Chapter 6

Szemerédi's Regularity Lemma

To ask the right question is harder than to answer it.
Georg Cantor

In Chapter 2 we saw that in any finite colouring of the positive integers, some colour class must contain arbitrarily long arithmetic progressions. In 1936, Erdős and Turán conjectured that the colouring here is just a distraction; the same conclusion should hold for the 'largest' colour class.

6.1 The Erdős-Turán Conjecture

Recall that the upper density of a set $A \subset \mathbb{N}$ is

$$\overline{d}(A) \;=\; \limsup_{n\to\infty} \frac{|A \cap [n]|}{n}.$$

Conjecture 6.1.1 (Erdős and Turán, 1936)**.** *If* $\overline{d}(A) > 0$*, then* A *contains arbitrarily long arithmetic progressions.*

Roth proved the case $k = 3$ of the conjecture in 1953, and Szemerédi proved the case $k = 4$ in 1969, before finally resolving the question in 1975. As mentioned in Section 6.3.4, Furstenburg (in 1977) and Gowers (in 2001) provided important alternative proofs of Szemerédi's Theorem. (According to Terry Tao, there are now at least 16 different proofs known!)

Although the conjecture of Erdős and Turán may look like little more than a curiosity, it is perhaps the greatest triumph of the 'Hungarian school' of mathematics, in which the aim is to pose (and solve) beautiful and difficult problems, and allow deep theories and connections between areas to present themselves. For more on the modern (and very active) study of Additive Combinatorics, which grew out of the proofs of Roth, Szemerédi, Furstenburg and Gowers, see the excellent recent book by Tao and Vu [5].

In this section we shall prove Roth's Theorem (the case $k = 3$ of Szemerédi's Theorem), using the Regularity Lemma introduced by Szemerédi. We shall then see how one may use this powerful tool to prove several other beautiful results. The lemma has proven so useful, it is little exaggeration to say that whenever you see a graph, the the first thing you should think to do is to take its Szemerédi partition.

6.2 The Regularity Lemma

Szemerédi's Regularity Lemma, perhaps the most powerful tool in Graph Theory, may be thought of as an answer to the following, fairly vague question.

Question 6.2.1. *How well can an arbitrary graph be approximated by a collection of random graphs?*

Szemerédi's answer to this question was as follows:

> "Given any graph G, we can partition the vertex set $V(G)$ into a bounded number of pieces (at most k, say), such that for almost every pair (A, B) of parts, the induced bipartite graph $G[A, B]$ is approximated well by a random graph."

We shall quantify the statements 'almost every' and 'is approximated well by a random graph' in terms of a parameter $\varepsilon > 0$. The crucial point is that the bound k depends on ε, but not on $n = |V(G)|$.

To quantify 'almost every' is easy: it simply means all but εk^2 pairs. To make precise the intuitive idea that a bipartite graph 'looks random' is more complicated, and we shall need the following technical definition. Given subsets $X, Y \subset V(G)$, we shall write $e(X,Y)$ for the number of edges from X to Y, i.e., the size of the set $\{xy \in E(G) : x \in X, y \in Y\}$.

Definition. *Let $\varepsilon > 0$ and let $A, B \subset V(G)$ be disjoint sets of vertices in a graph G. The pair (A, B) is said to be* ε-regular *if*

$$\left| \frac{e(A,B)}{|A||B|} - \frac{e(X,Y)}{|X||Y|} \right| \leqslant \varepsilon$$

for every $X \subset A$ and $Y \subset B$ with $|X| \geqslant \varepsilon |A|$ and $|Y| \geqslant \varepsilon |B|$.

In other words, for every pair (X, Y) of sufficiently large subsets of A and B, the density of (X, Y) is about the same as that of (A, B). It is easy to see (using Chernoff's inequality, see [1]) that in the random graph $G_{n,p}$, any (sufficiently large) pair of subsets (A, B) is ε-regular with high probability.

