| \leq |W_i| \leq |V_i| \leq 2n/k$, we have $k \leq 2\delta'^{-1} \leq k_0$.

Consider the coloring c of the complete graph with loops R on $[k]$ where a pair (i, j) of vertices is black if $d(W_i, W_j) \geq 1 - \epsilon$, white if $d(W_i, W_j) \leq \epsilon$ and grey if $\epsilon < d(W_i, W_j) < 1 - \epsilon$. Suppose, for the sake of contradiction, that there is a graph H with $H \to_c R$. From the definition of Ψ, there is a graph H on $h \leq \Psi_{\mathcal{H}}(k)$ vertices with $H \to_c R$. As $k \leq k_0$, the number of vertices of H satisfies $h \leq h_0$. As each pair (W_i, W_j) is $f(k)$-regular and $|W_i| \geq \delta'|V| \geq f(k)^{-1}$, applying the induced counting lemma, Lemma 3.3, with $\gamma = f(k)$, we get at least

$$\frac{1}{h!}\left(\frac{\epsilon}{4}\right)^{\binom{h}{2}}(\delta'|V|)^h \geq \delta n^h$$

induced copies of H in G, contradicting the supposition of the theorem. Thus, there is no graph H with $H \to_c R$.

We round the graph G by the partition P and the coloring c as described earlier in the proof to obtain a graph G'. By the key observation, for each graph H with $H \not\to_c R$, the graph G' does not contain H as an induced subgraph. Hence, G' is induced $\mathcal{H}$-free.

Moreover, not many edges were changed from G to obtain G'. Indeed, as there are at most ϵk^2 pairs $1 \leq i \leq j \leq k$ which satisfy $|d(V_i, V_j) - d(W_i, W_j)| > \epsilon$, the number of edge modifications made between such pairs is at most $\epsilon k^2 \cdot (2n/k)^2 = 4\epsilon n^2$. Between the other pairs we have made at most $2\epsilon\binom{n}{2} \leq \epsilon n^2$ edge modifications. In total, at most $5\epsilon n^2$ edge modifications were made to obtain G' from G. Replacing ϵ by $\epsilon/5$ in the above argument completes the proof. $\square$

4 Arithmetic removal

The notion of arithmetic removal was introduced by Green [57]. By establishing an appropriate variant of the regularity lemma in the context of abelian groups, he proved the following result.

Theorem 4.1 *For any natural number $k \geq 3$ and any $\epsilon > 0$, there exists $\delta > 0$ such that if G is an abelian group of order n and $A_1, \ldots, A_k$ are subsets of G such that there are at most δn^{k-1} solutions to the equation $a_1 + a_2 + \cdots + a_k = 0$ with $a_i \in A_i$ for all i then it is possible to remove at most ϵn elements from each set A_i to form sets A'_i so that there are no solutions to the equation $a'_1 + a'_2 + \cdots + a'_k = 0$ with $a'_i \in A'_i$ for all i.*

It is an exercise to show that Green's result implies Roth's theorem. While Green's proof of this result relied on Fourier analytic techniques, an alternative proof was found by Král', Serra, and Vena [70], who showed that the following more general result follows from an elegant reduction to the removal lemma in directed graphs.

Theorem 4.2 *For any natural number $k \geq 3$ and any $\epsilon > 0$, there exists $\delta > 0$ such that if G is a group of order n, $g \in G$ and $A_1, \ldots, A_k$ are subsets of G such that there are at most δn^{k-1} solutions to the equation $a_1 a_2 \cdots a_k = g$ with $a_i \in A_i$ for all i then it is possible to remove at most ϵn elements from each set A_i to form sets A_i' so that there are no solutions to the equation $a_1' a_2' \cdots a_k' = g$ with $a_i' \in A_i'$ for all i.*

This is stronger than Theorem 4.1 in two ways. Firstly, it applies to all groups and not just to abelian groups. Secondly, it applies to non-homogeneous equations, that is, $a_1 a_2 \cdots a_k = g$ for a general g, whereas Green only treats the homogeneous case where $g = 1$. To give some idea of their proof, we will need the following definition.

A directed graph is a graph where each edge has been given a direction. Formally, the edge set may be thought of as a collection of ordered pairs. We will always assume that the directed graph has no loops and does not contain parallel directed edges, though we do allow anti-parallel edges, that is, both the edge $\vec{uv}$ and the edge $\vec{vu}$. The following analogue of the graph removal lemma for directed graphs was proved by Alon and Shapira [9] as part of their study of property testing in directed graphs.

Theorem 4.3 *For any directed graph H and any $\epsilon > 0$, there exists $\delta > 0$ such that any directed graph on n vertices which contains at most $\delta n^{v(H)}$ copies of H may be made H-free by removing at most ϵn^2 edges.*

We will show how to prove Theorem 4.2 with $g = 1$ using Theorem 4.3. Suppose that G is a group of order n and $A_1, \ldots, A_k$ are subsets of G such that there are at most δn^{k-1} solutions to the equation $a_1 a_2 \cdots a_k = 1$ with $a_i \in A_i$ for all i. Consider the auxiliary directed graph Γ whose vertex set is $G \times \{1, 2, \ldots, k\}$. We place an edge from (x, i) to $(y, i+1)$, where addition is taken modulo k, if there exists $a_i \in A_i$ such that $xa_i = y$. It is easy to see that any directed cycle in Γ corresponds to a solution of the equation $a_1 a_2 \cdots a_k = 1$. Moreover, every such solution will result in n different directed cycles in Γ, namely, those with vertices $(x, 1), (xa_1, 2), (xa_1 a_2, 3), \ldots, (xa_1 \cdots a_{k-1}, k)$.

Since G has at most δn^{k-1} solutions to $a_1 a_2 \cdots a_k = 1$, this implies that there are at most δn^k directed cycles in Γ. By Theorem 4.3, for an appropriately chosen δ, we may therefore remove at most $\frac{\epsilon}{k} n^2$ edges to make it free of directed cycles of length k. In A_i, we now remove the element a_i if at least $\frac{n}{k}$ edges of the form $(x, i)(xa_i, i+1)$ have been removed. Note that this results in us removing at most ϵn elements from each A_i. Suppose now that the remaining sets A_i' are such that there is a solution $a_1' a_2' \ldots a_k' = 1$ with $a_i' \in A_i'$ for all i. Then, as above, there are at least n cycles $(x, 1), (xa_1', 2), \ldots, (xa_1' \cdots a_{k-1}', k)$ corresponding to this solution. Since we must have removed one edge from each of these cycles, we must have removed at least $\frac{n}{k}$ edges of the form $(y, i)(ya_i', i+1)$ for some i. But this implies that $a_i' \notin A_i'$, yielding the required contradiction.

