Szemerédi's Lemma for the Analyst

László Lovász and Balázs Szegedy
Microsoft Research

February 2006

Abstract

Szemerédi's Regularity Lemma is a fundamental tool in graph theory: it has many applications to extremal graph theory, graph property testing, combinatorial number theory, etc. The goal of this paper is to point out that Szemerédi's Lemma can be thought of as a result in analysis. We show three different analytic interpretations.

1 Introduction

Szemerédi's Regularity Lemma was first used in his celebrated proof of the Erdős–Turán Conjecture on arithmetic progressions in dense sets of integers. Since then, the Lemma has emerged as a fundamental tool in graph theory: it has many applications in extremal graph theory, in the area of "Property Testing" in computer science, combinatorial number theory, etc.

Roughly speaking, the Szemereédi Regularity Lemma says that the node set of every (large) graph can be partitioned into a small number of parts so that the subgraphs between the parts are "random-like". There are several ways to make this precise, some equivalent to the original version, some not (see Section 2 for an exact statement).

The goal of this paper is not to describe the many applications of this Lemma (see [16, 15] for surveys and discussions of such applications); nor will we discuss extensions to sparse graphs by [11], or hypergraphs by Frankl–Rödl, Gowers and Tao [9, 14, 20]. Tao [21] describes the lemma as a result in probability. Our goal is to point out that Szemerédi's Lemma can be thought of as a result in analysis. We show three different analytic interpretations. The first one is a general statement about approximating elements in Hilbert spaces which implies many different versions of the Regularity Lemma, and also potentially other approximation results. The second one presents the Regularity Lemma as the compactness of an important metric space of 2-variable functions. We prove the compactness by using a weak version of the Regularity Lemma but (somewhat surprisingly) this compactness implies strong versions of the Regularity Lemma very easily. The third analytic interpretation shows the connection between a weak version of the Regularity Lemma and covering by small diameter sets, i.e., dimensionality.

We describe two applications of this third version: a lower bound on the number of classes in the "weak" version of the Szemerédi Lemma, and an algorithm that constructs the "weak" Szemerédi partition as Voronoi cells in a metric space.

2 Strong and Weak Regularity Lemma

We start with stating a standard version of the Lemma. For a graph $G = (V, E)$ and for $X, Y \subseteq V$, let $e_G(X, Y)$ denote the number of edges with one endnode in X and another in Y; edges with both endnodes in $X \cap Y$ are counted twice.

Let G be a bipartite graph G with bipartition $\{U, W\}$. The ratio $d = d_G(U, W) = \frac{e_G(U,W)}{|U|\cdot|W|}$ can be thought of as the density of edges between U and W. On the average, we expect that for $X \subseteq U$ and $Y \subseteq W$,

$$e_G(X, Y) \approx d|X| \cdot |Y|.$$

For two arbitrary subsets of the nodes, $e_G(X, Y)$ may be very far from this "expected value". If G is a random graph, then, however, it will be close; random graphs are very "homogeneous" in this respect. So the following definition captures how "random-like" the bipartite graph G is: We say that G is *ε-regular*, if

$$\left| \frac{e_G(X, Y)}{|X| \cdot |Y|} - d \right| \leq \varepsilon$$

holds for all subsets $X \subseteq U$ and $Y \subseteq W$ such that $|X| > \varepsilon|U|$ and $|Y| > \varepsilon|W|$. Notice that we could not require the condition to hold for small X and Y: for example, if both have one element, then the quotient $e_G(X, Y)/(|X| \cdot |Y|)$ is either 0 or 1.

Let $G = (V, E)$ be a graph (not necessarily bipartite) and let S, T be disjoint subsets of V. We denote by $G[S, T]$ the bipartite graph on $S \cup T$ obtained by keeping just those edges of G that connect S and T.

A partition $\{V_1, \dots, V_k\}$ of V is called an *equipartition* if $\lfloor |V|/k \rfloor \leq |V_i| \leq \lceil |V_i|/k \rceil$ for all $1 \leq i \leq k$.

With these definitions, the Regularity Lemma can be stated as follows:

Lemma 2.1 (Szemerédi Regularity Lemma, usual form) *For every $\varepsilon > 0$ and $m > 0$ there is a $k = k(\varepsilon, m)$ such that every graph $G = (V, E)$ on at least k nodes has an equipartition $\{V_1, \dots, V_k\}$ ($m \leq k \leq k(\varepsilon, l)$) such that for all but εk^2 pairs of indices $1 \leq i < j \leq k$, the bipartite graph $G[V_i, V_j]$ is ε-regular.*

Let us restate the Regularity Lemma in a form that is more suited for our discussions. Consider a graph $G = (V, E)$ and two subsets $U, W \subseteq V$ (not necessarily disjoint). We can measure how non-random the graph between U and W is by its *irregularity*

$$\text{irreg}_G(U, W) = \max_{X \subseteq U, Y \subseteq V} \big| e_G(X, Y) - d|X| \cdot |Y| \big|.$$

(Note that by scaling up by $|X| \cdot |Y|$, we can maximize over all sets $\subseteq U$ and $Y \subseteq W$.) Clearly $\text{irreg}_G(U, W) \leq |U| \cdot |W|$.

Lemma 2.2 (Szemerédi Regularity Lemma, second form) *For every $\varepsilon > 0$ there is a $k(\varepsilon) > 0$ such that every graph $G = (V, E)$ has an equipartition $\mathcal{P}$ into $k \leq k(\varepsilon)$ classes $V_1, \dots, V_k$ such that*

$$\sum_{1 \leq i < j \leq k} \text{irreg}_G(V_i, V_j) \leq \varepsilon |V|^2.$$

The equivalence of the two forms is easy to prove. One can add further requirements (at the cost of increasing $k(\varepsilon)$), like the requirement that $\{V_1, \dots, V_k\}$ refines a given partition.

We give one more reformulation for further reference. For $u, v \in V$, let $a_G(u, v) = 1$ if $uv \in E$ and $a_G u, v = 0$ otherwise. For a partition $\mathcal{P} = \{S_1, \dots, S_k\}$ of V and $u, v \in V$, let $a_{\mathcal{P}}(u, v) = d_G(S_i, S_j)$ where $u \in S_i$ and $v \in S_j$.

