The hypergraph regularity method and its applications

V. Rödl*, B. Nagle†, J. Skokan‡, M. Schacht§¶, and Y. Kohayakawa‡

*Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322; †Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557; ‡Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, CEP 05508-090, São Paulo, Brazil; and §Institut für Informatik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany

Szemerédi's regularity lemma asserts that every graph can be decomposed into relatively few random-like subgraphs. This random-like behavior enables one to find and enumerate subgraphs of a given isomorphism type, yielding the so-called counting lemma for graphs. The combined application of these two lemmas is known as the regularity method for graphs and has proved useful in graph theory, combinatorial geometry, combinatorial number theory, and theoretical computer science. Here, we report on recent advances in the regularity method for k-uniform hypergraphs, for arbitrary $k \geq 2$. This method, purely combinatorial in nature, gives alternative proofs of density theorems originally due to E. Szemerédi, H. Furstenberg, and Y. Katznelson. Further results in extremal combinatorics also have been obtained with this approach. The two main components of the regularity method for k-uniform hypergraphs, the regularity lemma and the counting lemma, have been obtained recently: Rödl and Skokan (based on earlier work of Frankl and Rödl) generalized Szemerédi's regularity lemma to k-uniform hypergraphs, and Nagle, Rödl, and Schacht succeeded in proving a counting lemma accompanying the Rödl–Skokan hypergraph regularity lemma. The counting lemma is proved by reducing the counting problem to a simpler one previously investigated by Kohayakawa, Rödl, and Skokan. Similar results were obtained independently by W. T. Gowers, following a different approach.

Szemerédi's theorem | regularity lemma | counting lemma | removal lemma

In 1975, Szemerédi (1) confirmed a long-standing conjecture of Erdős and Turán (2) concerning the upper density of sets containing no arithmetic progression with a fixed number of elements.

Theorem 1 [Szemerédi's theorem (1)]. *For every positive integer t and every $\delta > 0$, there exists an integer $N_0 = N_0(t, \delta)$ such that, for $N \geq N_0$, any subset $Z \subset [N] = \{1, \ldots, N\}$ with $|Z| \geq \delta N$ contains an arithmetic progression with t elements.*

In 1977, shortly after Szemerédi's combinatorial proof appeared, Furstenberg (3) gave an alternative proof of *Theorem 1* using methods of ergodic theory. Refining the techniques of that proof, Furstenberg and Katznelson later were able to derive several other density versions of combinatorial partition theorems. The following theorem, which may be viewed as a density version of the Gallai–Witt theorem, is one of them (in what follows, we denote by $[-t; t]$ the set of all integers i satisfying $-t \leq i \leq t$).

Theorem 2 [Furstenberg and Katznelson (4)]. *Let T be a finite subset of $\mathbb{R}^d$ and let $\delta > 0$ be given. Then there exists a finite subset C of $\mathbb{R}^d$ such that any subset $Z \subset C$ with $|Z| > \delta|C|$ contains a homothetic copy of T, i.e., a set of the form $z + \lambda T$ for some $z \in \mathbb{R}^d$ and some $\lambda \neq 0$. Moreover, if $T \subset [-t; t]^d$ for some positive integer t, then $C = [-N; N]^d$ has the above property for every sufficiently large $N \geq N_0(t, d, \delta)$.*

Note that the special case of *Theorem 2* for $d = 1$ implies *Theorem 1*. More generally, for fixed d, the above result allows us to find a homothetic copy of a full-dimensional cube $[-t; t]^d$ in any dense subset of a sufficiently large cube of the same dimension. Two other results in a similar vein address the complementary case in which the dimension is allowed to grow.

Theorem 3 [Furstenberg and Katznelson (5)]. *Let $\mathbb{F}_q$ be the finite field with q elements. For every positive integer d and every $\delta > 0$, there exists $M_0 = M_0(q, d, \delta)$ such that, for $M \geq M_0$, any subset $Z \subset \mathbb{F}_q^M$ with $|Z| > \delta|\mathbb{F}_q^M| = \delta q^M$ contains a d-dimensional affine subspace.*

Theorem 4 [Furstenberg and Katznelson (5)]. *Let G be a finite abelian group and let $\delta > 0$ be given. Then there exists $M_0 = M_0(G, \delta)$ such that if $M \geq M_0$ and Z is a subset of G^M with $|Z| > \delta|G|^M$, then Z contains a coset of a subgroup of G^M isomorphic to G.*

The techniques introduced by Furstenberg and Katznelson have been extended to prove other generalizations of *Theorems 1–4*, among which are a density version of the Hales–Jewett theorem (6), again due to Furstenberg and Katznelson (7), and polynomial extensions of Szemerédi's theorem, due to Bergelson and Leibman (8) and Bergelson and McCutcheon (9).