It was observed by Fox [38] that δ^{-1} in Theorem 4.3 may, like the graph removal lemma, be taken to be at most a tower of twos of height logarithmic in ϵ^{-1}. This may in turn be used to give a similar bound for δ^{-1} in Theorem 4.2.

In [70], Král', Serra and Vena also showed how to prove a removal lemma for systems of equations which are graph representable, in the sense that they can be put in a natural correspondence with a

directed graph. An example of such a system is

$$x_1 x_2 x_4^{-1} x_3^{-1} = 1$$
$$x_1 x_2 x_5^{-1} = 1.$$

This idea of associating a system of linear equations with a directed graph representation was extended to hypergraphs independently by Král', Serra and Vena [71] and by Shapira [100, 101] in order to prove the following theorem (some partial results had been obtained earlier by Král', Serra and Vena [69], Szegedy [108] and Candela [20]).

Theorem 4.4 *For any natural numbers k and ℓ and any $\epsilon > 0$, there exists $\delta > 0$ such that if F is the field of size n, M is an $\ell \times k$ matrix with coefficients in F, $b \in F^\ell$ and $A_1, \ldots, A_k$ are subsets of F such that there are at most $\delta n^{k-\ell}$ solutions $a = (a_1, \ldots, a_k)$ of the system $Ma = b$ then it is possible to remove at most ϵn elements from each set A_i to form sets A_i' so that there are no solutions $a' = (a_1', \ldots, a_k')$ to the equation $Ma' = b$ with $a_i' \in A_i'$ for all i.*

An easy application of this result shows that a removal lemma for systems of linear equations holds in the set $[n]$, confirming a conjecture of Green [57]. We remark that this result easily implies Szemerédi's theorem. Both proofs use a colored variant of the hypergraph removal lemma due to Austin and Tao [13], though the representations which they use to transfer the problem to hypergraphs are different.

It would be interesting to know whether an analogous statement holds for all groups. A partial extension of these results to abelian groups is proved in [72] (see also [108]) but already in this case there are technical difficulties which do not arise for finite fields.

5 Sparse removal

Given graphs Γ and H, let $N_H(\Gamma)$ be the number of copies of H in Γ. A possible generalization of the graph removal lemma, which corresponds to the case $\Gamma = K_n$, could state that if G is a subgraph of Γ with $N_H(G) \leq \delta N_H(\Gamma)$ then G may be made H-free by deleting at most $\epsilon e(\Gamma)$ edges. Unfortunately, this is too much to hope in general. However, if the graph Γ is sufficiently well-behaved, such an extension does hold. We will discuss two such results here.

5.1 Removal in random graphs

The binomial random graph $G_{n,p}$ is formed by taking n vertices and considering each pair of vertices in turn, choosing each connecting edge to be in the graph independently with probability p. These graphs were introduced by Erdős and Rényi [33, 34] in the late fifties[10] and their study has grown enormously since then (see, for example, the monographs [17, 60]).

Usually, one is interested in finding a threshold function $p^* := p^*(n)$ where the probability that the random graph $G_{n,p}$ has a particular property $\mathcal{P}$ changes from $o(1)$ to $1 - o(1)$ as we pass from random graphs chosen with probability $p \ll p^*$ to those chosen with probability $p \gg p^*$. For example, the threshold for the random graph to be connected is at $p^*(n) = \frac{\ln n}{n}$.

---

[10]The notion was also introduced independently by several other authors at about the same time but, quoting Bollobás [17], "Erdős and Rényi introduced the methods which underlie the probabilistic treatment of random graphs. The other authors were all concerned with enumeration problems and their techniques were essentially deterministic."

One theme that has received a lot of attention in recent years is the question of determining thresholds for the appearance of certain combinatorial properties. One well-studied example is the Ramsey property. Given a graph H and a natural number $r \geq 2$, we say that a graph G is (H, r)-*Ramsey* if in any r-coloring of the edges of G there is guaranteed to be a monochromatic copy of H. Ramsey's theorem [82] is itself the statement that K_n is (H, r)-Ramsey for n sufficiently large. The following celebrated result of Rödl and Ruciński [83, 84] from 1995 (see also [60], Chapter 8) determines the threshold for the appearance of the Ramsey property in random graphs.

Theorem 5.1 *For any graph H that is not a forest consisting of stars and paths of length 3 and every positive integer $r \geq 2$, there exist constants $c, C > 0$ such that*

$$\lim_{n \to \infty} \mathbb{P}\big(G_{n,p} \text{ is } (H, r)\text{-Ramsey}\big) = \begin{cases} 0, & \text{if } p < cn^{-1/m_2(H)}, \\ 1, & \text{if } p > Cn^{-1/m_2(H)}, \end{cases}$$

where

$$m_2(H) = \max \left\{ \frac{e(H') - 1}{v(H') - 2} : H' \subseteq H \text{ and } v(H') \geq 3 \right\}.$$

The threshold occurs at the largest value of p^* such that there is some subgraph H' of H for which the number of copies of H' is approximately the same as the number of edges. For p significantly smaller than p^*, the number of copies of H' will also be significantly smaller than the number of edges. This property allows us (by a rather long and difficult argument [83]) to show that the edges of the graph may be colored in such a way as to avoid any monochromatic copies of H'. For p significantly larger than p^*, every edge of the random graph is contained in many copies of every subgraph of H. The intuition, which takes substantial effort to make rigorous [84], is that these overlaps are enough to force the graph to be Ramsey.

Many related questions were studied in the late nineties. In particular, people were interested in determining the threshold for the following Turán property. Given a graph H and a real number $\epsilon > 0$, we say that a graph G is (H, ϵ)-*Turán* if every subgraph of G with at least

$$\left(1 - \frac{1}{\chi(H) - 1} + \epsilon\right) e(G)$$

edges contains a copy of H. The classical Erdős-Stone-Simonovits theorem [35, 36, 115] states that the graph K_n is (H, ϵ)-Turán for n sufficiently large. Resolving a conjecture of Haxell, Kohayakawa, Łuczak and Rödl [58, 65], Conlon and Gowers [25] and, independently, Schacht [98] proved the following theorem. It is worth noting that the result of Conlon and Gowers applies in the strictly balanced case, that is, when $m_2(H') < m_2(H)$ for all $H' \subset H$, while Schacht's result applies to all graphs. However, the class of strictly balanced graphs includes most of the graphs one would naturally consider, such as cliques or cycles.

Theorem 5.2 *For any graph H[11] and any $\epsilon > 0$, there exist positive constants c and C such that*

$$\lim_{n \to \infty} \mathbb{P}\big(G_{n,p} \text{ is } (H, \epsilon)\text{-Turán}\big) = \begin{cases} 0, & \text{if } p < cn^{-1/m_2(H)}, \\ 1, & \text{if } p > Cn^{-1/m_2(H)}. \end{cases}$$

---

[11]Note that if $H = K_2$, we take $m_2(H) = \frac{1}{2}$.