Lemma 2.3 (Szemerédi Regularity Lemma, third form) *For every $\varepsilon > 0$ there is a $k(\varepsilon) > 0$ such that every graph $G = (V, E)$ has an equipartition $\mathcal{P}$ into $k \leq k(\varepsilon)$ classes such that*

$$\Big| \sum_{uv \in E(H)} (a_G(u, v) - a_{\mathcal{P}}(u, v) \Big| \leq \varepsilon$$

for every graph H on V that is the union of at most k^2 complete bipartite graphs.

To see how this implies the previous form, let $X = X_{ij} \subseteq V_i$ and $Y = X_{ji} \subseteq V_j$ attain the maximum in the definition of $\mathrm{irreg}_G(V_i, V_j)$, and let H_{ij} be a complete bipartite graph between X_{ij} and X_{ji}. Let H be the union of those H_{ij} for which $e_G(X_{ij}, X_{ji}) > d_G(V_i, V_j)$ and let H' be the union of the rest. Applying Lemma 2.3 to both H and H', we obtain Lemma 2.2.

One feature of the Regularity Lemma, which unfortunately forbids practical applications, is that $k(\varepsilon)$ is very large: the best proof gives a tower of height about $1/\varepsilon^2$, and unfortunately this is not far from the truth, as was shown by Gowers [13].

A related result with a more reasonable threshold was proved by Frieze and Kannan [10], but they measure irregularity in a different way. For a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of V, define $d_{ij} = \frac{e_G(V_i,V_j)}{|V_i|\cdot|V_j|}$. For any two sets $S, T \subseteq V(G)$, we expect that the number of edges of G connecting S to T is about

$$e_{\mathcal{P}}(S,T) = \sum_{i=1}^{k}\sum_{j=1}^{k} d_{ij}|V_i \cap S| \cdot |V_j \cap T|.$$

So we can measure the irregularity of the partition by $\max_{S,T} |e_G(S,T) - e_{\mathcal{P}}(S,T)|$. The Weak Regularity Lemma [10] says the following.

Lemma 2.4 (Weak Regularity Lemma) *For every $\varepsilon > 0$ and every graph $G = (V, E)$, V has a partition $\mathcal{P}$ into $k \le 2^{2/\varepsilon^2}$ classes $V_1, \dots, V_k$ such that for all $S, T \subseteq V$,*

$$|e_G(S,T) - e_{\mathcal{P}}(S,T)| \le \varepsilon |V|^2.$$

Note that we do not require here that $\mathcal{P}$ is an equipartition; it is not hard to see that this version implies that we could require $\mathcal{P}$ to be an equipartition, at the cost of increasing the bound on k to $2^{c/\varepsilon^2}$ with a larger absolute constant c.

The partition in the weak lemma has substantially weaker properties than the partition in the strong lemma; these properties are sufficient in some, but not all, applications. The bound on the number of partition classes is still rather large (exponential), but at least not a tower. We'll see that the proof obtains the partition as an "overlay" of only $2/\varepsilon^2$ sets, which in some applications can be treated as if there were only about $1/\varepsilon^2$ classes, which makes the weak lemma quite efficient (see e.g. its applications in [3]). We'll come back to the sharpness of the threshold in Section 6.

Other versions of the Regularity Lemma strengthen, rather than weaken, the conclusion (of course, at the cost of replacing the tower function by an even more formidable value). Such a "super-strong" Regularity Lemma was proved by Alon, Fisher, Krivelevich and Szegedy [1, 2]. Alon and Shapira [5] used this to obtain very general results in the theory of "Property Testing" in computer science.

It turns out that the Regularity Lemma has reformulations in other branches of mathematics. A probabilistic and information theoretic version was given by Tao [21]. Our goal is to describe three reformulations in analysis.

3 The analytic language

A two-variable function $W : [0,1]^2 \to \mathbb{R}$ is called *symmetric* if $W(x,y) = W(y,x)$ for all $0 \le x, y \le 1$. Let $\mathcal{W}$ denote the set of all bounded symmetric measurable functions $W : [0,1]^2 \to \mathbb{R}$ and let $\mathcal{W}_0$ denote the set of symmetric measurable functions $W : [0,1]^2 \to [0,1]$. We call a function $U \in \mathcal{W}$ *a stepfunction with at most m steps* if there is a partition $\{S_1, \dots, S_m\}$ of $[0,1]$ such that U is constant on every $S_i \times S_j$. From the analytic point of view, we think of graphs as $0-1$ valued stepfunctions in $\mathcal{W}_0$ such that the steps S_i have equal sizes. It is clear that every such function represents a graph on the vertex set $\{S_i\}$ and every graph arises this way.

Every $W \in \mathcal{W}$ can be considered as a kernel operator on the Hilbert space $L_2([0,1]^2)$ by

$$(Wf)(x) = \int_0^1 W(x,y)f(y)\,dy.$$

Besides the standard L_2 and L_1 norms, we'll need the following norm on $\mathcal{W}$:

$$\|W\|_\square = \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} W(x,y)\,dx\,dy \right|.$$

For the case of matrices, and up to scaling, this norm is called the "cut norm"; various important properties of it were proved by Alon and Naor [4] and by Alon, Fernandez de la Vega, Kannan and Karpinski [3]. Many of these extend to the infinite case without any change. In particular, $\|W\|_\square$ is within absolute constant factors to the $L_1 \to L_\infty$ norm of W as a kernel operator. Furthermore, the following useful equations and inequalities are easy to verify:

$$\begin{aligned} \|W\|_\square &= \sup_{f,g:[0,1]\to[0,1]} |\langle f, Wg\rangle| \geq \sup_{f:[0,1]\to[0,1]} |\langle f, Wf\rangle| \\ &\geq \sup_{S\subseteq[0,1]} \left| \int_{S\times S} W(x,y)\,dx\,dy \right| \geq \frac{1}{2}\|W\|_\square. \end{aligned} \tag{1}$$

The Weak Regularity Lemma in these terms asserts the following:

Lemma 3.1 (Weak Regularity Lemma, Analytic Form) *For every function $W \in \mathcal{W}_0$ and $\varepsilon > 0$ there is a stepfunction $W' \in \mathcal{W}_0$ with at most $\lceil 2^{2/\varepsilon^2}\rceil$ steps such that $\|W - W'\|_\square \leq \varepsilon$.*

For every $W \in \mathcal{W}$ and every partition $\mathcal{P} = \{P_1, \dots, P_k\}$ of $[0,1]$ into measurable sets, let $W_\mathcal{P} : [0,1]^2 \to \mathbb{R}$ denote the stepfunction obtained from W by replacing its value at $(x,y) \in P_i \times P_j$ by the average of W over $P_i \times P_j$. (This is not defined when $\lambda(P_i) = 0$ or $\lambda(P_j) = 0$, but this is of measure 0; here λ denotes the Lebesgue measure.)