Before stating the Regularity Lemma, let us see why the definition above is useful by proving a couple of easy consequences of ε-regularity. Indeed, if we believe that ε-regular pairs 'look like' random graphs, then we would like ε-regular graphs and random graphs to share some basic properties.

The simplest property of a random graph is that most vertices have roughly the same number of neighbours. This property is shared by ε-regular pairs.

Property 1. *Let (A, B) be an ε-regular pair of density d in a graph G. Then all but at most $2\varepsilon|A|$ vertices of A have degree between $(d - \varepsilon)|B|$ and $(d + \varepsilon)|B|$ in $G[A, B]$.*

Proof. Let $X \subset A$ be the set of low degree vertices in A, i.e.,

$$X := \Big\{ v \in A : |N_G(v) \cap B| < (d - \varepsilon)|B| \Big\},$$

and let $Y = B$. Then the density of the pair (X, Y) is less than $d - \varepsilon$, so $|X| < \varepsilon |A|$, by the definition of ε-regularity. By symmetry, the same holds for the high degree vertices. □

Another nice property of $G_{n,p}$ is that a (randomly chosen) induced subgraph of $G_{n,p}$ is also a random graph, with the same density.

Property 2. *Let (A, B) be an ε-regular pair of density d in a graph G. If $X \subset A$ and $Y \subset B$, with $|X| \geqslant \delta|A|$ and $|Y| \geqslant \delta|B|$, then (X, Y) is an $2\varepsilon/\delta$-regular pair, of density between $d - \varepsilon$ and $d + \varepsilon$.*

Proof. This follows easily from the definitions, so we leave its proof as an exercise. □

The properties above give us some idea of the motivation behind the definition of ε-regularity. However, the main reason that the definition is useful is the following 'embedding lemma'. It says that, for the purpose of finding small subgraphs in a graph G, we can treat 'dense' ε-regular pairs like complete bipartite graphs.

Given a graph H and an integer m, let $H(m)$ denote the graph obtained by 'blowing up' each vertex of H to size m, i.e., each vertex $j \in V(H)$ is replaced by a set A_j of size m. Thus $H(m)$ has vertex set $\bigcup_j A_j$, and edge set $\{uv : u \in A_i, v \in A_j \text{ for some } ij \in E(H)\}$.

Given a graph H, an integer m and $\delta > \varepsilon > 0$, let $\mathcal{G}(H, m, \varepsilon, \delta)$ denote the family of graphs G such that $V(G) = V(H(m))$, $G \subset H(m)$, and $G[A_i, A_j]$ is ε-regular and has density at least δ whenever $ij \in E(H)$.

The Embedding Lemma (simple version). *Let H be a graph, and let $\delta > 0$. There exists $\varepsilon > 0$ and $M \in \mathbb{N}$ such that if $m \geqslant M$ and $G \in \mathcal{G}(H, m, \varepsilon, \delta)$, then $H \subset G$.*

For most of our applications the simple version of the embedding lemma will be sufficient. The full version below is a bit harder to remember, but is not much harder to prove.

The Embedding Lemma. *Let $\Delta \in \mathbb{N}$, let $\delta > 0$, and let $\varepsilon_0 = \delta^{\Delta}/(\Delta + 2)$. Let R be a graph, let $m, t \in \mathbb{N}$ with $t \leqslant \varepsilon_0 m$, and let $H \subset R(t)$ with maximum degree at most Δ.*

If $\varepsilon_0 > \varepsilon > 0$ and $G \in \mathcal{G}(R, m, \varepsilon, \delta + \varepsilon)$, then G contains at least $(\varepsilon_0 m)^{|V(H)|}$ copies of H.

Proof. We shall prove the simple embedding lemma for the triangle $H = K_3$, and leave the proof of the full statement to the reader. To find a triangle in $G \in \mathcal{G}(K_3, m, \varepsilon, \delta)$, we simply pick vertices one by one. Indeed, choose $v_1 \in A_1$ arbitrarily amongst the vertices with at least $(\delta - \varepsilon)m$ neighbours in both A_2 and A_3.