The results of [25] and [98] (see also [43]) allow one to prove thresholds for the appearance of many different combinatorial properties. For example, the results extend without difficulty to prove analogues of Theorems 5.1 and 5.2 for hypergraphs. The results also apply to give thresholds in different contexts - one example is an extension of Szemerédi's theorem to random subsets of the integers.

Perhaps surprisingly, the methods used in [25] and [98] are very different and have different strengths and weaknesses. We have already mentioned that Schacht's results applied to all graphs while the results of Conlon and Gowers only applied to strictly balanced graphs. On the other hand, the results of [25] also allowed one to transfer structural statements to the sparse setting, including the stability version of the Erdős-Stone-Simonovits theorem [103] and the graph removal lemma. More recently, Samotij [94] modified Schacht's method to extend this sparse stability theorem to all graphs. The result is the following theorem.

Theorem 5.3 *For any graph H and any $\epsilon > 0$, there exist positive constants δ and C such that if $p \geq Cn^{-1/m_2(H)}$ then the following holds a.a.s. in $G_{n,p}$. Every H-free subgraph of $G_{n,p}$ with at least $\left(1 - \frac{1}{\chi(H)-1} - \delta\right)p\binom{n}{2}$ edges may be made $(\chi(H) - 1)$-partite by deleting at most ϵpn^2 edges.*

Recently, a third method was developed by Balogh, Morris and Samotij [15] and, simultaneously and independently, by Saxton and Thomason [97] for proving sparse random analogues of combinatorial theorems. One of the results of their research is a proof of the KŁR conjecture of Kohayakawa, Łuczak and Rödl [65]. This is a technical statement which allows one to prove an embedding lemma complementing the sparse regularity lemma of Kohayakawa [63] and Rödl. A variant of this conjecture has also been proved by Conlon, Gowers, Samotij and Schacht [26] using the methods of [25, 98]. One of the applications of this latter result is the following sparse random analogue of the graph removal lemma (this was already proved for triangles in [64] and for strictly balanced graphs in [25]).

Theorem 5.4 *For any graph H and any $\epsilon > 0$, there exist positive constants δ and C such that if $p \geq Cn^{-1/m_2(H)}$ then the following holds a.a.s. in $G_{n,p}$. Every subgraph of $G_{n,p}$ which contains at most $\delta p^{e(H)}n^{v(H)}$ copies of H may be made H-free by removing at most ϵpn^2 edges.*

Note that for any ϵ there exists a positive constant c such that if $p \leq cn^{-1/m_2(H)}$, the removal lemma is trivial. This is because, for c sufficiently small, the number of copies of the densest subgraph H' of H will a.a.s. be smaller than ϵpn^2. Theorem 5.4 shows that it also holds for $p \geq Cn^{-1/m_2(H)}$. This leaves a small intermediate range of p where it might also be expected that a sparse removal lemma a.a.s. holds. That this is so was conjectured by Łuczak [76].

For balanced graphs H, we may close the gap by letting δ be sufficiently small depending on C, ϵ and H. Indeed, as $p \leq Cn^{-1/m_2(H)}$, the number of copies of H is a.a.s. on the order of $p^{e(H)}n^{v(H)} \leq C^{e(H)}pn^2$. Therefore, taking $\delta < \epsilon C^{-e(H)}$, we see that the number of copies of H is a.a.s. less than ϵpn^2. Deleting one edge from each copy of H in the graph then makes it H-free.

A sparse random analogue of the hypergraph removal lemma was proved in [25] when $\mathcal{H} = K_{k+1}^{(k)}$. This result also extends to cover all strictly balanced hypergraphs.[12] It would be interesting to extend this result to all hypergraphs.

It is worth noting that the sparse random version of the triangle removal lemma does not imply a sparse random version of Roth's theorem. This is because the reduction which allows us to pass from

---

[12]We note that for k-uniform hypergraphs the relevant function is $m_k(\mathcal{H}) = \max\left\{\frac{e(\mathcal{H}')-1}{v(\mathcal{H}')-k}\right\}$, where the maximum is taken over all subgraphs $\mathcal{H}'$ of $\mathcal{H}$ with at least $k+1$ vertices.

a subset of the integers with no arithmetic progressions of length 3 to a graph containing few triangles gives us a graph with dependencies between its edges. This issue does not occur with pseudorandom graphs, which we discuss in the next section.

5.2   Removal in pseudorandom graphs

Though there have long been explicit examples of graphs which behave like the random graph $G_{n,p}$, the first systematic study of what it means for a given graph to be like a random graph was initiated by Thomason [113, 114]. Following him,[13] we say that a graph on vertex set V is (p, β)-*jumbled* if, for all vertex subsets $X, Y \subseteq V$,
$$|e(X, Y) - p|X||Y|| \leq \beta \sqrt{|X||Y|}.$$

The random graph $G_{n,p}$ is, with high probability, (p, β)-jumbled with $\beta = O(\sqrt{pn})$. This is also optimal in that a graph on n vertices with $p \leq 1/2$ cannot be (p, β)-jumbled with $\beta = o(\sqrt{pn})$. The Paley graph is an example of an explicit graph which is optimally jumbled. This graph has vertex set $\mathbb{Z}_p$, where $p \equiv 1 (\mathrm{mod}\ 4)$ is prime, and edge set given by connecting x and y if their difference is a quadratic residue. It is (p, β)-jumbled with $p = \frac{1}{2}$ and $\beta = O(\sqrt{n})$. Many more examples are given in the excellent survey [73].

A fundamental result of Chung, Graham and Wilson [22] states that for graphs of density p, where p is a fixed positive constant, the property of being $(p, o(n))$-jumbled is equivalent to a number of other properties that one would typically expect in a random graph. For example, if the number of cycles of length 4 is as one would expect in a binomial random graph then, surprisingly, this is enough to imply that the edges are very well-spread.

For sparser graphs, the equivalences are less clear cut, but the notion of jumbledness defined above is a natural property to study. Given a graph property $\mathcal{P}$ that one would expect of a random graph, one can ask for the range of p and β for which a (p, β)-jumbled graph satisfies $\mathcal{P}$.

To give an example, it is known that there is a constant c such that if $\beta \leq cp^2 n$ then any (p, β)-jumbled graph contains a triangle. It is also known that this is sharp, since an example of Alon [2] gives a triangle-free graph with $p = \Omega(n^{-1/3})$ which is optimally jumbled, so that $\beta = O(\sqrt{pn}) = O(p^2 n)$.