It was observed in [3] that we can replace the stepfunction W' in Lemma 3.1 by the stepfunction $W_\mathcal{P}$, where $\mathcal{P}$ is the partition into the steps of W', at the cost of increasing the error ε by a factor of at most 2. Furthermore, at the cost of replacing the bound $2^{\lceil 2/\varepsilon^2\rceil}$ on the number of steps by $2^{\lceil 20/\varepsilon^2\rceil}$, we could require that the steps have the same measure.

It can also be shown [8] that finite simple graphs (0-1 valued symmetric stepfunctions) are dense in the set $\mathcal{W}_0$ with respect to the $\|.\|_\square$-norm.

The norm $\|.\|_\square$ relates to other norms by the following inequalities. It is trivial that

$$\|W\|_\square \leq \|W\|_1. \tag{2}$$

The following inequalities, proved in [8], are still simple but less obvious. For every $W \in \mathcal{W}$, let $W \circ W$ denote its square as a kernel operator, i.e.,

$$(W \circ W)(x,y) = \int_0^1 W(x,t)W(t,y)\,dt.$$

Then

$$\|W\|_\square^4 \leq \|W \circ W\|_2^2 \leq \|W\|_\square \|W\|_\infty^2 \|W\|_1. \tag{3}$$

So for functions in $\mathcal{W}$,

$$\|W \circ W\|_2^{1/2} \leq \|W\|_\square \leq \|W \circ W\|_2^2.$$

It can be checked that the left hand side, as a function of W, is a norm. Due to its more explicit form, this is often easier to handle than $\|W\|_\square$.

We conclude this section with formulating an analytic version of the strong Szemerédi Lemma (third version). A *rectangle* in $[0,1]$ is any set of the form $S\times T$, where S and T are measurable subsets of $[0,1]$.

Lemma 3.2 (Strong Regularity Lemma, Analytic Form) *For every $\varepsilon>0$ there is an integer $k(\varepsilon)>0$ such that for every function $W\in\mathcal{W}_0$ there is a partition $\mathcal{P}=\{S_1,\dots,S_k\}$ of $[0,1]$ into $k\le k(\varepsilon)$ sets of equal measure with the following property: For every set $R\subseteq[0,1]^2$ that is the union of at most k^2 rectangles, we have*

$$\Big|\int_R (W-W_{\mathcal{P}})\,dx\,dy\Big|\le\varepsilon.$$

4 The Regularity Lemma in Hilbert space

The following lemma is an extension of the Regularity lemma to a very general setting of Hilbert spaces.

Lemma 4.1 (Regularity Lemma in Hilbert Space) *Let $\mathcal{K}_1,\mathcal{K}_2,\dots$ be arbitrary nonempty subsets of a Hilbert space $\mathcal{H}$. Then for every $\varepsilon>0$ and $f\in\mathcal{H}$ there is an $m\le\lceil 1/\varepsilon^2\rceil$ and there are $f_i\in\mathcal{K}_i$ $(1\le i\le m)$ and $\gamma_1,\gamma_2,\dots,\gamma_m\in\mathbb{R}$ such that for every $g\in\mathcal{K}_{m+1}$*

$$|\langle g, f-(\gamma_1 f_1+\dots+\gamma_m f_m)\rangle|\le\varepsilon\cdot\|g\|\cdot\|f\|.$$

Before proving this lemma, a little discussion is in order. Assume that the sets $\mathcal{K}_n$ are subspaces. Then a natural choice for the function $\gamma_1 f_1+\dots+\gamma_m f_m$ is the best approximation of f in the subspace $\mathcal{K}_1+\dots+\mathcal{K}_m$ (or an approximately best approximation, if the best does not exist), and the error $f-(\gamma_1 f_1+\dots+\gamma_m f_m)$ is orthogonal (or almost orthogonal) to every $g\in\mathcal{K}_1+\dots+\mathcal{K}_m$. The main point in this lemma is that it is also almost orthogonal to the *next* set K_{m+1}.

Proof. Let

$$\eta_k=\inf_{\{\gamma_i\},\{f_i\}}\|f-\sum_{i=1}^k\gamma_i f_i\|^2,$$

where the infimum is taken over all $\gamma_1,\dots,\gamma_k\in\mathbb{R}$ and $f_i\in\mathcal{K}_i$. Clearly we have $\|f\|^2\ge\eta_1\ge\eta_2\ge\dots\ge0$. Hence there is an $m\le\lceil 1/\varepsilon^2\rceil$ such that $\eta_m<\eta_{m+1}+\varepsilon^2\|f\|^2$. So there are $\gamma_1,\dots,\gamma_m\in\mathbb{R}$ and $f_i\in\mathcal{K}_i$ such that

$$\|f-\sum_{i=1}^m\gamma_i f_i\|^2\le\eta_{m+1}+\varepsilon^2\|f\|^2.$$

Let $f^*=\sum_i\gamma_i f_i$, and consider any $g\in\mathcal{K}_{m+1}$. By the definition of η_{m+1}, we have for every real t that

$$\|f-(f^*+tg)\|^2\ge\eta_{m+1}\ge\|f-f^*\|^2-\varepsilon^2\|f\|^2,$$

or

$$\|g\|^2t^2-2\langle g,f-f^*\rangle t+\varepsilon^2\|f\|^2\ge0.$$

The discriminant of this quadratic polynomial must be nonpositive, which proves the lemma. □

We derive some consequences of this Lemma. First, let us apply this lemma to the case when the Hilbert space is $L^2([0,1]^2)$, and each $\mathcal{K}_n$ is the set of indicator functions of product sets $S\times S$, where S is a measurable subset of $[0,1]$. Let $f\in\mathcal{W}_0$, then $f^*=\sum_{i=1}^k\gamma_i f_i$ is a stepfunction with at most 2^k steps, and so we get a stepfunction $W^* \in \mathcal{W}$ with at most $2^{\lceil 1/\varepsilon^2 \rceil}$ steps such that for every measurable set $S \subseteq [0,1]$,

$$\left| \int_{S\times S} (W - W^*) \right| \le \varepsilon.$$

It is easy to see that the conclusion implies that for any two measurable sets $S, T \subseteq [0,1]$,

$$\left| \int_{S\times T} (W - W^*) \right| \le 2\varepsilon,$$

which implies Lemma 3.1 (up to the factor of 2).