Another area of investigation concerns estimates on N_0 and M_0 in *Theorems 1–4*. Szemerédi's original proof of *Theorem 1* uses the regularity lemma for graphs (10), upcoming *Theorem 9*, which forces (see ref. 11) the upper bound on $N_0 = N_0(t, \delta)$ to exceed a tower function of height polynomial in $1/\delta$. Using, among others, methods of Fourier analysis, Gowers (12) gave an alternative proof of *Theorem 1* rendering the immensely improved estimate

$$N_0 \leq \exp\left(\exp\left((1/\delta)^{2^{2t+9}}\right)\right).$$

The original proofs of *Theorems 2–4* and their generalizations rely on ergodic theory and do not yield any upper bounds on N_0 and M_0.

The aim of this work is to report on an extremal result for hypergraphs, which can be called "removal lemma" (see *Theorem 5*). This lemma has a number of applications, including purely combinatorial proofs of *Theorems 1–4*. This approach also yields the first quantitative bounds on N_0 and M_0 in *Theorems 2–4*. The bounds are, however, poor: they belong to a level of the Ackermann hierarchy that depends on the input parameters.

The proof of the removal lemma is based on an extension of the "regularity method" from graphs to "uniform hypergraphs." We present the removal lemma and some of its consequences below. The fundamentals of the regularity method for hypergraphs are discussed in the remainder of the work.

Main Result: The Removal Lemma

For a set V and an integer $k \geq 1$, let $\binom{V}{k}$ be the set of all k-element subsets of V. A subset $\mathcal{H}^{(k)} \subset \binom{V}{k}$ is a k-uniform hypergraph or k-graph for short, on the vertex set V.

The following result was conjectured by Erdős, Frankl, and Rödl (13) in the case where $\mathcal{F}^{(k)}$ is a clique.

Theorem 5 (Removal Lemma). *For all fixed integers $l \geq k \geq 2$ and every $\mu > 0$ there exist $\zeta = \zeta(l, k, \mu) > 0$ and $n_0 = n_0(l, k, \mu)$ so that the following holds. Suppose $\mathcal{F}^{(k)}$ is a k-graph on l vertices and $\mathcal{H}^{(k)}$ is a k-graph on $n \geq n_0$ vertices. If $\mathcal{H}^{(k)}$ contains at most ζn^l copies of $\mathcal{F}^{(k)}$, then one can delete μn^k edges of $\mathcal{H}^{(k)}$ to make it $\mathcal{F}^{(k)}$-free.*

Ruzsa and Szemerédi (14) proved *Theorem 5* for $k = 2$ and $\mathcal{F}^{(2)}$ being the triangle $K_3^{(2)}$, the graph consisting of three edges on three vertices. Frankl and Rödl (15) proved *Theorem 5* for $k = 3$ and $\mathcal{F}^{(3)} = K_4^{(3)}$, the 3-graph consisting of four triples on four vertices. These proofs follow what we call the regularity method for graphs and 3-graphs, respectively. Extending concepts developed in ref. 15, the regularity method for k-graphs was established in refs. 16 and 17. The proof of *Theorem 5* for general k follows lines similar to those in refs. 14 and 15 [see ref. 17 for the case $\mathcal{F}^{(k)} = K_{k+1}^{(k)}$ and ref. 18 for general $\mathcal{F}^{(k)}$]. For illustrative purposes, we later sketch the derivation of Ruzsa and Szemerédi's result by using the regularity method for graphs.

We mention that Gowers (W. T. Gowers, personal communication; see also ref. 19) has also independently developed similar regularity techniques for proving *Theorem 5*. Very recently, Tao (T. Tao, personal communication) announced another proof of *Theorem 5*.

Applications

Before discussing the components of the regularity method for k-graphs, we turn to some applications of the removal lemma. As mentioned earlier, the removal lemma implies *Theorems 1–4*. This was shown for *Theorem 1* by Frankl and Rödl in ref. 15. Subsequently, Solymosi (20) established a similar relation for *Theorem 2*. Elaborating on an idea from ref. 15, Tengan, Tokushige, V.R., and M.S. (21) derived the following common generalization of *Theorems 3* and *4*.

Theorem 6. *Let A be a finite ring with q elements. For every $\delta > 0$, there exists $M_0 = M_0(q, \delta)$ such that, for $M \geq M_0$, any subset $Z \subset A^M$ with $|Z| > \delta|A^M| = \delta q^M$ contains a coset of an isomorphic copy of A (as a left A-module). In other words, there exist $r, u \in A^M$ such that $r + \varphi(A) \subset Z$, where $\varphi: A \longhookrightarrow A^M$, $\varphi(\alpha) = \alpha u$ for $\alpha \in A$, is an injection.*

Theorem 5 also implies the affirmative answer to a geometric problem of Székely (ref. 22, p. 226): For a point $c = (c_1, \ldots, c_k)$ in the k-fold cross-product of $[n]$ with itself, we define the jack $J(c)$ with center c as the set of points that differ from c in at most one coordinate. For $1 \leq i \leq k$ and fixed $c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_k \in [n]$, we also define a line as a set of n points of the form

$$\{(c_1, \ldots, c_{i-1}, x, c_{i+1}, \ldots, c_k): 1 \leq x \leq n\}.$$

Let $\mathrm{LS}(n, k)$ be the maximum cardinality of a system J of jacks for which

(*i*) no two distinct jacks share a common line; and
(*ii*) $\cap_{i=1}^{k} J_i = \varnothing$ for all distinct jacks $J_1, \ldots, J_k \in J$.