As in the previous section, one can ask for conditions on p and β which guarantee that a (p, β)-jumbled graph satisfies certain combinatorial properties. For the property of being (K_3, ϵ)-Turán, this question was addressed by Sudakov, Szabó and Vu [107] (see also [21]), who showed that it was enough that $\beta \leq cp^2 n$ for an appropriate c. This is clearly sharp, since for larger values of β we cannot even guarantee that the graph contains a triangle. More generally, they proved the following theorem.[14]

Theorem 5.5 *For any natural number $t \geq 3$ and any $\epsilon > 0$, there exists $c > 0$ such that if $\beta \leq cp^{t-1}n$ then any (p, β)-jumbled graph is (K_t, ϵ)-Turán.*

Except in the case of triangles, there are no known constructions which demonstrate that this theorem is tight. However, it is conjectured [107] that the bound on β in Theorem 5.5 is the correct condition for finding copies of K_t in a (p, β)-jumbled graph. This would in turn imply that Theorem 5.5 is tight.

---

[13]Strictly speaking, Thomason considered a slightly different notion, namely, that $|e(X) - p\binom{|X|}{2}| \leq \beta|X|$ for all $X \subseteq V$, but the two are closely related.

[14]Their results were only stated for the special class of (p, β)-jumbled graphs known as (n, d, λ)-graphs. These are graphs on n vertices which are d-regular and such that all eigenvalues of the adjacency matrix, save the largest, have absolute value at most λ. The expander mixing lemma implies that these graphs are (p, β)-jumbled with $p = \frac{d}{n}$ and $\beta = \lambda$. However, it is not hard to verify that their method applies in the more general case.

For the triangle removal lemma, the following pseudorandom analogue was recently proved by Ko-hayakawa, Rödl, Schacht and Skokan [68].

Theorem 5.6 *For any $\epsilon > 0$, there exist positive constants δ and c such that if $\beta \leq cp^3 n$ then any (p, β)-jumbled graph G on n vertices has the following property. Any subgraph of G containing at most $\delta p^3 n^3$ triangles may be made triangle-free by removing at most $\epsilon p n^2$ edges.*

The condition on β in this theorem is stronger than that employed for triangles in Theorem 5.5. As a result, Alon's construction does not apply and it is an open problem to determine whether the condition $\beta \leq cp^3 n$ is optimal or if it can be improved to $\beta \leq cp^2 n$. Kohayakawa, Rödl, Schacht and Skokan conjecture the latter, though we feel that the former is a genuine possibility.

In a recent paper, Conlon, Fox and Zhao [27] found a way to prove a counting lemma for embedding any fixed small graph into a regular subgraph of a sufficiently pseudorandom host graph. Like the KŁR conjecture for random graphs, this serves to complement the sparse regularity lemma of Kohayakawa [63] and Rödl in the pseudorandom context. As corollaries, they extended Theorems 5.5 and 5.6 to all graphs and proved sparse pseudorandom extensions of several other theorems, including Ramsey's theorem and the Erdős-Simonovits stability theorem.

To state these theorems, we define the *degeneracy* $d(H)$ of a graph H to be the smallest nonnegative integer d for which there exists an ordering of the vertices of H such that each vertex has at most d neighbors which appear earlier in the ordering. Equivalently, it may be defined as $d(H) = \max\{\delta(H') : H' \subseteq H\}$, where $\delta(H)$ is the minimum degree of H.[15]

The pseudorandom analogue of the graph removal lemma proved in [27] is now as follows.[16]

Theorem 5.7 *For any graph H and any $\epsilon > 0$, there exist positive constants δ and c such that if $\beta \leq cp^{d(H)+\frac{5}{2}} n$ then any (p, β)-jumbled graph G on n vertices has the following property. Any subgraph of G containing at most $\delta p^{e(H)} n^{v(H)}$ copies of H may be made H-free by removing at most $\epsilon p n^2$ edges.*

It is not hard to show, by using the random graph, that there are (p, β)-jumbled graphs with $\beta = O(p^{(d(H)+2)/4} n)$ which contain no copies of H. We therefore see that the exponent of p is sharp up to a multiplicative constant. However, in many cases, we expect it to be sharp up to an additive constant.

For certain classes of graph, Theorem 5.7 can be improved. For example, if we know that the degeneracy of the graph is the same as the maximum degree, such as what happens for the complete graph K_t, it is sufficient that $\beta \leq cp^{d(H)+1} n$. In particular, for K_3, we reprove Theorem 5.6. For cycles, the improvement is even more pronounced, since $\beta \leq cp^{t_\ell} n$, where $t_3 = 3$, $t_4 = 2$, $t_\ell = 1 + \frac{1}{\ell-3}$ if $\ell \geq 5$ is odd and $t_\ell = 1 + \frac{1}{\ell-4}$ if $\ell \geq 6$ is even, is sufficient for removing the cycle C_ℓ.

By following the proof of Král', Serra and Vena [70], these bounds on the cycle removal lemma in pseudorandom graphs[17] allow us to prove an analogue of Theorem 4.2 for pseudorandom subsets of any group G. The *Cayley graph* $G(S)$ of a subset S of a group G has vertex set G and (x, y) is an edge of G if $x^{-1}y \in S$. We say that a subset S of a group G is (p, β)-jumbled if the Cayley graph $G(S)$ is

---

[15]In [27], a slightly different parameter, the 2-degeneracy $d_2(H)$, is used. Though there are many cases in which this parameter is more appropriate, the degeneracy will be sufficient for the purposes of our discussion here.

[16]For other properties, such as that of being (H, r)-Ramsey or that of being (H, ϵ)-Turán, an exactly analogous theorem holds with the same condition $\beta \leq cp^{d(H)+\frac{5}{2}} n$. Any of the improvements subsequently discussed for specific graphs H also apply for these properties.

[17]Rather, a colored or directed version of this theorem.

(p, β)-jumbled. When G is abelian, if $\left| \sum_{x \in S} \chi(x) \right| \leq \beta$ for all nontrivial characters $\chi \colon G \to \mathbb{C}$, then S is $(\frac{|S|}{|G|}, \beta)$-jumbled (see [68, Lemma 16]).