We say that a partition $\mathcal{P}$ of $[0,1]$ is a *weak Szemerédi partition* for W with error ε, if

$$\left| \int_{S\times S} (W - W_{\mathcal{P}}) \right| \le \varepsilon$$

holds for every subset $S \subseteq [0,1]$. So every function has a weak Szemerédi partition with error ε, with at most $2^{2/\varepsilon^2}$ classes.

To derive the graph theoretic form of the (weak) Regularity Lemma from lemma 4.1, we represent the graph G on n nodes by a stepfunction W_G: we consider the adjacency matrix $A = (a_{ij})$ of G, and replace each entry a_{ij} by a square of size $(1/n) \times (1/n)$ with the constant function a_{ij} on this square. Let $\mathcal{A}$ be the algebra of subsets of $[0,1]$ generated by the intervals corresponding to nodes of G. We let $\mathcal{K}_n$ be the set of indicator functions of product sets $S \times S$ ($S \in \mathcal{A}$). Analogously to the proof of Lemma 3.1 above, we get a partition $\mathcal{P} = \{S_1, \dots, S_m\}$ of $[0,1]$ into sets in $\mathcal{A}$ such that

$$\left| \int_{S\times T} (W_G(x,y) - (W_G)_{\mathcal{P}}(x,y))\, dx\, dy \right| \le 2\varepsilon$$

for all sets $S, T \in \mathcal{A}$. This translates into the conclusion of Lemma 2.4.

Next we show how to get the strong analytic form of the Regularity Lemma 3.2; the graph theoretic form can be obtained similarly (just there is a little extra trouble because of divisibilities). Let us define a sequence $s(1), s(2), \dots$ of positive integers by $s(1) = 1$ and $s(k+1) = 2^{s(1)^4 \cdots s(k)^4}$. Let us apply Lemma 4.1 to the Hilbert space $L_2([0,1]^2)$ and the function W as before, but choose $\mathcal{K}_n$ to be the set of stepfunctions with at most $s(n)$ steps. Lemma 4.1 gives us a function W^*, which is a stepfunction with at most $m = s(1)s(2)\dots s(k)$ steps; let $S_1, \dots S_m$ be these steps. This stepfunction has the property that for every stepfunction U with at most $s(k+1)$ steps,

$$\Big| \int_{[0,1]^2} U(W - W^*)\, dx\, dy \le \varepsilon.$$

We further partition each S_i into an appropriate number of sets of measure $1/m^2$ (called *good* sets and a "remainder" of measure less than $1/m^2$. We combine these remainders into single set, whose measure is less than $1/m$. We partition this into sets of size $1/m^2$; there will be at most m such sets, which we call *bad* sets. So we get a partition $\mathcal{Q} = \{T_1, \dots, T_{m^2}\}$ of $[0,1]$ into m^2 equal parts, out of which (say) $T_1, \dots, T_{m^2-m}$ are good sets.

Let $R \subseteq [0,1]^2$ be a set that is the union of m^2 rectangles. We claim that

$$\Big| \int_R (W - W_{\mathcal{Q}}) \Big| \le 3\varepsilon.$$

We start with removing from R all points in sets $T_i \times T_j$, where either T_i or T_j is bad. The remaining set R' is again the union of at most m^2 rectangles, and since the measure of $R \setminus R'$ is less than $2/m < \varepsilon$, it suffices to prove that

$$\Big| \int_{R'} (W - W_{\mathcal{Q}}) \Big| \le 2\varepsilon.$$

Clearly, the indicator function of R' is a stepfunction with at most $2^{m^4} \le s(k+1)$ steps, and hence by the conclusion of Lemma 4.1, we have

$$\Big|\int_{R'} (W - W^*)\Big| \le \varepsilon.$$

So it suffices to verify that

$$\Big|\int_{R'} (W_{\mathcal{Q}} - W^*)\Big| \le \varepsilon. \tag{4}$$

If T_i and T_j are good sets, then both W^* and $W_{\mathcal{Q}}$ are constant on $T_i \times T_j$, so we can either include or exclude the rectangle $T_i \times T_j$ from R', and not decrease the left hand side of (4). Doing so for every pair of good sets, we obtain a set R'', which is the union of certain sets $T_i \times T_j$, where both T_i and T_j are good. Thus by the definition of $W_{\mathcal{Q}}$, we have

$$\Big|\int_{R''} (W_{\mathcal{Q}} - W^*)\Big| = \Big|\int_{R''} (W - W^*)\Big| \le \varepsilon$$

(by the assertion of Lemma 4.1). This concludes the proof of the strong Szemerédi Lemma.

There may be further interesting choices of the Hilbert space $\mathcal{H}$ and subsets $\mathcal{K}_n$. For example, let $\mathcal{H} = L_2[0,1]$, and let $\mathcal{K}_n$ be the set of polynomials of degree at most 2^n. Then we get:

Corollary 4.2 *For every $\varepsilon > 0$ and every function $f \in L_2[0,1]$ there is a polynomial $p \in \mathcal{R}[x]$ of degree $d \le 2^{\lceil 1/\varepsilon^2 \rceil}$ such that*

$$\langle g, f - p\rangle \le \varepsilon \|f\| \cdot \|g\|$$

for every polynomial g of degree at most $2d$.

5 The Regularity Lemma as compactness

In this chapter we show that the Regularity Lemma can be formulated as a compactness theorem.

Recall that a map $\phi:\ [0,1] \to [0,1]$ is *measure preserving* if $\lambda(\phi^{-1}(U)) = \lambda(U)$ for every measurable set $U \subseteq [0,1]$. We say that ϕ is a *measure preserving bijection* if it is bijective and its inverse is also measure preserving.