The following result (18) gives a positive answer to Székely's problem.

Theorem 7. $\mathrm{LS}(n, k) = o(n^{k-1})$.

As one would expect from the literature for graphs (see, e.g., ref. 23), the hypergraph regularity method also may be used to obtain further results in the areas of extremal and asymptotic hypergraph problems. Here we mention one such result.

Let $\mathcal{F}$ be a finite family of k-graphs and let $\mathrm{Forb}(\mathcal{F})$ be the set of all labeled k-graphs not containing any $\mathcal{F}^{(k)} \in \mathcal{F}$ as a subhypergraph. Moreover, let $\mathrm{Forb}(n, \mathcal{F})$ be those hypergraphs of $\mathrm{Forb}(\mathcal{F})$ with vertex set $[n]$ and set

$$\mathrm{ex}(n, \mathcal{F}) = \max\{|\mathcal{H}^{(k)}| : \mathcal{H}^{(k)} \in \mathrm{Forb}(n, \mathcal{F})\}.$$

Note that all subhypergraphs of any fixed $\mathcal{H}^{(k)} \in \mathrm{Forb}(n, \mathcal{F})$ also belong to $\mathrm{Forb}(n, \mathcal{F})$, and in particular, this holds for $\mathcal{H}_0^{(k)} \in \mathrm{Forb}(n, \mathcal{F})$ achieving $|\mathcal{H}_0^{(k)}| = \mathrm{ex}(n, \mathcal{F})$. The inequality $\log_2|\mathrm{Forb}(n, \mathcal{F})| \geq \mathrm{ex}(n, \mathcal{F})$ thus follows. The next theorem (B.N., V.R., and M.S., unpublished data) asserts this bound is essentially best possible whenever $\mathcal{F}$ satisfies $\mathrm{ex}(n, \mathcal{F}) = \Omega(n^k)$.

Theorem 8. *For any finite family $\mathcal{F}$ of k-graphs,*

$$\log_2|\mathrm{Forb}(n, \mathcal{F})| = \mathrm{ex}(n, \mathcal{F}) + o(n^k).$$

Theorem 8 complements a collection of analogous theorems already proven for graphs (i.e., $k = 2$), and generalizes results from refs. 13, 24, and 25.

We now discuss the regularity method for graphs and hypergraphs.

Szemerédi's Regularity Lemma

In the course of proving *Theorem 1*, Szemerédi established a lemma that decomposes the edge set of any graph into relatively few "blocks," almost all of which are random-like (10). In what follows, we give a precise account of Szemerédi's lemma.

A pair (A, B) of two disjoint vertex subsets $A, B \subset V$ of a graph $G = (V, E)$ is said to be ε-regular if for all $A' \subset A, B' \subset B$, where $|A'| > \varepsilon|A|$, $|B'| > \varepsilon|B|$, we have $|d(A, B) - d(A', B')| < \varepsilon$ where $d(A', B') = |E(A', B')|/(|A'||B'|)$ and $E(A', B')$ denotes the set of edges of G with an endpoint in each of A' and B'. We denote by $G[A, B]$ the bipartite subgraph of G induced on A and B, and we say that $G[A, B]$ is ε-regular if (A, B) is ε-regular. Szemerédi's lemma may then be stated as follows.

Theorem 9 (Szemerédi's Regularity Lemma). *Let $\varepsilon > 0$ be given and let t_0 be a positive integer. There exists a positive integer $T_0 = T_0(\varepsilon, t_0)$ such that any graph $G = (V, E)$ admits a partition $V = V_1 \cup \cdots \cup V_t$, $t_0 \leq t \leq T_0$, satisfying*

(*i*) $|V_1| \leq \cdots \leq |V_t| \leq |V_1| + 1$, *and*
(*ii*) *all but at most $\varepsilon\binom{t}{2}$ pairs (V_i, V_j), $1 \leq i < j \leq t$, are ε-regular.*

Szemerédi's regularity lemma is a powerful tool in extremal graph theory. One of its most important consequences is that, in appropriate circumstances, it can be used to show that a given graph contains a fixed subgraph. This observation follows from the following well-known and easily proved fact, which may be called the "counting lemma" for graphs.