Theorem 5.8 *For any natural number $k \geq 3$ and any $\epsilon > 0$, there exist positive constants δ and c such that the following holds. Suppose $B_1, \ldots, B_k$ are subsets of a group G of order n such that each B_i is (p, β)-jumbled with $\beta \leq cp^{t_k} n$. If subsets $A_i \subseteq B_i$ for $i = 1, \ldots, k$ are such that there are at most $\delta |B_1| \cdots |B_k|/n$ solutions to the equation $x_1 x_2 \cdots x_k = 1$ with $x_i \in A_i$ for all i, then it is possible to remove at most $\epsilon |B_i|$ elements from each set A_i so as to obtain sets A_i' for which there are no solutions to $x_1 x_2 \cdots x_k = 1$ with $x_i \in A_i'$ for all i.*

This result easily implies a Roth-type theorem in quite sparse pseudorandom subsets of a group. We say that a subset B of a group G is (ϵ, k)-*Roth* if, for all integers $a_1, \ldots, a_k$ which satisfy $a_1 + \cdots + a_k = 0$ and $\gcd(a_i, |G|) = 1$ for $1 \leq i \leq k$, every subset $A \subseteq B$ which has no nontrivial solution to $x_1^{a_1} x_2^{a_2} \cdots x_k^{a_k} = 1$ has $|A| \leq \epsilon |B|$.

Corollary 5.1 *For any natural number $k \geq 3$ and any $\epsilon > 0$, there exists $c > 0$ such that the following holds. If G is a group of order n and B is a (p, β)-jumbled subset of G with $\beta \leq cp^{t_k} n$, then B is (ϵ, k)-Roth.*

Note that Roth's theorem on 3-term arithmetic progressions in dense sets of integers follows from the special case of this result with $B = G = \mathbb{Z}_n$, $k = 3$ and $a_1 = a_2 = 1$, $a_3 = -2$. The rather weak pseudorandomness condition in Corollary 5.1 shows that even quite sparse pseudorandom subsets of a group have the Roth property.

6 Further topics

6.1 The Erdős-Rothschild problem

A problem of Erdős and Rothschild [30] asks one to estimate the maximum number $h(n, c)$ such that every n-vertex graph with at least cn^2 edges, each of which is contained in at least one triangle, must contain an edge that is in at least $h(n, c)$ edges. Here, and throughout this subsection, we assume $c > 0$ is a fixed absolute constant. The fact that $h(n, c)$ tends to infinity already follows from the triangle removal lemma.[18]

To see this, suppose that G is an n-vertex graph with cn^2 edges such that every edge is in at least one and at most $h := h(n, c)$ triangles. The total number of triangles in G is at most $hcn^2/3$. Therefore, if h does not tend to infinity, the triangle removal lemma tells us that there is a collection E of $o(n^2)$ edges such that every triangle contains at least one of them. Since each edge in G is in at least one triangle, we know that there are at least $cn^2/3$ triangles. It follows that some edge in E is contained in at least $\omega(1)$ edges.

Using Fox's bound [38] for the triangle removal lemma, this implies that $h(n, c) \geq e^{a \log^* n}$, where $\log^* n$ is the iterated logarithm. This is defined by $\log^* x = 0$ if $x \leq 1$ and $\log^* x = \log^*(\log x) + 1$ otherwise. This improves on the bound $h(n, c) \geq (\log^* n)^a$ which follows from Ruzsa and Szemerédi's original proof of the triangle removal lemma.

---

[18]Even the statement that $h(n, c) > 1$ is already enough to imply Roth's theorem.

On the other hand, Alon and Trotter (see [31]) showed that for any positive $c < \frac{1}{4}$ there is $c' > 0$ such that $h(n,c) < c'\sqrt{n}$. The condition $c < \frac{1}{4}$ is easily seen to be best possible since any n-vertex graph with more than $n^2/4$ edges contains an edge in at least $n/6$ triangles [29, 62]. Erdős conjectured that perhaps this behaviour is correct. That is, that for any positive $c < \frac{1}{4}$ there exists $\epsilon > 0$ such that $h(n,c) > n^\epsilon$ for all sufficiently large n. This was recently disproved by Fox and Loh [39] as follows.

Theorem 6.1 *For n sufficiently large, there is an n-vertex graph with $\frac{n^2}{4}(1 - e^{-(\log n)^{1/6}})$ edges such that every edge is in a triangle and no edge is in more than $n^{14/\log\log n}$ triangles.*

To give some idea of the construction, consider a tripartite graph between sets A, B and C, each of which is a copy of a lattice cube with appropriate sidelength r and dimension d. We join points in A and B if their distance is close to the expected distance between random points in A and B. By concentration, this implies that the density of edges between A and B is close to 1. We join points in C to points in A or B if their distance is close to half the expected distance. It is not hard to see that every edge between A and B is then contained in few triangles. At the same time, every edge will be in at least one triangle, as can be seen by considering the midpoint of any two connected points a and b. This yields a construction with roughly $\frac{n^2}{9}$ edges but the result of Fox and Loh may be obtained by shrinking the vertex set C (or blowing up A and B) in an appropriate fashion.

6.2 Induced matchings

Call a graph $G = (V, E)$ an (r, t)-*Ruzsa-Szemerédi graph* ((r, t)-RS graph for short) if its edge set can be partitioned into t induced matchings in G, each of size r. The total number of edges of such a graph is rt. The most interesting problem concerns the existence of such graphs when r and t are both relatively large as a function of the number of vertices. The construction of Ruzsa and Szemerédi [93] using Behrend's construction demonstrates that such a graph on n vertices exists with $r = e^{-c\sqrt{\log n}}n$ and $t = n/3$. The Ruzsa-Szemerédi result on the $(6, 3)$-problem is equivalent to showing that no (r, t)-RS graph on n vertices exists with r and t linear in n.

For r linear in the number n of vertices, it is still an open problem if there exists an (r, t)-RS graph with $t = n^\epsilon$. The best known construction in this case, due to Fischer et al. [37], is an example with $r = n/3$ and $t = n^{c/\log\log n}$. However, for $r = n^{1-o(1)}$, substantial progress was made recently by Alon, Moitra and Sudakov [8] by extending ideas used in the construction of Fox and Loh [39] discussed in the previous subsection. They give a construction of n-vertex graphs with $rt = (1 - o(1))\binom{n}{2}$ and $r = n^{1-o(1)}$. That is, there are nearly complete graphs, with edge density $1 - o(1)$, such that its edge set can be partitioned into large induced matchings, each of order $n^{1-o(1)}$. They give several applications of this construction to combinatorics, complexity theory and information theory.

6.3 Testing small graphs

A *property* of graphs is a family of graphs closed under isomorphism. A graph G on n vertices is ϵ-*far* from satisfying a property P if no graph which can be constructed from G by adding and/or removing at most ϵn^2 edges satisfies P. An ϵ-*tester* for P is a randomized algorithm which, given the quantity n and the ability to make queries whether a desired pair of vertices spans an edge in G, distinguishes with probability at least $2/3$ between the case that G satisfies P and the case that G is ϵ-far from satisfying P. Such an ϵ-tester is a *one-sided ϵ-tester* if when G satisfies P the ϵ-tester determines that

this is the case. The property P is called *testable* if, for every fixed $\epsilon > 0$, there exists a one-sided ϵ-tester for P whose total number of queries is bounded only by a function of ϵ which is independent of the size of the input graph. This means that the running time of the algorithm is also bounded by a function of ϵ only and is independent of the input size. We measure query-complexity by the number of vertices sampled, assuming we always examine all edges spanned by them. The infinite removal lemma, Theorem 3.2, of Alon and Shapira [12] shows that every hereditary graph property, that is, a graph property closed under taking induced subgraphs, is testable. Many of the best studied graph properties are hereditary.