Let W be a function from $\mathcal{W}$. We define W^ϕ by $W^\phi(x,y) = W(\phi(x),\phi(y))$. We define a "distance" on the space $\mathcal{W}$ by

$$\delta_\square(U,W) = \inf_\phi \|U^\phi - W\|_\square,$$

where ϕ ranges over all measure preserving bijections $[0,1] \to [0,1]$. It is not hard to check that $\delta_\square(U,W) = \delta_\square(W,U)$, and that this distance satisfies the triangle inequality. Furthermore, $\delta_\square(U,W) = \delta_\square(U^\phi,W)$ for every measure preserving bijection ϕ.

The distance of two different functions can be 0; various characterizations of when the $\delta_\square$ distance is 0 are given in [7] and [18].

We construct a metric space $\mathcal{X}$ from $(\mathcal{W},\delta_\square)$ by identifying functions U and W with $\delta_\square(U,W) = 0$. Let $\mathcal{X}_0$ denote the image of $\mathcal{W}_0$ under this identification. Informally speaking, the elements of $\mathcal{X}_0$ are the isomorphism classes of functions in $\mathcal{W}_0$. Clearly the distance $\delta_\square$ is well defined on $\mathcal{X}_0$.

The following fact can be regarded as a topological interpretation of the Regularity Lemma. (We prove it by the methods in [17].)

Theorem 5.1 *The metric space $\mathcal{X}_0$ is compact.*

Proof. Let $W_1, W_2, \dots$ be a sequence of functions in $\mathcal{W}_0$. We want to construct a subsequence that has a limit in $\mathcal{X}_0$.

Using Lemma 3.1 and the remarks after it, for each k and n we construct a partition $\mathcal{P}_{n,k}$ such that these partitions and the corresponding stepfunctions $W_{n,k} = W_{\mathcal{P}_{n,k}} \in \mathcal{W}_0$ satisfy the following.

- $\|W_n - W_{n,k}\|_\square \le 1/k$.
- $|\mathcal{P}_{n,k}| = m_k$ (where m_k depends only on k).
- The partition $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$ for every k.

We'll only use that $\delta_\square(W_n, W_{n,k}) \le 1/k$, which means that we can rearrange the range of $W_{n,k}$ as we wish; in particular, we may assume that all steps are intervals.

Now we can select a subsequence of the W_n for which the length of the i-th interval of $W_{n,1}$ converges for every i, and also the value of $W_{n,1}$ on the product of the i-th and j-th intervals converges for every i and j (as $n \to \infty$). It follows then that the sequence $W_{n,1}$ converges to a limit U_1 almost everywhere, which itself is a stepfunction with m_1 steps that are intervals.

We repeat this for $k = 2, 3, \dots$, to get subsequences for which $W_{k,n} \to U_k$ almost everywhere, where U_k is a stepfunction with m_k steps that are intervals.

For every $k < l$, the partition into the steps of $W_{n,l}$ is a refinement of the partition into the steps of $W_{n,k}$, and hence it is easy to see that the same relation holds for the partitions into the steps of U_l and U_k. Furthermore, the function $W_{n,k}$ can be obtained from the function $W_{n,l}$ by averaging its value over each step, and it follows that a similar relation holds for U_l and U_k.

Let (X, Y) be a random point in $[0,1]^2$ chosen uniformly, then this property of the functions U_k implies that the sequence $(U_1(X,Y), U_2(X,Y), \dots)$ is a martingale. Since the random variables $U_i(X,Y)$ remain bounded, the Martingale Convergence Theorem (see e.g. [22], Theorem 11.5) implies that this martingale is convergent with probability 1. In other words, the sequence $(U_1, U_2, \dots)$ is convergent almost everywhere. Let U be its limit.

Fix any $\varepsilon > 0$. Then there is a $k > 3/\varepsilon$ such that $\|U - U_k\|_1 < \varepsilon/3$. Fixing this k, there is an n_0 such that $\|U_k - W_{n,k}\|_1 < \varepsilon/3$ for all $n \ge n_0$. Then

$$\begin{aligned}
\delta_\square(U, W_n) &\le \delta_\square(U, U_k) + \delta_\square(U_k, W_{n,k}) + \delta_\square(W_{n,k}, W_n)\\
&\le \|U - U_k\|_1 + \|U_k - W_{n,k}\|_1 + \delta_\square(W_{n,k}, W_n)\\
&\le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.
\end{aligned}$$

This proves that $W_n \to U$ in the metric space $\mathcal{X}_0$. □

Note that in the proof above, the explicit bound on the number of partition classes in the Regularity Lemma was not used, only that their number is bounded by a function of ε, independent of the function. This is quite often the case with applications of the Lemma.

Now we show how this compactness statement implies the following strong form of the Regularity Lemma.

Lemma 5.2 (Very Strong Regularity Lemma) *Let $h(\varepsilon, t) > 0$, $(\varepsilon > 0, t \in \mathbb{N})$ be an arbitrary fixed function. Then for every $\varepsilon > 0$ there is a threshold $k(\varepsilon)$ such that for every function $W \in \mathcal{W}_0$ there are two functions $W', U \in \mathcal{W}_0$ such that U is a stepfunction with $l \le k(\varepsilon)$ steps, and*

$$\|W - W'\|_\square \le h(\varepsilon, l), \qquad \|W' - U\|_1 \le \varepsilon.$$

The role of the two norms could be interchanged: the function $W'' = U - W' + W$ satisfies

$$\|W - W''\|_1 \le \varepsilon, \qquad \|W'' - U\|_\square \le h(\varepsilon, 1).$$

Choosing $h(\varepsilon, t) = \varepsilon$, we get Lemma 3.1. Choosing $h(\varepsilon, t) = \varepsilon/t^2$, it is not hard to see that the strong form of Szemerédi's Lemma follows. Choosing $h(\varepsilon, t)$ appropriately small, we get the "super-strong" Regularity Lemma from [1, 2] mentioned in the introduction. Our Lemma is very closely related to a version of the regularity lemma given by Tao [21].