Let $X = N(v) \cap A_2$ and $Y = N(v) \cap A_3$. Since $|X| \geqslant \varepsilon|A_2|$ and $|Y| \geqslant \varepsilon|A_3|$, it follows that there exists an edge between X and Y, so we are done. Alternatively (and more instructively for the general case), note that, by Property 2 above, the pair (X, Y) is ε'-regular and has density at least $\delta - \varepsilon$ (where $\varepsilon' = 2\varepsilon/(\delta - \varepsilon)$). Hence, by Property 1, all but $2\varepsilon'|X|$ vertices of X have degree at least $\delta - 2\varepsilon$ in $G[X, Y]$, and so we have found the desired triangle. □

Problem 6.2.2. *Prove the full embedding lemma. [Hint: choose vertices one by one, and keep track of their common neighbourhoods.]*

Inspired by the Embedding Lemma, we shall (throughout this section) refer to a graph on n vertices as *sparse* if it has $o(n^2)$ edges, and *dense* otherwise, i.e., if it has at least δn^2 edges for some $\delta > 0$. Note that (as stated) this definition is not precise; to understand it we should think of a sequence of graphs (G_n), where $|G_n| = n$ for each $n \in \mathbb{N}$, and consider the limit $n \to \infty$. However, our precise statements will hold for (large) fixed graphs.

Having (we hope!) convinced the reader that our definition of ε-regularity is a useful one, we are ready to state the Regularity Lemma.

Theorem 6.2.3 (The Szemerédi Regularity Lemma, 1975)**.** *Let $\varepsilon > 0$, and let $m \in \mathbb{N}$. There exists a constant $M = M(m, \varepsilon)$ such that the following holds.*

For any graph G, there exists a partition $V(G) = A_0 \cup \ldots \cup A_k$ of the vertex set into $m \leqslant k \leqslant M$ parts, such that

- $|A_1| = \ldots = |A_k|$,
- $|A_0| \leqslant \varepsilon|V(G)|$,
- *all but εk^2 of the the pairs (A_i, A_j) are ε-regular.*

The Regularity Lemma can be a little difficult to fully grasp at first sight; the reader is encouraged not to worry, and to study the applications below. We remark that the proof of the lemma (see

Section 6.7) is not difficult; the genius of Szemerédi was to imagine that such a statement could be true!

We note also that although the lemma holds for *all* graphs, it is only useful for large and (fairly) dense graphs. Indeed, if $n := |V(G)| < M(m, \varepsilon)$ then the statement is vacuous (since each part has size at most one), and if $e(G) = o(n^2)$ then G is well-approximated by the empty graph. Finally, we note that the best possible function $M(m, \varepsilon)$ is known to be (approximately) a tower of height $f(\varepsilon)$, for some function $\log(1/\varepsilon) \leqslant f(\varepsilon) \leqslant (1/\varepsilon)^5$.

6.3 Applications of the Regularity Lemma

The easiest way to understand the Regularity Lemma is to use it. In this section we shall give three simple (and canonical) applications: the Erdős-Stone Theorem, which we proved in Chapter 3; the 'triangle removal lemma', which we shall use to deduce Roth's Theorem; and bounds on Ramsey-Turán numbers.

6.3.1 The Erdős-Stone Theorem

Recall the statement of Theorem 3.3.

The Erdős-Stone Theorem. *Let H be an arbitrary graph. Then*

$$\mathrm{ex}(n, H) \;=\; \left(1 - \frac{1}{\chi(H) - 1} + o(1)\right)\binom{n}{2}.$$

We shall first give a sketch proof, and then fill in some of the details. The reader is encouraged to spend some time turning this sketch into a rigorous proof. Given a graph G, a typical application of the Regularity Lemma (SzRL) goes as follows:

1. Apply the SzRL (for some sufficiently small $\varepsilon > 0$). We obtain a partition $(A_0, \ldots, A_k)$ of the vertex set as described above.