If the query complexity of an ϵ-tester is polynomial in ϵ^{-1}, we say that the property is *easily testable*. It is an interesting open problem to characterize the easily testable hereditary properties. Alon [3] considered the case where $P = P_H$ is the property that the graph does not contain H as a subgraph. He showed that P_H is easily testable if and only if H is bipartite. Alon and Shapira [10] considered the case where $P = P_H^*$ is the property that the graph does not contain H as an *induced* subgraph. They showed that for any graph H except for the path with at most four vertices, the cycle of length four and their complements, the property P_H^* is not easily testable. The problem of determining whether the property P_H^* is easily testable for the path with four vertices or the cycle of length four (or equivalently its complement) was left open. The case where H is a path with four vertices was recently shown to be easily testable by Alon and Fox [7]. The case where H is a cycle of length four is still open. Alon and Fox also showed that if P is the family of perfect graphs, then P is not easily testable and, in a certain sense, testing for P is at least as hard as testing triangle-freeness.

6.4   Local repairability

The standard proof of the regularity lemma contains a procedure for turning a graph which is almost triangle-free into a graph which is triangle-free. We simply delete the edges between all vertex sets of low density and between all vertex sets which do not form a regular pair. This procedure can be made more explicit still by using an algorithmic version of the regularity lemma [4].

A surprising observation of Austin and Tao [13] is that this repair procedure can be determined in a local fashion. They show that for any graph H and any $\epsilon > 0$ there exists $\delta > 0$ and a natural number m such that if G is a graph containing at most $\delta n^{v(H)}$ copies of H then there exists a set A of size at most m such that G may be made H-free by removing at most ϵn^2 edges and the decision of whether to delete a given edge uv may be determined solely by considering the restriction of G to the set $A \cup \{u, v\}$.[19]

The key point, first observed by Ishigami [59], is that the regular partition can be determined in a local fashion by randomly selecting vertex neighborhoods to create the partition. Since a finite set of points determine the partition, this may in turn be used to create a local modification rule which results in an H-free graph.

Similar ideas may also be applied to show that any hereditary graph property, including the property of being induced H-free, is locally repairable in the same sense. This again follows from the observation that random neighborhoods can be used to construct the partitions arising in the strong regularity lemma.

Surprisingly, Austin and Tao show that, even though all hereditary hypergraph properties are testable,

---

[19]Strictly speaking, Austin and Tao [13] consider two forms of local repairability. Here we are considering only the weak version.

there are hereditary properties which are not locally repairable. On the other hand, they show that many natural hypergraph properties, including the property of being $\mathcal{H}$-free, are locally repairable.

6.5 Linear hypergraphs

A *linear hypergraph* is a hypergraph where any pair of edges overlap in at most one vertex. For this special class of hypergraphs, it is not necessary to apply the full strength of hypergraph regularity to prove a corresponding removal lemma [66]. Instead, a straightforward analogue of the usual regularity lemma is sufficient. This results in bounds for δ^{-1} in the linear hypergraph removal lemma which are of tower-type in a power of ϵ^{-1}.

While this is already a substantial improvement on general hypergraphs, where the best known bounds are Ackermannian,[20] it can be improved further by using the ideas of [38]. This results in a bound of the form $T(a_{\mathcal{H}} \log \epsilon^{-1})$.

A similar reduction does not exist for induced removal of linear hypergraphs. Because we need to consider all edges, whether present or not, between the vertices of the hypergraph, we must apply the full strength of the strong hypergraph regularity lemma. This results in Ackermannian bounds.

It is plausible that an extension of the methods of Section 2.2 could be used to give a primitive recursive, or even tower-type, bound for hypergraph removal. We believe that such an improvement would be of great interest, not least because it would give the first primitive recursive bound for the multidimensional extension of Szemerédi's theorem. Such an improvement would also be likely to lead to an analogous improvement of the bounds for induced hypergraph removal.

References

[1] M. Ajtai and E. Szemerédi, Sets of lattice points that form no squares, *Stud. Sci. Math. Hungar.* **9** (1974), 9–11.

[2] N. Alon, Explicit Ramsey graphs and orthonormal labellings, *Electron. J. Combin.* **1** (1994), R12, 8pp.

[3] N. Alon, Testing subgraphs in large graphs, *Random Structures Algorithms* **21** (2002), 359–370.

[4] N. Alon, R. A. Duke, H. Lefmann, V. Rödl and R. Yuster, The algorithmic aspects of the regularity lemma, *J. Algorithms* **16** (1994), 80–109.

[5] N. Alon, W. Fernandez de la Vega, R. Kannan and M. Karpinski, Random sampling and approximation of MAX-CSPs, *J. Comput. System Sci.* **67** (2003), 212–243.

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

---

[20]We have already seen two levels of the Ackermann function, the tower function and the wowzer function. Generally, the kth level is defined by $A_k(1) = 2$ and $A_k(i+1) = A_{k-1}(A_k(i))$. Taking $A_1(i) = 2^i$, we see that $A_2(i) = T(i)$ and $A_3(i) = W(i)$. The upper bound on δ^{-1} in the k-uniform hypergraph removal lemma given by the hypergraph regularity proofs are of the form $A_k(\epsilon^{-O(1)})$ or worse.

[7] N. Alon and J. Fox, Testing perfectness is hard, submitted.

[8] N. Alon A. Moitra and B. Sudakov, Nearly complete graphs decomposable into large induced matchings and their applications, submitted.

[9] N. Alon and A. Shapira, Testing subgraphs in directed graphs, *J. Comput. System Sci.* **69** (2004), 353–382.

[10] N. Alon and A. Shapira, A characterization of easily testable induced subgraphs, *Combin. Probab. Comput.* **15** (2006), 791–805.

[11] N. Alon and A. Shapira, Every monotone graph property is testable, in Proc. of STOC 2005, 128–137, *SIAM J. Comput.* (Special Issue of STOC'05) **38** (2008), 505–522.

[12] N. Alon and A. Shapira, A characterization of the (natural) graph properties testable with one-sided error, in Proc. of FOCS 2005, 429–438, *SIAM J. Comput.* (Special Issue on FOCS '05) **37** (2008), 1703–1727.

[13] T. Austin and T. Tao, Testability and repair of hereditary hypergraph properties, *Random Structures Algorithms* **36** (2010), 373–463.