Proof. We may assume that h is monotone decreasing in its second variable. Let us fix a number $\varepsilon > 0$. Every function $U \in \mathcal{W}_0$ is the limit of stepfunctions in the $\|.\|_1$ norm, hence there is a stepfunction $U' \in \mathcal{W}_0$ with $\|U - U'\|_1 \le \varepsilon$. Let $f(U)$ denote the minimum number of steps in such a stepfunction U'. For a function $U \in \mathcal{W}_0$, let $B(U)$ denote the open ball $\{W \mid \delta_\square(U, W) < h(\varepsilon, f(U))\}$. Using Theorem 5.1, we obtain that there is a finite set of functions $W_1, W_2, \dots, W_t \in \mathcal{W}_0$ with $\cup_{i=1}^t B(W_i) = \mathcal{W}_0$. This means that for every function $W \in \mathcal{W}_0$ there is a function W_m $(1 \le m \le t)$ and a stepfunction $U_0 \in \mathcal{W}_0$ with $f(W_m)$ steps such that $\delta_\square(W, W_m) < h(\varepsilon, f(W_m))$ and $\|W_m - U_0\|_1 < \varepsilon$.

Set $l = f(W_m)$ and $k(\varepsilon) = \max_{i=1}^t f(W_i)$. There is a measure preserving bijection $\phi : [0,1] \mapsto [0,1]$ such that $\|W - W_m^\phi\|_\square < h(\varepsilon, l)$. Then $U = U_0^\phi$ is a stepfunction with l steps, and $W' = W_m^\phi$ satisfies

$$\|W' - U\|_1 = \|W_m^\phi - U_0^\phi\|_1 = \|W_m - U_0\|_1 < \varepsilon$$

and

$$\delta_\square(W, W') = \delta_\square(W, W_m^\phi) = \delta_\square(W, W_m) \le h(\varepsilon, l),$$

which completes the proof. □

6 The Regularity Lemma and covering by small balls

Every function $W \in \mathcal{W}$ gives rise to a metric on $[0,1]$ by

$$d_W^1(x_1, x_2) = \|W(x_1, .) - W(x_2, .)\|_2 = \left(\int_0^1 (W(x_1, y) - W(x_2, y))^2 \, dy \right)^{1/2}.$$

It turns out that for our purposes, the following distance function is more important: we square W as a kernel operator, and then consider the above distance. More precisely, we define

$$\begin{aligned} d_W(x_1, x_2) &= d^1_{W \circ W}(x_1, x_2) \\ &= \left(\int_0^1 \left(\int_0^1 W(x_1, y) W(y, z) \, dy - \int_0^1 W(x_2, y) W(y, z) \, dy \right)^2 dz \right)^{1/2}. \end{aligned}$$

Our goal is to prove that the (weak) Regularity Lemma is equivalent to the assertion that most of the metric space $([0,1], \delta_W)$ can be covered by a bounded number of small balls. More exactly:

Theorem 6.1 *Let $W \in \mathcal{W}_0$ and let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a partition of $[0,1]$ into measurable sets.*

(a) *If $\mathcal{P}$ is a weak Szemerédi partition with error $\varepsilon^2/8$, then there is a set $S \subseteq [0,1]$ with $\lambda(S) \le \varepsilon$ such that for each partition class, $P_i \setminus S$ has diameter at most ε in the d_W metric.*

(b) *If there is a set $S \subseteq [0,1]$ with $\lambda(S) \le (\varepsilon/5)^4$ such that for each partition class, $P_i \setminus S$ has diameter at most $(\varepsilon/5)^2$ in the d_W metric, then $\mathcal{P}$ is a weak Szemerédi partition with error ε.*

Combining this fact with the existence of weak Szemerédi partitions, we get the following:

Corollary 6.2 *For every function $W \in \mathcal{W}$ and every $\varepsilon > 0$ there is a partition $\mathcal{P} = \{P_0, P_1, \dots, P_k\}$ of $[0,1]$ into measurable sets with $k \le 2^{\lceil 64/\varepsilon^4 \rceil}$ such that $\lambda(P_0) \le \varepsilon$ and for $1 \le i \le k$, P_i has diameter at most ε in the d_W metric.*

It is straightforward to formulate this theorem for graphs instead of functions $W \in \mathcal{W}$: We define the distance of two nodes u, v of a graph G by squaring the adjacency matrix, and taking the euclidean distance between the row vectors corresponding to u and v, divided by $n^{3/2}$. Then the statement of the Theorem is analogous, and the proof is the same.

Proof. (a) Suppose that $\mathcal{P}$ is a weak Szemerédi partition with error $\varepsilon^2/8$. Let $R = W - W_{\mathcal{P}}$, then we know that $\|R\|_\square \le \varepsilon^2/8$.

For every $x \in [0,1]$, define

$$F(x) = \int_0^1 \Big(\int_0^1 R(x,s)W(s,z)\,ds \Big)^2 \,dz.$$

Then we have

$$\int_0^1 F(x)\,dx = \int_0^1 \int_0^1 \int_0^1 \int_0^1 R(x,t)R(x,s)W(s,z)W(t,z)\,dx\,ds\,dt\,dz.$$

Fix z and t, then since $-1 \le R(x,t) \le 1$ and $0 \le W(s,z) \le 1$, we have

$$\int_0^1 \int_0^1 R(x,t)R(x,s)W(s,z)\,dx\,ds \le \varepsilon^2/4,$$

and so

$$\int_0^1 F(x)\,dx \le \varepsilon^2/4.$$

Hence there is a set $S \subseteq [0,1]$ with measure at most ε such that for $x \in [0,1] \setminus S$, we have $F(x) \le \varepsilon/4$.

Let $x, y \in [0,1] \setminus S$ be two points in the same partition class of $\mathcal{P}$. Then $W_{\mathcal{P}}(x,s) = W_{\mathcal{P}}(y,s)$ for every $s \in [0,1]$, and hence

$$\begin{aligned}
d_W(x,y)^2 &= \int_0^1 \Big(\int_0^1 (W(x,s) - W(y,s))W(s,z)\,ds \Big)^2 \,dz \\
&= \int_0^1 \Big(\int_0^1 (R(x,s) - R(y,s))W(s,z)\,ds \Big)^2 \,dz \\
&= \int_0^1 \Big(\int_0^1 R(x,s)W(s,z)\,ds - \int_0^1 R(y,s))W(s,z)\,ds \Big)^2 \,dz \\
&\le 2\int_0^1 \Big(\int_0^1 R(x,s)W(s,z)\,ds \Big)^2 \,dz + 2\int_0^1 \Big(\int_0^1 R(y,s)W(s,z)\,ds \Big)^2 \,dz \\
&= 2F(x) + 2F(y) \le \varepsilon.
\end{aligned}$$