Fact 10 (Counting Lemma). *For all positive integers l and $d > 0$ and $\gamma > 0$ there exists $\varepsilon = \varepsilon(l, d, \gamma) > 0$ so that the following holds. Let $G = \cup_{1 \leq i < j \leq l} G^{ij}$ be an l-partite graph with l-partition $V_1 \cup \cdots \cup V_l$, where $G^{ij} = G[V_i, V_j]$, $1 \leq i < j \leq l$, and $|V_1| = \cdots = |V_l| = n$. Suppose further that each graph G^{ij}, $1 \leq i < j \leq l$, is ε-regular with density d. Then the number of copies of the l-clique K_l in G is within the interval $(1 \pm \gamma)d^{\binom{l}{2}}n^l$.*

We refer to the combined application of *Theorem 9* and *Fact 10* as the regularity method for graphs. This method yields *Theorem 5* in the case of graphs ($k = 2$), the proof of which can be traced to Ruzsa and Szemerédi (14). We now sketch this proof in the special case in which $\mathcal{F}^{(2)} = K_3^{(2)}$ is the triangle.

We first address the promised constants. To that end, let $\mu > 0$ be given. To define $\zeta = \zeta(3, 2, \mu) > 0$, we first let $\gamma = 1/2$, $d_0 =$

$\mu/3$, and $t_0 = \lceil 3/\mu \rceil$ and take $\varepsilon = \varepsilon(3, d_0, \gamma) > 0$ to be the constant guaranteed by *Fact 10*. Without loss of generality, we may assume that $\varepsilon < \mu/3$. For $t_0 = \lceil 3/\mu \rceil$ and ε defined above, let $T_0 = T_0(\varepsilon, t_0)$ be the constant guaranteed by *Theorem 9*. We define $\zeta = d_0^3/(4T_0^3)$. We take n_0 sufficiently large whenever needed.

Suppose $H = \mathcal{H}^{(2)}$ is a graph on $n \geq n_0$ vertices and suppose H contains at most ζn^3 copies of the triangle $K_3^{(2)}$. We exhibit a triangle-free subgraph H' of H with only μn^2 fewer edges and do so by appealing to the regularity lemma, *Theorem 9*.

Indeed, with input parameters $\varepsilon > 0$ and t_0 previously defined, apply *Theorem 9* to the graph H to obtain a partition $V(H) = V_1 \cup \cdots \cup V_t$, $t_0 \leq t \leq T_0$, satisfying the properties in the conclusion of that theorem. Note that the constant T_0, guaranteed by *Theorem 9*, is the same as we considered earlier when defining ζ. To obtain the promised triangle-free subgraph H', delete from H any edge $\{v_i, v_j\} \in E(H)$, $v_i \in V_i$, $v_j \in V_j$, $1 \leq i, j \leq t$, for which either $i = j$, or $d_H(V_i, V_j) < d_0$, or the pair (V_i, V_j) is not ε-regular.

It is easy to see that the process above deletes at most

$$t \binom{\lceil n/t \rceil}{2} + \binom{t}{2} d_0 \left[\frac{n}{t} \right]^2 + \varepsilon \binom{t}{2} \left[\frac{n}{t} \right]^2 \leq \mu n^2$$

edges, where the last inequality follows from our choice of constants. We claim the resulting subgraph H' is triangle-free. Suppose, on the contrary, that H' contains a copy of $K_3^{(2)}$ with vertex set $\{v_h, v_i, v_j\}$, where, by construction of H', we may assume $v_h \in V_h$, $v_i \in V_i$ and $v_j \in V_j$, for $1 \leq h < i < j \leq t$. Then, it must also be the case that each of (V_h, V_i), (V_i, V_j), and (V_h, V_j) are ε-regular with respective densities at least d_0. *Fact 10* then implies that the 3-partite subgraph $H'[V_h, V_i, V_j]$ induced on $V_h \cup V_i \cup V_j$ contains at least

$$\left(1 - \frac{1}{2}\right) d_0^3 \left[\frac{n}{t} \right]^3 > \left(\frac{d_0^3}{4T_0^3} \right) n^3 = \zeta n^3$$

copies of $K_3^{(2)}$, contradicting our hypothesis that H had at most ζn^3 such copies.

Regularity Method for Hypergraphs

Szemerédi's regularity lemma decomposes any given graph into pseudorandom blocks, the ε-regular pairs. This notion of pseudorandomness admits the companion counting statement in *Fact 10*. To develop a regularity method for hypergraphs, one needs a concept of pseudorandomness that allows one to prove both a regularity lemma (which decomposes any given hypergraph into pseudorandom blocks) and a counting lemma (which estimates the number of hypergraphs of a given isomorphism type in an appropriate collection of pseudorandom blocks).

Various concepts capturing the notion of pseudorandomness for hypergraphs have been studied; see Haviland and Thomason (26) and Chung and Graham (27–30). "Deviation," the central concept in refs. 29 and 30, admits a companion counting statement, as does "discrepancy," a different notion of pseudorandomness for hypergraphs, studied in ref. 31. No matching regularity lemma, however, is known for either deviation or discrepancy.

Before we introduce the notion of pseudorandomness to be employed, we note that, to extend *Theorem 9* to hypergraphs, one also needs to establish a suitable concept of partition. Recall that in *Theorem 9* the main structure that undergoes regularization is the edge set of a graph, and a certain partition of the vertex set is an auxiliary structure. Briefly, the 2-tuples (edges) are regularized versus the 1-tuples (vertices). If for k-graphs, $k \geq 3$, we just regularize the k-tuples versus 1-tuples, as, e.g., in refs. 32–34,

then the natural analogue to the counting lemma fails to be true (see ref. 35 for a counterexample).