[14] C. Avart, V. Rödl and M. Schacht, Every monotone 3-graph property is testable, *SIAM J. Discrete Math.* **21** (2007), 73–92.

[15] J. Balogh, R. Morris and W. Samotij, Independent sets in hypergraphs, *submitted.*

[16] F. Behrend, On sets of integers which contain no three terms in arithmetic progression, *Proc. Nat. Acad. Sci.* **32** (1946), 331–332.

[17] B. Bollobás, **Random graphs**, second edition, Cambridge Studies in Advanced Mathematics 73, Cambridge University Press, Cambridge, 2001.

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

[19] W. G. Brown, P. Erdős and V. T. Sós, On the existence of triangulated spheres in 3-graphs, and related problems, *Period. Math. Hungar.* **3** (1973), 221–228.

[20] P. Candela, Developments at the interface between combinatorics and Fourier analysis, PhD thesis, University of Cambridge, 2009.

[21] F. R. K. Chung, A spectral Turán theorem, *Combin. Probab. Comput.* **14** (2005), 755–767.

[22] F. R. K. Chung, R. L. Graham and R. M. Wilson, Quasi-random graphs, *Combinatorica* **9** (1989), 345–362.

[23] L. H. Clark, R. C. Entringer, J. E. McCanna and L. A. Székely, Extremal problems for local properties of graphs, in Combinatorial mathematics and combinatorial computing (Palmerston North, 1990), *Australas. J. Combin.* **4** (1991), 25–31.

[24] D. Conlon and J. Fox, Bounds for graph regularity and removal lemmas, *Geom. Funct. Anal.* **22** (2012), 1192–1256.

[25] D. Conlon and W. T. Gowers, Combinatorial theorems in sparse random sets, *submitted.*

[26] D. Conlon, W. T. Gowers, W. Samotij and M. Schacht, On the KŁR conjecture in random graphs, *submitted.*

[27] D. Conlon, J. Fox and Y. Zhao, Extremal results in sparse pseudorandom graphs, *submitted.*

[28] R. A. Duke, H. Lefmann and V. Rödl, A fast approximation algorithm for computing the frequencies of subgraphs in a given graph, *SIAM J. Comput.* **24** (1995), 598–620.

[29] C. S. Edwards, A lower bound for the largest number of triangles with a common edge, 1977, unpublished manuscript.

[30] P. Erdős, Some problems on finite and infinite graphs, in Logic and combinatorics (Arcata, Calif., 1985), 223–228, Contemp. Math. 65, Amer. Math. Soc., Providence, RI, 1987.

[31] P. Erdős, Some of my favourite problems in various branches of combinatorics, in Combinatorics 92 (Catania, 1992), *Matematiche (Catania)* **47** (1992), 231–240.

[32] 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.

[33] P. Erdős and A. Rényi, On random graphs I, *Publ. Math. Debrecen* **6** (1959), 290–297.

[34] P. Erdős and A. Rényi, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17–61.

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

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

[37] E. Fischer, I. Newman, S. Raskhodnikova, R. Rubinfeld and A. Samorodnitsky, Monotonicity testing over general poset domains, in Proceedings of the 2002 ACM Symposium on Theory of Computing, 474–483, ACM, New York, 2002.

[38] J. Fox, A new proof of the graph removal lemma, *Ann. of Math.* **174** (2011), 561–579.

[39] J. Fox and P. Loh, On a problem of Erdős and Rothschild on edges in triangles, to appear in *Combinatorica.*

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

[41] P. Frankl, R. L. Graham and V. Rödl, On subsets of abelian groups with no 3-term arithmetic progression, *J. Combin. Theory Ser. A* **45** (1987), 157–161.

[42] P. Frankl and V. Rödl, Extremal problems on set systems, *Random Structures Algorithms* **20** (2002), 131–164.

[43] E. Friedgut, V. Rödl and M. Schacht, Ramsey properties of discrete random structures, *Random Structures Algorithms*, **37** (2010), 407–436.

[44] A. Frieze and R. Kannan, The regularity lemma and approximation schemes for dense problems, *Proceedings of the 37th IEEE FOCS* (1996), 12–20.

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

[46] Z. Füredi, The maximum number of edges in a minimal graph of diameter 2, *J. Graph Theory* **16** (1992), 81–98.

[47] Z. Füredi, Extremal hypergraphs and combinatorial geometry, in Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), 1343–1352, Birkhäuser, Basel, 1995.

[48] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* **31** (1977), 204–256.

[49] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, *J. Analyse Math.* **34** (1978), 275–291.

[50] H. Furstenberg, Y. Katznelson and D. Ornstein, The ergodic theoretical proof of Szemerédi's theorem, *Bull. Amer. Math. Soc.* **7** (1982), 527–552.

[51] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its applications to learning and approximation, *J. ACM* **45** (1998), 653–750.

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

[53] W. T. Gowers, A new proof of Szemerédi's theorem for arithmetic progressions of length four, *Geom. Funct. Anal.* **8** (1998), 529–551.

[54] W.T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11** (2001), 465–588.

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

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

[57] B. Green, A Szemerédi-type regularity lemma in abelian groups, with applications, *Geom. Funct. Anal.* **15** (2005), 340–376.

[58] P. E. Haxell, Y. Kohayakawa and T. Łuczak, Turán's extremal problem in random graphs: forbidding odd cycles, *Combinatorica* **16** (1996), 107–122.

[59] Y. Ishigami, A simple regularization of hypergraphs, *submitted*.

[60] S. Janson, T. Łuczak and A. Ruciński, **Random graphs,** Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[61] S. Kalyanasundaram and A. Shapira, A wowzer-type lower bound for the strong regularity lemma, *Proc. London Math. Soc.,* to appear.

[62] N. Khadžiivanov and V. Nikiforov, Solution of a problem of P. Erdős about the maximum number of triangles with a common edge in a graph, *C. R. Acad. Bulgare Sci.* **32** (1979), 1315–1318.

[63] Y. Kohayakawa, Szemerédi's regularity lemma for sparse graphs, in Foundations of computational mathematics (Rio de Janeiro, 1997), Springer, Berlin, 1997, 216–230.

[64] Y. Kohayakawa, T. Łuczak and V. Rödl, Arithmetic progressions of length three in subsets of a random set, *Acta Arith.* **75** (1996), 133–163.

[65] Y. Kohayakawa, T. Łuczak and V. Rödl, On K^4-free subgraphs of random graphs, *Combinatorica* **17** (1997), 173–213.

[66] Y. Kohayakawa, B. Nagle, V. Rödl and M. Schacht, Weak regularity and linear hypergraphs, *J. Combin. Theory Ser. B* **100** (2010), 151–160.