(b) We want to show that $\|W - W_{\mathcal{P}}\|_\square < \varepsilon$. By (1), it suffices to show that for any 0-1 valued function f,

$$\langle f, (W - W_{\mathcal{P}})f \rangle \le \frac{1}{2}\varepsilon. \tag{5}$$

Let us write $f = f_{\mathcal{P}} + g$, where $f_{\mathcal{P}}(x)$ is obtained by replacing $f(x)$ by the average of f over the class P_i containing x. It is easy to check that we have

$$\langle f, (W - W_{\mathcal{P}})f \rangle = \langle f + f_{\mathcal{P}}, Wg \rangle. \tag{6}$$

By Cauchy-Schwartz,

$$\langle f + f_{\mathcal{P}}, Wg \rangle \le \|f + f_{\mathcal{P}}\| \cdot \|Wg\| \le 2\|Wg\|. \tag{7}$$

We have

$$\|Wg\|^2 = \langle g, W^2 g\rangle = \int_{[0,1]^3} g(x)W(x,y)W(y,z)g(z)\,dx\,dy\,dz.$$

For each x, let $\phi(x)$ be an arbitrary, but fixed, element of the class P_i containing x such that $x \notin S$ (if $P_i \subseteq S$ then we define $\phi(x)$ to be 0). Then

$$\begin{aligned}\int_{[0,1]^3} & g(x)W(x,y)W(y,z)g(z)\,dx\,dy\,dz \\ &= \int_{[0,1]^3} g(x)\big(W(x,y)W(y,z) - W(x,y)W(y,\phi(z))\big)g(z)\,dx\,dy\,dz \\ &\quad + \int_{[0,1]^3} g(x)W(x,y)W(y,\phi(z))g(z)\,dx\,dy\,dz\,.\end{aligned}$$

Here the last integral is 0, since the integral of g over each partition class is 0. Furthermore,

$$\begin{aligned}\int_{[0,1]^3} & g(x)\big(W(x,y)W(y,z) - W(x,y)W(y,\phi(z))\big)g(z)\,dx\,dy\,dz \\ &\le \left(\int_{[0,1]^3} g(x)^2 g(z)^2\,dx\,dy\,dz\right)^{1/2} \\ &\quad \times \left(\int_{[0,1]^3} \big(W(x,y)W(y,z) - W(x,y)W(y,\phi(z))\big)^2\,dx\,dy\,dz\right)^{1/2}.\end{aligned}$$

Here the first factor is at most 1, and

$$\begin{aligned}\int_{[0,1]^3} \big(W(x,y)W(y,z) - W(x,y)W(y,\phi(z))\big)^2\,dx\,dy\,dz &= \int_0^1 d_W(z,\phi(z))^2\,dz \\ = \int_{[0,1]/S} d_W(z,\phi(z))^2\,dz + \int_S d_W(z,\phi(z))^2\,dz &\le 2(\varepsilon/5)^4.\end{aligned}$$

Thus

$$\int_{[0,1]^3} g(x)W(x,y)W(y,z)g(z)\,dx\,dy\,dz \le 2^{1/2}(\varepsilon/5)^2,$$

and so $\|Wg\| \le \frac{1}{5}2^{1/4}\varepsilon < \frac{1}{4}\varepsilon$. By (6) and (7), this proves (5), and completes the proof. □

7 Two applications

We conclude with two applications of this characterization of weak Szemerédi partitions. First we prove that the exponential dependence of the number of classes on ε in Lemma 3.1 is necessary. Taking an appropriately dense finite subgraph of our construction, one can prove a similar bound on the threshold in the finite version Lemma 2.4.

Let S^d be the d-dimensional sphere, endowed with the uniform probability measure μ (it does not matter which probability space we consider as the domain of W, so we may consider S^d instead of $[0,1]$). For $x, y \in S^d$, let

$$W(x,y) = \begin{cases} 1, & \text{if } x^{\mathsf{T}}y \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

We prove:

Proposition 7.1 *Every weak Szemerédi partition of W with error $\varepsilon \le 1/(8d+8)$ contains at least 2^{d-1} classes.*

Proof. We may assume that $d \ge 3$. Let $\sphericalangle(x,y)$ denote the angle (spherical distance) between the points $x, y \in S^d$. Then clearly

$$W^{(2)}(x,y) = \frac{1}{2} - \frac{\sphericalangle(x,y)}{\pi}.$$

From this it is routine to verify that for any two points $x, y \in S^d$,

$$d_W(x,y) \ge \frac{2}{\sqrt{d+1}} \sphericalangle(x,y). \tag{8}$$

Let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a weak Szemerédi partition of Ω for the function W, with error ε. Then by Theorem 6.1(a), there is a set $T \subseteq S^d$ with $\lambda(T) \le (8\varepsilon)^{1/2}$ such that the diameter of $P_i \setminus T$ in the d_W metric is at most $(8\varepsilon)^{1/2}$ for every i. By (8), this implies that the diameter of $P_i \setminus T$ in spherical distance is at most $\sqrt{2\varepsilon(d+1)}$, and hence its measure satisfies $\lambda(P_i \setminus T) \le (\sqrt{2\varepsilon(d+1)})^d \le 2^{-d}$. Since the sets $P_i \setminus T$ $(i = 1, \dots, k)$ and T cover S^d, we get

$$k2^{-d} + (8\varepsilon)^{1/2} \ge 1,$$

which implies that $k \ge 2^{d-1}$. □

Thus for a given $\varepsilon > 0$, we get weak Szemerédi partitions with error at most ε with $2^{2/\varepsilon^2}$ classes, and for some functions we need at least $(1/4)2^{1/(8\varepsilon)}$ classes. It is not clear whether the best threshold has $1/\varepsilon$ or $1/\varepsilon^2$ in the exponent.

As a second application of Theorem 6.1, we sketch a (somewhat surprising) algorithm to construct a weak Szemerédi partition.