A more refined approach is to consider an auxiliary partition of the j-tuples for each $j < k$. This idea was pursued in refs. 36 and 37 with no attempt, however, to prove a companion counting statement. Building on the ideas from ref. 37, Frankl and Rödl improved the concept of hypergraph regularity for $k = 3$ in ref. 15 and proved a corresponding regularity lemma for 3-graphs. They also succeeded in proving a companion counting lemma for the special case $\mathcal{F} = K_4^{(3)}$, the complete 3-graph on four vertices. Subsequently, the regularity lemma from ref. 15 was extended to k-graphs for arbitrary $k \geq 3$ in ref. 16.

We shall now outline the regularity lemmas of refs. 15 and 16 and begin by fixing some notation. A k-uniform clique of order j, denoted by $K_j^{(k)}$, is a k-graph on $j \geq k$ vertices consisting of all $\binom{j}{k}$ many k-tuples [i.e., $K_j^{(k)}$ is isomorphic to $\binom{[j]}{k}$].

Let $l \geq k \geq 2$ be fixed integers. For each integer $2 \leq j < k$, let $\mathcal{G}^{(j)}$ be an l-partite j-graph with vertex partition $\mathcal{G}^{(1)} = V_1 \cup \cdots \cup V_l$. Let $\mathcal{K}_j(\mathcal{G}^{(1)}) = K_l^{(j)}(V_1, \ldots, V_l)$ be the complete l-partite j-graph, and, for $2 \leq i < j < k$, let $\mathcal{K}_j(\mathcal{G}^{(i)})$ be the family of all j-element vertex sets that span the clique $K_j^{(i)}$ in $\mathcal{G}^{(i)}$.

For $j \geq 3$, fix classes $V_{i_1}, \ldots, V_{i_j}$, $1 \leq i_1 < \cdots < i_j \leq l$. For an integer $r \geq 1$, let $Q^{(j-1)} = \{Q_1^{(j-1)}, \ldots, Q_r^{(j-1)}\}$ be a family of subhypergraphs of $\mathcal{G}^{(j-1)}[V_{i_1}, \ldots, V_{i_j}]$, the j-partite subgraph of $\mathcal{G}^{(j-1)}$ induced on $V_{i_1} \cup \cdots \cup V_{i_j}$. We define the density or relative density

$$d(\mathcal{G}^{(j)} | Q^{(j-1)})$$

of $\mathcal{G}^{(j)}$ with respect to $Q^{(j-1)}$ as

$$\frac{|\mathcal{G}^{(j)} \cap \bigcup_{s \in [r]} \mathcal{K}_j(Q_s^{(j-1)})|}{|\bigcup_{s \in [r]} \mathcal{K}_j(Q_s^{(j-1)})|}$$

if

$$\left| \bigcup_{s \in [r]} \mathcal{K}_j(Q_s^{(j-1)}) \right| > 0$$

and 0 otherwise.

We are now in position to introduce a notion of pseudorandomness for j-graphs. For positive reals δ_j and d_j and integer $r \geq 1$, we say that $\mathcal{G}^{(j)}$ is (δ_j, d_j, r)-regular with respect to (w.r.t.) $\mathcal{G}^{(j-1)}$ if for any choice of classes $V_{i_1}, \ldots, V_{i_j}$, $1 \leq i_1 < \cdots < i_j \leq l$, and any collection $Q^{(j-1)} = \{Q_1^{(j-1)}, \ldots, Q_r^{(j-1)}\}$ of subhypergraphs of $\mathcal{G}^{(j-1)}[V_{i_1}, \ldots, V_{i_j}]$ satisfying

$$\left| \bigcup_{s \in [r]} \mathcal{K}_j(Q_s^{(j-1)}) \right| \geq \delta_j |\mathcal{K}_j(\mathcal{G}^{(j-1)}[V_{i_1}, \ldots, V_{i_j}])|,$$

we have

$$d(\mathcal{G}^{(j)} | Q^{(j-1)}) = d_j \pm \delta_j.$$

This concept of regularity gives control over the structure of the hypergraph $\mathcal{G}^{(j)}$ with respect to the hypergraph $\mathcal{G}^{(j-1)}$. To gain more control on the structure of $\mathcal{G}^{(j)}$, one imposes additional structural assumptions on $\mathcal{G}^{(j-1)}$. In ref. 16, it is assumed that $\mathcal{G}^{(j-1)}$ is itself $(\delta_{j-1}, d_{j-1}, r)$-regular w.r.t. some $\mathcal{G}^{(j-2)}$, which again is $(\delta_{j-2}, d_{j-2}, r)$-regular w.r.t. some $\mathcal{G}^{(j-3)}$, etc.