2. Remove edges inside parts, between irregular pairs, and between sparse pairs. There are at most $O(\varepsilon n^2)$ such edges.

3. Consider the 'reduced graph' R which has vertex set $[k]$ and edge set

$$\big\{ij \,:\, \text{ the pair } (A_i, A_j) \text{ is dense and } \varepsilon\text{-regular}\big\}.$$

4. Apply a standard result from Graph Theory (e.g., Hall's Theorem, Dirac's Theorem, Turán's Theorem) to R.
5. Use the Embedding Lemma to find a copy of the desired subgraph H in G.

Now let us see how to apply this approach in the case of Erdős-Stone. The statement we are trying to prove is that, for every $\delta > 0$ there exists $N(\delta) \in \mathbb{N}$, such that if $n \geqslant N(\delta)$, $|G| = n$ and

$$e(G) \,\geqslant\, \left(1 - \frac{1}{\chi(H) - 1} + \delta\right)\binom{n}{2},$$

then $H \subset G$.

Proof of the Erdős-Stone Theorem. Let G be a graph on n vertices as described above, let $\varepsilon = \varepsilon(H, \delta) > 0$ be sufficiently small, and let $m = 1/\varepsilon$. Apply the SzRL to obtain a partition $(A_1, \dots, A_k)$, and form the reduced graph R.

<u>Claim</u>: $e(R) > \left(1 - \dfrac{1}{\chi(H) - 1} + \dfrac{\delta}{2}\right)\dbinom{k}{2}.$

Proof of Claim. Consider the edges inside parts, between irregular pairs, and between pairs of density at most $\delta/4$; we claim that there are at most $(\delta/3)\binom{n}{2}$ such edges. Indeed, there are at most $k(n/k)^2$ edges inside parts (recall that $k \geqslant m = 1/\varepsilon$); there are at most $\varepsilon k^2 (n/k)^2$ edges between irregular pairs (since there are at most εk^2 such pairs); and there are at most $\binom{k}{2}(\delta/4)(n/k)^2$ edges between 'sparse' pairs.

Thus only considering edges between dense, regular pairs, we obtain a graph $G' \subset G$ with

$$e(G') \,\geqslant\, \left(1 - \frac{1}{\chi(H) - 1} + \frac{2\delta}{3}\right)\binom{n}{2}.$$

Now, the edges of G' are all between pairs of parts which correspond to edges of R, so we have $e(G') \leqslant e(R)(n/k)^2$, and the claim follows. □

Applying Turán's Theorem to the reduced graph, we deduce from the claim that R contains a complete graph on $r = \chi(H)$ vertices. Let $i(1), \ldots, i(r)$ be the labels of the parts corresponding to the vertices of this complete graph, and let $A = \bigcup_{j=1}^{r} A_{i(j)}$.

Then $G'[A] \in \mathcal{G}(K_r, n', \varepsilon, \delta/2)$ for some $n' \geqslant n/2k$, by the definition of R. Also, since H is a fixed graph with $\chi(H) = r$, we have $H \subset R(t)$ for $t = |H|$. Hence, by the Embedding Lemma, $H \subset G$ as required. □

6.3.2 The Triangle Removal Lemma

The next application is a famous and important one; until recently, there was no known proof of it which avoided the Regularity Lemma.

The Triangle Removal Lemma (Ruzsa and Szemerédi, 1976)**.** *For every $\varepsilon > 0$ there exists a $\delta > 0$ such that the following holds:*

If G is a graph with at most δn^3 triangles, then all the triangles in G can be destroyed by removing εn^2 edges.

Proof. We shall use the same approach as in the proof above. Indeed, let $\varepsilon' = \varepsilon'(\varepsilon, \delta) > 0$ be sufficiently small, and apply the SzRL for ε'. Remove all edges inside parts, between irregular pairs, and between sparse pairs; there are at most εn^2 such edges.