Let $W \in \mathcal{W}_0$ and $\varepsilon^2/5 > 0$. Set

$$m = \left\lceil \frac{80}{\varepsilon^2} \ln \frac{80}{\varepsilon^2} \right\rceil 2^{\lceil 10^{12}/\varepsilon^{16} \rceil}.$$

Choose independent uniform random points $X_1, \dots, X_m$ from $[0,1]$. Let $S_1, \dots, S_m$ be the Voronoi cells of these points with respect to the metric δ_W; in other words, let $x \in S_i$ if x is closer to X_i than to any other X_j; if there are more than one points X_j at minimum distance from x, then we assign x to that with smallest subscript. This way get a partition $\mathcal{S}(X_1, \dots, X_m) = \{S_1, \dots, S_m\}$ of $[0,1]$.

Theorem 7.2 *With probability at least $3/4$, the partition $\mathcal{S}(X_1, \dots, X_m)$ is a weak Szemerédi partition with error at most ε.*

We have described this algorithm as applied to a function $W \in \mathcal{W}_0$, but it is straightforward to modify it so that it applies to a graph G. Our algorithm gives a larger number of classes than that of Frieze and Kannan [10], and it is also slower (primarily because of the cost of squaring the adjacency matrix at the beginning). Our purpose with this formulation is to illuminate this geometric connection.

Proof. Let $k = 2^{\lceil 10^{12}/\varepsilon^{16} \rceil}$. By Corollary 6.2, there is a partition $\{T_0, T_1, \dots, T_k\}$ of $[0,1]$ into $k+1$ measurable sets such that $\lambda(T_0) \le \varepsilon^2/10$ and for $1 \le i \le k$, T_i has diameter at most $\varepsilon^2/10$ in the d_W metric. let $\alpha_i = \lambda(T_i)$. Let I be the set of those indices $i \in \{1, \dots, k\}$ for which T_i contains at least one sample point X_j (we don't care whether T_0 contains a sample point). Then

$$\mathsf{E}\big(\lambda(\cup_{i \notin I} T_i)\big) = \sum_{i=1}^{k} \alpha_i (1-\alpha_i)^m.$$

To estimate this sum, let $c = \varepsilon^2/(80k)$. Then we have

$$\mathsf{E}\big(\lambda(\cup_{i\notin I} T_i)\big) = \sum_{i:\ \alpha_i \le c} \alpha_i(1-\alpha_i)^m + \sum_{i:\ \alpha_i > c} \alpha_i(1-\alpha_i)^m \le ck + (1-c)^m$$
$$\le \frac{\varepsilon^2}{80} + e^{-\varepsilon^2 m/(80k)} \le \frac{\varepsilon^2}{80} + \frac{\varepsilon^2}{80} = \frac{\varepsilon^2}{40}.$$

So with probability at least 3/4, we have $\lambda(\cup_{i\notin I} T_i) \le \varepsilon^2/10$. In such a case, the set $S = \cup_{i\in I} T_i \cup T_0$ has measure $\lambda(S) \le \varepsilon^2/5$.

We claim that for $j = 1, \dots, m$, the diameter of $S_j \setminus S$ is at most $\varepsilon^2/5$. It suffices to prove that $d_W(x, X_j) \le \varepsilon^2/10$ for every point $x \in S_j \setminus S$. Indeed, there is an $i \in I$ such that $x \in T_i$. The set T_i has diameter at most $\varepsilon^2/10$, and (since $i \in I$) there is a sample point $X_h \in T_i$. Thus $d_W(x, X_h) \le \varepsilon^2/10$, but since x belongs to the Voronoi cell of X_j, it follows that $d_W(x, X_j) \le \varepsilon^2/10$.

We are done by Theorem 6.1(b). □

Acknowledgement

We are grateful to Christian Borgs and the anonymous referee for pointing out many minor errors and inconsistencies in earlier versions of this paper.

References

[1] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy: Efficient testing of large graphs, *Proc. 40th Ann. Symp. on Found. of Comp. Sci.,* IEEE (1999), 656–666.

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

[3] 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.

[4] N. Alon and A. Naor: Approximating the Cut-Norm via Grothendieck's Inequality, *Proc. of the 36 ACM STOC* ACM Press (2004), 72–80.

[5] N. Alon and A. Shapira, Every monotone graph property is testable, *Proc. of the 37 ACM STOC*, Baltimore, ACM Press (2005), 128-137.

[6] N. Alon and A. Shapira, A characterization of the (natural) graph properties testable with one-sided error, *Proc. of the 46 IEEE FOCS*, IEEE (2005), 429-438.

[7] C. Borgs, J. Chayes, L. Lovász: Unique limits of dense graph sequences (manuscript)

[8] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi: Convergent sequences of dense graphs I: Subgraph Frequencies, Metric Properties and Testing (manuscript)

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

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

[11] S. Gerke, A. Steger: Stefanie Gerke, Angelika Steger: The sparse regularity lemma and its applications, in: *Surveys in Combinatorics* (ed. B.S.Webb), Cambridge Univ. Press (2005), 227–258.

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

[13] W.T. Gowers: Lower bounds of tower type for Szemerédi's Uniformity Lemma, *Geom. Func. Anal.* **7** (1997), 322–337.

[14] W.T. Gowers: Hypergraph regularity and the multidimensional Szemeredi theorem (manusript)

[15] Y. Kohayakawa, V. Rödl: Szemerd́i's regularity lemma and quasi-randomness, in: *Recent Advances in Algorithms and Combinatorics,* CMS Books Math./Ouvrages Math. SMC **11**, Springer, New York (2003), 289–351.

[16] J. Komlós and M. Simonovits: Szemerédi's Regularity Lemma and its applications in graph theory, in: *Combinatorics, Paul Erdos is Eighty* (D. Miklos et. al, eds.), Bolyai Society Mathematical Studies **2** (1996), pp. 295–352.

[17] L. Lovász and B. Szegedy: Limits of dense graph sequences (MSR Tech Report)
`ftp://ftp.research.microsoft.com/pub/tr/TR-2004-79.pdf`

[18] L. Lovász and B. Szegedy: Testing properties of graphs and functions (manuscript)

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

[20] T.C. Tao: A variant of the hypergraph removal lemma (preprint)
`http://arxiv.org/abs/math.CO/0503572`

[21] T.C. Tao: Szemerédi's regularity lemma revisited (preprint)
`http://arxiv.org/abs/math.CO/0504472`

[22] D. Williams: *Probability with Martingales*, Cambridge Univ. Press, 1991.