The discussion above leads to the definition of an (l, h)-complex. For $l \geq h$, an (l, h)-complex G is a system $\{\mathcal{G}^{(j)}\}_{j=1}^h$ of l-partite j-graphs $\mathcal{G}^{(j)}$ satisfying $\mathcal{G}^{(j)} \subset \mathcal{K}_j(\mathcal{G}^{(j-1)})$ for $2 \leq j \leq h$, with the same vertex partition $\mathcal{G}^{(1)} = V_1 \cup \cdots \cup V_l$. Given vectors of positive reals $\boldsymbol{\delta} = (\delta_2, \ldots, \delta_h)$ and $\boldsymbol{d} = (d_2, \ldots, d_h)$ and an integer $r \geq 1$, we say that an (l, h)-complex G is $(\boldsymbol{\delta}, \boldsymbol{d}, r)$-regular if

(i) for each $1 \leq i_1 < i_2 \leq l$, $\mathcal{G}^{(2)}[V_{i_1}, V_{i_2}]$ is δ_2-regular with density $d_2 \pm \delta_2$;

(ii) for each $3 \leq j \leq h$, $\mathcal{G}^{(j)}$ is (δ_j, d_j, r)-regular with respect to $\mathcal{G}^{(j-1)}$.

Informally speaking, regular complexes are the pseudorandom blocks into which the hypergraph regularity lemma decomposes any large enough hypergraph.

We now proceed to describe the partition analogue of the hypergraph regularity lemma in ref. 16. To this end, we shall need the concept of a $(\mu, \boldsymbol{\delta}, \boldsymbol{d}, r)$-equitable family of partitions, which, as we shall see momentarily, is a sequence of partitions of vertices, pairs, triples, ..., $(k-1)$-tuples. Let $\mu > 0$ be a real number, let $\boldsymbol{\delta} = (\delta_2, \ldots, \delta_{k-1})$ and $\boldsymbol{d} = (d_2, \ldots, d_{k-1})$ be vectors of positive reals, and let $r \geq 1$ be an integer. Let $\boldsymbol{a} = (a_1, \ldots, a_{k-1})$ be a vector of positive integers and V an n-element vertex set. We say that a family of partitions $\mathcal{P} = \mathcal{P}(k-1, \boldsymbol{a}) = \{\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(k-1)}\}$ on V is $(\mu, \boldsymbol{\delta}, \boldsymbol{d}, r)$-equitable if it satisfies the following conditions:

(P1) $\mathcal{P}^{(1)} = \{V_i : i \in [a_1]\}$ is an equitable vertex partition of V, i.e., $|V_1| \leq \cdots \leq |V_{a_1}| \leq |V_1| + 1$,

(P2) $\mathcal{P}^{(j)}$ partitions $\mathcal{K}_j(G^{(1)}) = K_{a_1}^{(j)}(V_1, \ldots, V_{a_1})$ so that if $P_1^{(j-1)}, \ldots, P_j^{(j-1)} \in \mathcal{P}^{(j-1)}$ and $\mathcal{K}_j(\cup_{i=1}^j P_i^{(j-1)}) \neq \varnothing$, then $\mathcal{K}_j(\cup_{i=1}^j P_i^{(j-1)})$ is partitioned into at most a_j parts, all of them members of $\mathcal{P}^{(j)}$, and, most importantly,

(P3) for all but at most μn^k k-tuples $K \in \binom{V}{k}$ there is a unique $(\boldsymbol{\delta}, \boldsymbol{d}, r)$-regular $(k, k-1)$-complex $P = P(K) = \{\mathcal{P}^{(j)}\}_{j=1}^{k-1}$ such that $\mathcal{P}^{(j)}$ has as members $\binom{k}{j}$ different partition classes from $\mathcal{P}^{(j)}$ and $K \in \mathcal{K}_k(\mathcal{P}^{(k-1)}) \subset \cdots \subset \mathcal{K}_k(\mathcal{P}^{(1)})$.

The complex $P = \{\mathcal{P}^{(j)}\}_{j=1}^{k-1}$ mentioned in (P3) takes the place of the pairs (V_i, V_j) in *Theorem 9*. We say that a k-graph $\mathcal{H}^{(k)}$ is (δ_k, r)-regular w.r.t. a family of partitions $\mathcal{P}$ if all but at most $\delta_k n^k$ edges K of $\mathcal{H}^{(k)}$ have the property that $K \in \mathcal{K}_k(G^{(1)})$ and if $P = P(K) = \{\mathcal{P}^{(j)}\}_{j=1}^{k-1}$ is the unique $(k, k-1)$-complex for which $K \in \mathcal{K}_k(\mathcal{P}^{(k-1)})$, then $\mathcal{H}^{(k)}$ is $(\delta_k, d(\mathcal{H}^{(k)}|\mathcal{P}^{(k-1)}), r)$-regular w.r.t. $\mathcal{P}^{(k-1)}$.

The hypergraph regularity lemma from ref. 16, stated below, asserts that for every k-graph $\mathcal{H}^{(k)}$ there exists a $(\mu, \boldsymbol{\delta}, \boldsymbol{d}, r)$-equitable family of partitions $\mathcal{P}$ with a bounded number of complexes so that $\mathcal{H}^{(k)}$ is (δ_k, r)-regular w.r.t. $\mathcal{P}$.