We claim that the remaining graph, G', is triangle-free. Indeed, the only edges remaining correspond to edges of the reduced graph, R. Thus, if G' contains a triangle, it follows that R contains a triangle. But then, by the Embedding Lemma, G must contain at least $f(\varepsilon)n^3 > \delta n^3$ triangles, which is a contradiction. □

6.3.3 Roth's Theorem

Finally, let's deduce Roth's Theorem from the triangle removal lemma.

Roth's Theorem (Roth, 1954)**.** *If $A \subset \mathbb{N}$ has positive upper density, then A contains an arithmetic progression of length three.*

Proof. Given A, we first form a graph G as follows. Choose $0 < \varepsilon < \overline{d}(A)$, and choose n sufficiently large, with $|A \cap [n]| > \varepsilon n$. Let $V(G) = X \cup Y \cup Z$, where X, Y and Z are disjoint copies of $[n]$, and let

$$\begin{aligned} E(G) &= \big\{\{x,y\} : x \in X,\ y \in Y, \text{ and } y = x + a \text{ where } a \in A\big\} \\ &\cup \big\{\{y,z\} : y \in Y,\ z \in Z, \text{ and } z = y + a \text{ where } a \in A\big\} \\ &\cup \big\{\{x,z\} : x \in X,\ z \in Z, \text{ and } z = x + 2a \text{ where } a \in A\big\} \end{aligned}$$

The key observation is that all but n^2 of the triangles in G correspond to 3-APs in A. Indeed, if $\{x,y,z\}$ is a triangle in G, then a, b and $(a+b)/2$ are in A, where $a = y - x$ and $b = z - y$. So if A contains no 3-AP, then the only triangles in G correspond to triples (a,a,a). There are $n|A \cap [n]| \leqslant n^2$ such triangles (each is determined by two of its vertices).

Apply the triangle removal lemma to G. Since $n^2 < \delta(\varepsilon)n^3$ (if n is sufficiently large), it follows that we can destroy all triangles in G by removing at most εn^2 edges. But the triangles in G corresponding to the triples (a,a,a) are *edge-disjoint*! Hence $n|A \cap [n]| \leqslant \varepsilon n^2$, which is a contradiction. Thus A must contain a 3-AP, as claimed. □

Problem 6.3.1. *Prove Roth's Theorem in an arbitrary abelian group.*

6.3.4 Ramsey-Turán numbers

In Turán's Theorem, the extremal K_r-free graphs (i.e., the Turán graphs) have very large independent sets. What happens to the extremal number if we require in addition that our K_r-free graphs have small independence number?

Question 6.3.2. *Let G be a triangle-free graph on n vertices, and suppose that G has no independent set of size $f(n)$. If $f(n) = o(n)$, does it follow that $e(G) = o(n^2)$?*

Answer: Yes!

Proof. The maximum degree of a vertex in G is at most $f(n)$. □

Question 6.3.3. *Let G be a K_r-free graph on n vertices, for some $r \geqslant 5$, and suppose that G has no independent set of size $f(n)$. If $f(n) = o(n)$, does it follow that $e(G) = o(n^2)$?*

Answer: No!

Proof. Recall from Theorem 4.1.2 that there exist arbitrarily large graphs with girth at least four and no linear size independent set; call such a graph an Erdős graph.

Now let $H = T_{\lfloor (r-1)/2 \rfloor}(n)$, the $\lfloor (r-1)/2 \rfloor$-partite Turán graph on n vertices, and let G be the graph obtained from H by placing an Erdős graph in each part. Then G contains no K_r, no independent set of linear size, and at least $n^2/4$ edges. □

Given a 'forbidden' graph H, and a function $f : \mathbb{N} \to \mathbb{N}$, define the Ramsey-Turán number of the pair (H, f) to be

$$RT\big(n, H, f(n)\big) \; := \; \max\big\{ e(G) : |G| = n, H \not\subset G \text{ and } \alpha(G) \leqslant f(n) \big\},$$

where $\alpha(G)$ denotes the independence number of G.