[67] Y. Kohayakawa, B. Nagle, V. Rödl, J. Skokan and M. Schacht, The hypergraph regularity method and its applications, *Proc. Natl. Acad. Sci. USA* **102** (2005), 8109–8113.

[68] Y. Kohayakawa, V. Rödl, M. Schacht and J. Skokan, On the triangle removal lemma for subgraphs of sparse pseudorandom graphs, in An Irregular Mind (Szemerédi is 70), Bolyai Society Math. Studies 21, Springer, 2010, 359–404.

[69] D. Král', O. Serra and L. Vena, A removal lemma for linear systems over finite fields, in Sixth conference on discrete mathematics and computer science (Spanish), 417–423, Univ. Lleida, Lleida, 2008.

[70] D. Král', O. Serra and L. Vena, A combinatorial proof of the removal lemma for groups, *J. Combin. Theory Ser. A* **116** (2009), 971–978.

[71] D. Král', O. Serra and L. Vena, A removal lemma for systems of linear equations over finite fields, *Israel J. Math.* **187** (2012), 193–207.

[72] D. Král', O. Serra and L. Vena, On the removal lemma for linear systems over abelian groups, *European J. Combin.* **34** (2013), 248–259.

[73] M. Krivelevich and B. Sudakov, Pseudo-random graphs, in More sets, graphs and numbers, Bolyai Soc. Math. Stud. 15, Springer, Berlin, 2006, 199–262.

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

[75] L. Lovász and B. Szegedy, Testing properties of graphs and functions, *Israel J. Math.* **178** (2010), 113–156.

[76] T. Łuczak, Randomness and regularity, in International Congress of Mathematicians, Vol. III, 899–909, Eur. Math. Soc., Zürich, 2006.

[77] B. Nagle and V. Rödl, Regularity properties for triple systems, *Random Structures Algorithms* 23 (2003), 264–332.

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

[79] Y. Peng, V. Rödl, and A. Ruciński, Holes in graphs, *Electron. J. Combin.* **9** (2002), R1, 18pp.

[80] Y. Peng, V. Rödl and J. Skokan, Counting small cliques in 3-uniform hypergraphs, *Combin. Probab. Comput.* **14** (2005), 371–413.

[81] D. H. J. Polymath, A new proof of the density Hales-Jewett theorem, *Ann. of Math.* **175** (2012), 1283–1327.

[82] F. P. Ramsey, On a problem of formal logic, *Proc. London Math. Soc.* **30** (1930), 264–286.

[83] V. Rödl and A. Ruciński, Lower bounds on probability thresholds for Ramsey properties, in Combinatorics, Paul Erdős is eighty, Vol. 1, 317–346, Bolyai Soc. Math. Stud., János Bolyai Math. Soc., Budapest, 1993.

[84] V. Rödl and A. Ruciński, Threshold functions for Ramsey properties, *J. Amer. Math. Soc.* **8** (1995), 917–942.

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

[86] V. Rödl and M. Schacht, Generalizations of the removal lemma, *Combinatorica* **29** (2009), 467–501.

[87] V. Rödl and M. Schacht, Regularity lemmas for graphs, in Fete of Combinatorics and Computer Science, Bolyai Soc. Math. Stud. 20, Springer, 2010, 287–325.

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

[89] V. Rödl and J. Skokan, Counting subgraphs in quasi-random 4-uniform hypergraphs, Random Structures Algorithms **26** (2005), 160–203.

[90] V. Rödl and J. Skokan, Applications of the regularity lemma for uniform hypergraphs, *Random Structures Algorithms* **28** (2006), 180–194.

[91] K. F. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953), 104–109.

[92] R. Rubinfield and M. Sudan, Robust characterization of polynomials with applications to program testing, *SIAM J. Comput.* **25** (1996), 252–271.

[93] I. Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai 18, Volume II, 939–945.

[94] W. Samotij, Stability results for random discrete structures, to appear in *Random Structures Algorithms*.

[95] T. Sanders, On Roth's theorem on progressions, *Ann. of Math.* **174** (2011), 619–636.

[96] T. Sanders, On the Bogolyubov-Ruzsa lemma, *Anal. PDE*, to appear.

[97] D. Saxton and A. Thomason, Hypergraph containers, *submitted*.

[98] M. Schacht, Extremal results for random discrete structures, *submitted*.

[99] T. Schoen and I. Shkredov, Roth's theorem in many variables, *submitted.*

[100] A. Shapira, Green's conjecture and testing linear-invariant properties, in Proceedings of the 2009 ACM International Symposium on Theory of Computing, 159–166, ACM, New York, 2009.

[101] A. Shapira, A proof of Green's conjecture regarding the removal properties of sets of linear equations, *J. London Math. Soc.* **81** (2010), 355–373.

[102] I. Shkredov, On a generalization of Szemerédi's theorem, *Proc. London Math. Soc.* **93** (2006), 723–760.

[103] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, in Theory of graphs (Proc. Colloq. Tihany,1966), Academic Press, New York, 1968, 279–319.

[104] J. Solymosi, Note on a generalization of Roth's theorem, in Discrete and computational geometry, Algorithms Combin. Vol. 25, Springer, 2003, 825–827.

[105] J. Solymosi, A note on a question of Erdős and Graham, *Combin. Probab. Comput.* **13** (2004), 263–267.

[106] J. Solymosi, Regularity, uniformity, and quasirandomness, *Proc. Natl. Acad. Sci. USA* **102** (2005), 8075–8076

[107] B. Sudakov, T. Szabó and V.H. Vu, A generalization of Turán's theorem, *J. Graph Theory* **49** (2005), 187–195.

[108] B. Szegedy, The symmetry preserving removal lemma, *Proc. Amer. Math. Soc.* **138** (2010), 405–408.

[109] E. Szemerédi, Integer sets containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 299–345.

[110] E. Szemerédi, Regular partitions of graphs, in Colloques Internationaux CNRS 260 - Problèmes Combinatoires et Théorie des Graphes, Orsay, (1976), 399–401.

[111] T. Tao, Szemerédi's regularity lemma revisited, *Contrib. Discrete Math.* **1** (2006), 8–28.

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

[113] A. Thomason, Pseudorandom graphs, in Random graphs '85 (Poznań, 1985), 307–331, North-Holland Math. Stud. 144, North-Holland, Amsterdam, 1987.

[114] A. Thomason, Random graphs, strongly regular graphs and pseudorandom graphs, Surveys in combinatorics 1987 (New Cross, 1987), 173–195, London Math. Soc. Lecture Note Ser. 123, Cambridge Univ. Press, Cambridge, 1987.

[115] P. Turán, Eine Extremalaufgabe aus der Graphentheorie, *Mat. Fiz. Lapok* **48** (1941), 436–452.