Theorem 11 (Hypergraph Regularity Lemma). *For all positive reals μ and δ_k and functions*

$$\delta_j : (0, 1]^{k-j} \to (0, 1] \quad \text{for } j = 2, \ldots, k-1,$$

$$\text{and } r : \mathbb{N} \times (0, 1]^{k-2} \to \mathbb{N},$$

there exist T_0 and n_0 so that the following holds. For every k-graph $\mathcal{H}^{(k)}$ on $n \geq n_0$ vertices, there exist a family of partitions $\mathcal{P} = \mathcal{P}(k-1, \boldsymbol{a})$ and a vector $\boldsymbol{d} = (d_2, \ldots, d_{k-1})$ so that, for $\boldsymbol{\delta} = (\delta_2, \ldots, \delta_{k-1})$, where $\delta_j = \delta_j(d_j, \ldots, d_{k-1})$ for all j, and $r = r(a_1, \boldsymbol{d})$, the following holds:

(i) *$\mathcal{P}$ is a $(\mu, \boldsymbol{\delta}, \boldsymbol{d}, r)$-equitable family of partitions and $a_i \leq T_0$ for every $i = 1, \ldots, k-1$ and*

(ii) *$\mathcal{H}^{(k)}$ is (δ_k, r)-regular w.r.t. $\mathcal{P}$.*

In essence, this is similar to the regularity lemma for graphs, *Theorem 9*, and the regularity lemma for 3-graphs (15). For instance, for a graph G, Szemerédi's regularity lemma implies

that most of the edges of G belong to ε-regular, bipartite subgraphs $G[V_i, V_j]$. In the above concept of hypergraph regularity, the $(\boldsymbol{\delta}, \boldsymbol{d}, r)$-regular (k, k)-complexes $\{\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(k-1)}, \mathcal{H}^{(k)} \cap \mathcal{K}_k(\mathcal{P}^{(k-1)})\}$ correspond to the ε-regular pairs in Szemerédi's partition. More specifically, the auxiliary structure $\mathcal{P} = \{\mathcal{P}^{(j)}\}_{j=1}^{k-1}$ corresponds to the pairs of vertex sets (V_i, V_j), while $\mathcal{H}^{(k)} \cap \mathcal{K}_k(\mathcal{P}^{(k-1)})$ corresponds to the edge set $G[V_i, V_j]$ induced by V_i and V_j.

Note that in *Theorem 11*, for each $2 \leq j \leq k-1$, the constant δ_j may be chosen as a function of $d_j, \ldots, d_{k-1}$. Therefore, we may ensure $\delta_j \ll \min\{d_j, \ldots, d_{k-1}\}$. However, we have no control over the relation between δ_j and d_{j-1}. In particular, *Theorem 11* cannot avoid the outcome $\delta_j \gg d_{j-1}$, thus forcing us to face the following hierarchy concerning the constant δ_k and the constant entries of the vectors $\boldsymbol{\delta}$ and $\boldsymbol{d}$:

$$\gamma, d_k \gg \delta_k \gg d_{k-1} \gg \delta_{k-1} \gg \cdots$$

$$\gg d_h \gg \delta_h \gg d_{h-1} \gg \cdots \gg d_2 \gg \delta_2. \qquad [1]$$

For a counting lemma for hypergraphs to be an appropriate counterpart of the regularity lemma in *Theorem 11*, it necessarily has to match the quantification stated in Eq. **1**. Such a result has been obtained in ref. 17.

Theorem 12 (Hypergraph Counting Lemma). *For all integers $2 \leq k \leq l$ the following holds: $\forall \gamma > 0 \; \forall d_k > 0 \; \exists \delta_k > 0 \; \forall d_{k-1} > 0 \; \exists \delta_{k-1} > 0 \cdots \forall d_2 > 0 \; \exists \delta_2 > 0$ and there are integers r and m_0 so that, with $\boldsymbol{d} = (d_2, \ldots, d_k)$, $\boldsymbol{\delta} = (\delta_2, \ldots, \delta_k)$ and $m \geq m_0$, if $G = \{G^{(j)}\}_{j=1}^k$ is a $(\boldsymbol{\delta}, \boldsymbol{d}, r)$-regular (l, k)-complex with vertex partition $G^{(1)} = V_1 \cup \cdots \cup V_l$ and $|V_i| = m$, then*

$$|\mathcal{K}_l(G^{(k)})| = (1 \pm \gamma) \prod_{h=2}^k d_h^{\binom{l}{h}} \times m^l.$$

Unlike the case of graphs ($k = 2$), *Theorem 12* for $k \geq 3$ was the most technical part of the approach outlined here. The proof of *Theorem 12* establishes a reduction to a simpler counting problem, previously addressed in corollary 6.11 of ref. 31. The special case $k = 3$ of *Theorem 12* has been considered by three of us (B.N., V.R., and M.S., unpublished data) following a similar but simpler approach.