From Questions 6.3.2 and 6.3.3, we know that $RT(n, K_r, o(n))$ is $o(n^2)$ if $r \leqslant 3$, and is $\Theta(n^2)$ if $r \geqslant 5$. But what about $r = 4$?

Theorem 6.3.4 (Szemerédi, 1972; Bollobás and Erdős, 1976)**.**

$$RT(n, K_4, o(n)) \; = \; \frac{n^2}{8} + o(n^2).$$

We shall prove the upper bound using the SzRL; the lower bound follows from an ingenious construction of Bollobás and Erdős, which we shall describe briefly below.

Proof of the upper bound in Theorem 6.3.4. We shall follow the usual strategy, but this time we shall need a couple of extra (fairly simple) ideas. Indeed, let G be a graph with n vertices, and apply the SzRL to G, for some sufficiently small $\varepsilon > 0$. Form the reduced graph R of ε-regular pairs with density at least δ, in the usual way.

We make the following two claims.

Claim 1: R is triangle-free.

Proof of Claim 1. Let $\{A, B, C\}$ be a triangle of parts in R, and choose $a \in A$ with $|N(a) \cap B|, |N(a) \cap C| \geqslant (\delta/2)|C|$. Since the pair $(N(a) \cap B, N(a) \cap C)$ is ε'-regular (by Property 6.2), we can choose a vertex $b \in N(a) \cap B$ such that $|N(b) \cap N(a) \cap C| \geqslant (\delta/2)^2|C|$.

But $\alpha(G) = o(n)$, so $N(a) \cap N(b) \cap C$ is not an independent set. But then G contains a copy of K_4, so we are done. □

By Claim 1 and Mantel's Theorem, R has at most $k^2/4$ edges. The upper bound is thus an immediate consequence of the following claim.

Claim 2: G contains no ε-regular pair with density bigger than $1/2+\delta$.

Proof of Claim 2. Let (A, B) be an ε-regular pair with density $1/2+\delta$. Since A has no linear size independent set, we can choose a pair $x, y \in A$ such that $xy \in E(G)$, and $|N(x) \cap B|, |N(y) \cap B| \geqslant (1/2 + \delta/2)|B|$.

But then $|N(x) \cap N(y) \cap B| \geqslant \delta|B|$, and so either $\alpha(G) \geqslant \delta|B|$, or the set $N(x) \cap N(y) \cap B$ contains an edge, in which case $K_4 \subset G$. In either case, we have a contradiction. □

The result follows by counting edges. There are at most

$$\frac{k^2}{4}\left(\frac{1}{2} + \delta\right)\left(\frac{n}{k}\right)^2 = \frac{n^2}{8} + o(n^2)$$

edges between dense regular pairs, and $o(n^2)$ other edges, so G has at most $n^2/8 + o(n^2)$ edges, as required. □

When Szemerédi proved the upper bound in Theorem 6.3.4 in 1972, most people expected the correct answer to be $o(n^2)$. It was therefore very surprising when Bollobás and Erdős gave the following construction, which shows that Szemerédi's bounds is in fact sharp!

The Bollobás-Erdős graph. *Let $U = S^d$ be a sufficiently high-dimensional sphere, and scatter $n/2$ red points and $n/2$ blue points at random on U. Join two points of the same colour if they are at distance less than $\sqrt{2} - \varepsilon$, and join points of different colours if they are at distance greater than $2 - \varepsilon$, for some suitably chosen $\varepsilon = \varepsilon(d) > 0$.*

Problem 6.3.5. *Show that the Bollobás-Erdős graph has $n^2/8+o(n^2)$ edges, contains no copy of K_4, and has independence number $o(n)$.*

6.4 Graph Limits

Another way of viewing the Regularity Lemma is topological in nature: it allows us to compactify the space of all large dense graphs.

Definition (Convergence of graph sequences). *A sequence $G_n = ([n], E_n)$ of graphs is* convergent *if for any graph H, the density of copies of H in G_n converges to a limit as $n \to \infty$.*

Theorem 6.4.1 (Lovász and Szegedy, 2004). *Every sequence of graphs has a convergent subsequence.*

In fact, one can say something stronger.

Definition (Graphons). *A* graphon *is a symmetric measurable function $p : [0,1] \times [0,1] \to [0,1]$.*

For each graphon, define a generalised Erdős-Rényi graph $G(n,p)$ on n vertices as follows: first give each vertex v a 'colour' $x_v \in [0,1]$, chosen uniformly at random; then add each edge vw with probability $p(x_v, x_w)$, all independently.

Let $G(\infty, p)$ denote the (formal) limit of the random graphs $G(n,p)$.

Theorem 6.4.2 (Lovász and Szegedy, 2004). *Every sequence of graphs has a subsequence converging to $G(\infty, p)$ for some graphon p.*

Various other metrics have been proposed for the space of graph sequences; interestingly, although the definitions are quite different from one another, almost all have turned out to be equivalent.

A closely related topic is that of *graph testing*: roughly speaking, a property of graphs $\mathcal{P}$ is *testable* if we can (with high probability as $k \to \infty$) test whether or not a large graph G is in $\mathcal{P}$ by looking at a random subgraph of G on k vertices. (More precisely, it accepts any graph satisfying $\mathcal{P}$ with probability 1, and rejects any which is $\varepsilon = \varepsilon(k)$-far from $\mathcal{P}$ (in a given metric) with probability $1 - \varepsilon$.)

The strongest result along these lines is due to Austin and Tao, who generalized results of Alon and Shapira (for graphs) and Rödl and Schacht (for hypergraphs).

Theorem 6.4.3 (Rödl and Schacht, 2007). *Every hereditary property of hypergraphs is testable.*

This theorem has some surprising consequences: for example, it implies Szemerédi's Theorem.

Another related question asks for a fast algorithm which produces a Szemerédi partition of a graph G. This problem has been studied by various authors, and is (at least partly) so challenging because, as noted earlier, the bound on the number of parts in a Szemerédi partition is *very* large.

Motivated by this shortcoming, Frieze and Kannan introduced the following notion of 'weak regularity': given a partition $\mathcal{P}$, with given edge densities d_{ij}, let

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

denote the expected number of edges of G between S and T, and say that the partition $\mathcal{P}$ is *weakly regular* if $|e(S,T) - e_{\mathcal{P}}(S,T)| \leqslant \varepsilon n^2$ for every $S,T \subset V(G)$.

Theorem 6.4.4 (Frieze and Kannan, 1999)**.** *For every $\varepsilon > 0$ and every graph G, there exists a weakly regular partition of $V(G)$ into at most $2^{2/\varepsilon^2}$ classes.*

Lovász and Szegedy showed that by iterating the Weak Regularity Lemma, one can obtain the usual Regularity Lemma as a corollary. They also proved a generalization in the setting of an arbitrary Hilbert space, and described an algorithm which constructs the weak Szemerédi partition as Voronoi cells in a metric space.

6.5 Recommended further reading

An excellent introduction to the area is provided by the survey:

J. Komlos and M. Simonovits, Szemerédi's Regularity Lemma and its applications in graph theory, DIMACS Technical Report (1996).

For a very different perspective, see

L. Lovász and B. Szegedy, Szemerédi's Lemma for the analyst, *J. Geom. and Func. Anal.*, **17** (2007), 252–270.

6.6 Exercises

1. Use Szemerédi's Regularity Lemma to prove:

The $(6,3)$**-Theorem** (Ruzsa and Szemerédi, 1976). *Let $\mathcal{A} \subset \mathcal{P}(n)$ be a set system such that $|A| = 3$ for every $A \in \mathcal{A}$, and such that for each set $B \subset [n]$ with $|B| = 6$, we have*

$$\big|\big\{A \in \mathcal{A} \,:\, A \subset B\big\}\big| \leqslant 2.$$

Then $|\mathcal{A}| = o(n^2)$.