1. Szemerédi, E. (1975) *Acta Arithmetica* **27,** 199–245.
2. Erdős, P. & Turán, P. (1936) *J. London Math. Soc.* **11,** 261–264.
3. Furstenberg, H. (1977) *J. Analyse Math.* **31,** 204–256.
4. Furstenberg, H. & Katznelson, Y. (1978) *J. Analyse Math.* **34,** 275–291.
5. Furstenberg, H. & Katznelson, Y. (1985) *J. Analyse Math.* **45,** 117–168.
6. Hales, A. W. & Jewett, R. I. (1963) *Trans. Am. Math. Soc.* **106,** 222–229.
7. Furstenberg, H. & Katznelson, Y. (1991) *J. Anal. Math.* **57,** 64–119.
8. Bergelson, V. & Leibman, A. (1996) *J. Am. Math. Soc.* **9,** 725–753.
9. Bergelson, V. & McCutcheon, R. (2000) *Memoirs Am. Math. Soc.* **146,** viii–106.
10. Szemerédi, E. (1978) in *Problèmes Combinatoires et Théorie des Graphes, Colloquium International Centre National de la Recherche Scientifique, University of Orsay, Orsay, 1976*, eds. Bermond, J. C., Fournier, J. C., Las Vergnas, M. & Sotteau, D. (Centre National de la Recherche Scientifique, Paris), pp. 399–401.

11. Gowers, W. T. (1997) *Geometric Funct. Anal.* **7,** 322–337.
12. Gowers, W. T. (2001) *Geometric Funct. Anal.* **11,** 465–588.
13. Erdős, P., Frankl, P. & Rödl, V. (1986) *Graphs Combinatorics* **2,** 113–121.
14. Ruzsa, I. Z. & Szemerédi, E. (1978) in *Combinatorics: Proceedings of the Fifth Hungarian Colloquium, Keszthely, 1976*, eds. Hajnal, A. & Sós, V. T. (North–Holland, Amsterdam, Vol. II), pp. 939–945.
15. Frankl, P. & Rödl, V. (2002) *Rand. Struct. Algorithms* **20,** 131–164.
16. Rödl, V. & Skokan, J. (2004) *Rand. Struct. Algorithms* **25,** 1–42.
17. Nagle, B., Rödl, V. & Schacht, M. (2005) *Rand. Struct. Algorithms*, in press.
18. Rödl, V. & Skokan, J. (2005) *Rand. Struct. Algorithms*, in press.
19. Gowers, W. T. (2005) *Combin. Probab. Comput.*, in press.
20. Solymosi, J. (2004) *Combin. Probab. Comput.* **13,** 263–267.
21. Rödl, V., Schacht, M., Tengan, E. & Tokushige, N. (2005) *Isr. J. Math.*, in press.

22. Matoušek, J. (2002) *Lectures on Discrete Geometry*, Graduate Texts in Mathematics (Springer, New York).
23. Komlós, J., Shokoufandeh, A., Simonovits, M. & Szemerédi, E. (2002) in *Theoretical Aspects of Computer Science (Tehran, 2000)*, Lecture Notes in Computer Science, eds. Khosrovshahi, G. B., Shokoufandeh, A. & Shokrollahi, M. A. (Springer, Berlin), Vol. 2292, pp. 84–112.
24. Erdős, P., Kleitman, D. J. & Rothschild, B. L. (1976) in *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973)*, Atti dei Convegni Lincei (Accad. Naz. Lincei, Rome), Vol. II, No. 17, pp. 19–27.
25. Nagle, B. & Rödl, V. (2001) *Discrete Math.* **235,** 271–290.
26. Haviland, J. & Thomason, A. (1989) *Discrete Math.* **75,** 255–278.
27. Chung, F. R. K. & Graham, R. L. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8175–8177.
28. Chung, F. R. K. (1990) *Rand. Struct. Algorithms* **1,** 363–382.
29. Chung, F. R. K. & Graham, R. L. (1990) *Rand. Struct. Algorithms* **1,** 105–124.
30. Chung, F. R. K. & Graham, R. L. (1991) *J. Am. Math. Soc.* **4,** 151–196.
31. Kohayakawa, Y., Rödl, V. & Skokan, J. (2002) *J. Combin. Theory Ser. A* **97,** 307–352.
32. Prömel, H. J. & Steger, A. (1992) *Rand. Struct. Algorithms* **3,** 19–31.
33. Frieze, A. & Kannan, R. (1999) *Combinatorica* **19,** 175–220.
34. Czygrinow, A. & Rödl, V. (2000) *SIAM J. Comput.* **30,** 1041–1066.
35. Nagle, B. & Rödl, V. (2003) *Rand. Struct. Algorithms* **23,** 264–332.
36. Chung, F. R. K. (1991) *Rand. Structures Algorithms* **2,** 241–252.
37. Frankl, P. & Rödl, V. (1992) *Graphs Combinatorics* **8,** 309–312.