2. Let $r_3(n)$ be the size of the largest subset of $[n]$ with no 3-AP, and let $f(k,n)$ denote the maximum number of edges in a graph on n vertices which is the union of k induced matchings.

   (a) Prove that $r_3(n) \leqslant \dfrac{f(n,5n)}{n}$.

   [Hint: consider a graph with edge set $(x + a_i, x + 2a_i)$.]

   (b) Using Szemerédi's Regularity Lemma, deduce Roth's Theorem.

In the next exercise, we shall prove the following theorem of Thomassen.

Thomassen's Theorem. *If G is a triangle-free graph with minimum degree $\big(\frac{1}{3} + \varepsilon\big)|G|$, then $\chi(G) \leqslant C$, for some constant $C = C(\varepsilon)$.*

Given a Szemerédi partition $A_0 \cup \ldots \cup A_k$ of $V(G)$, and $d > 0$, consider the auxiliary partition $V(G) = \bigcup_{I \subset [k]} X_I$, where

$$X_I \,:=\, \Big\{v \in V(G) : \, i \in I \Leftrightarrow |N(v) \cap A_i| \geqslant d|V_i|\Big\}.$$

3. (a) Show that if $|I| \geqslant 2k/3$, then X_I is empty.

   (b) Show that if $|I| \leqslant 2k/3$, then X_I is an independent set.

   (c) Deduce Thomassen's Theorem.

[You may assume the following strengthening of the Regularity Lemma: that the reduced graph R has minimum degree $\big(1/3 + \varepsilon/2\big)|R|$.]

4. Use the Kneser graph to show that Thomassen's Theorem is sharp.

6.7 Proof the of Regularity Lemma

The Regularity Lemma is not very hard to prove; the hard part was imagining that it could be true. In this section we shall help the reader to prove the lemma for himself.

The idea is as follows: starting with an arbitrary equipartition P, we shall repeatedly refine P, each time getting 'closer' to a Szemerédi partition. The key point is to find he right notion of the 'distance' of a partition from an ideal partition.

Given a partition P of V into parts $V_0, \ldots, V_k$, we define the index of P to be

$$\text{ind}(P) \;:=\; \frac{1}{k^2} \sum_{i \neq j} \Big(d(V_i, V_j) \Big)^2,$$

where $d(X,Y)$ denotes the density of the pair (X,Y), i.e., $\frac{e(X,Y)}{|X||Y|}$.

Problem 6.7.1. *Prove that if P is not a Szemerédi partition, i.e., more than εk^2 of the pairs are irregular, then there exists a refinement Q of P such that*

$$\text{ind}(Q) \;\geqslant\; \text{ind}(P) + \delta$$

for some $\delta = \delta(\varepsilon)$.

In order to solve Problem 6.7.1, you may need to use the following 'defect' form of the Cauchy-Schwarz inequality.

Improved Cauchy-Schwarz Inequality. *Let $a_1, \ldots, a_n > 0$, and let $m \in [n]$ and $\delta \in \mathbb{R}$. Suppose that*

$$\sum_{k=1}^{m} a_k \;=\; \frac{m}{n} \sum_{k=1}^{n} a_k \;+\; \delta.$$

Then

$$\sum_{k=1}^{n} a_k^2 \;\geqslant\; \frac{1}{n} \left(\sum_{k=1}^{n} a_k \right)^2 \;+\; \frac{\delta^2 n}{m(n-m)}.$$

We leave the remainder of the proof of Theorem 6.2.3 to the reader.

Problem 6.7.2. *Deduce Szemerédi's Regularity Lemma.*