9.4 SZEMERÉDI'S REGULARITY LEMMA

In this section we describe a fundamental result, the *Regularity Lemma*, proved by Endre Szemerédi in the 70s. The original motivation for proving it has been an application in Combinatorial Number Theory, leading, together with several additional deep ideas, to a complete solution of the Erdős-Turán conjecture discussed in Appendix B.2: every set of integers of positive upper density contains arbitrarily long arithmetic progressions. It took some time to realize that the lemma is an extremely powerful tool in Extremal Graph Theory, Combinatorics and Theoretical Computer Science. Stated informally, the regularity lemma asserts that the vertices of *every* large graph can be decomposed into a finite number of parts, so that the edges between almost every pair of parts form a random-looking graph. The power of the lemma is in the fact it deals with an arbitrary graph, making no assumptions, and yet it supplies much useful information about its structure. A detailed survey of the lemma and some of its many variants and fascinating consequences can be found in Komlós and Simonovits (1996).

Let $G = (V, E)$ be a graph. For two disjoint nonempty subsets of vertices $A, B \subset V$, let $e(A, B)$ denote the number of edges of G with one end in A and one in B, and let $d(A, B) = \frac{e(A,B)}{|A||B|}$ denote the *density* of the pair (A, B). For a real $\varepsilon > 0$, a pair (A, B) as above is called *ε-regular* if for every $X \subset A$ and $Y \subset B$ that satisfy $|X| \geq \varepsilon|A|,\ |Y| \geq \varepsilon|B|$ the inequality $|d(A, B) - d(X, Y)| \leq \varepsilon$ holds. It is not difficult to see that for every fixed positive ε, p, a fixed pair of two sufficiently large disjoint subsets A and B of a random graph $G = G(n, p)$ are very likely to be ε-regular of density roughly p. (This is stated in one of the exercises at the end of the chapter.) Conversely, an ε-regular pair A, B with a sufficiently small positive ε is random-looking in the sense that it shares many properties satisfied by random (bipartite) graphs.

A partition $V = V_0 \cup V_1 \cup \cdots \cup V_k$ of V into pairwise disjoint sets in which V_0 is called the *exceptional set* is an *equipartition* if $|V_1| = |V_2| = \cdots = |V_k|$. We view the exceptional set as $|V_0|$ distinct parts, each consisting of a single vertex. For two partitions $\mathcal{P}$ and $\mathcal{P}'$ as above, $\mathcal{P}'$ is a *refinement* of $\mathcal{P}$, if every part in $\mathcal{P}$ is a union of some of the parts of $\mathcal{P}'$. By the last comment on the exceptional set this means, in particular, that if $\mathcal{P}'$ is obtained from $\mathcal{P}$ by shifting vertices from the other sets in the partition to the exceptional set, then $\mathcal{P}'$ is a refinement of $\mathcal{P}$. An equipartition is called *ε-regular* if $|V_0| \leq \varepsilon|V|$ and all pairs (V_i, V_j) with $1 \leq i < j \leq k$, except at most εk^2 of them, are ε-regular.

Theorem 9.4.1 (The Regularity Lemma of Szemerédi (1978)) *For every $\varepsilon > 0$ and every integer t there exists an integer $T = T(\varepsilon, t)$ so that every graph with at least T vertices has an ε-regular partition $(V_0, V_1, \ldots, V_k)$, where $t \leq k \leq T$.*

The basic idea in the proof is simple. Start with an arbitrary partition of the set of vertices into t disjoint classes of equal sizes (with a few vertices in the exceptional set, if needed, to ensure divisibility by t). Proceed by showing that as long as the existing partition is not ε-regular, it can be refined in a way that increases the weighted average of the square of the density between a pair of classes of the partition by at least a constant depending only on ε. As this average cannot exceed 1, the process has to terminate after a bounded number of refinement steps. Since in each step we control the growth in the number of parts as well as the number of extra vertices thrown to the exceptional set, the desired result follows. The precise details require some care, and are given in what follows.

Let $G = (V, E)$ be a graph on $|V| = n$ vertices. For two disjoint subsets $U, W \subset V$, define $q(U, W) = \frac{|U||W|}{n^2} d^2(U, W)$. For partitions $\mathcal{U}$ of U and $\mathcal{W}$ of W, define

$$q(\mathcal{U}, \mathcal{W}) = \sum_{U' \in \mathcal{U}, W' \in \mathcal{W}} q(U', W').$$

Finally, for a partition $\mathcal{P}$ of V, with an exceptional set V_0, define $q(\mathcal{P}) = \sum q(U, W)$, where the sum ranges over all unordered pairs of distinct parts U, W in the partition, with each vertex of the exceptional set V_0 forming a singleton part in its own. Therefore, $q(\mathcal{P})$ is a sum of $\binom{k+|V_0|}{2}$ terms of the form $q(U, W)$. The quantity $q(\mathcal{P})$ is called the *index* of the partition $\mathcal{P}$. Since $d^2(U, W) \leq 1$ for all U, W, and since the sum $\sum |U||W|$ over all unordered pairs of distinct parts U, W is at most the number of unordered pairs of vertices, it follows that the index of any partition is smaller than $1/2$.

Lemma 9.4.2
(i) Let U, W be disjoint nonempty subsets of V, let $\mathcal{U}$ be a partition of U and $\mathcal{W}$ a partition of W. Then $q(\mathcal{U}, \mathcal{W}) \geq q(U, W)$.
(ii) If $\mathcal{P}'$ and $\mathcal{P}$ are partitions of V and $\mathcal{P}'$ is a refinement of $\mathcal{P}$, then $q(\mathcal{P}') \geq q(\mathcal{P})$.
(iii) Suppose $\varepsilon > 0$, and suppose U, W are disjoint nonempty subsets of V and the pair (U, W) is not ε-regular. Then there are partitions $\mathcal{U} = \{U_1, U_2\}$ of U and $\mathcal{W} = \{W_1, W_2\}$ of W so that $q(\mathcal{U}, \mathcal{W}) > q(U, W) + \varepsilon^4 \frac{|U||W|}{n^2}$.

Proof.
(i) Define a random variable Z as follows. Let u be a uniformly chosen random element of U, and let w be a uniformly chosen random element of W. Let $U' \in \mathcal{U}$ and $W' \in \mathcal{W}$ be those members of the partition so that $u \in U'$, $w \in W'$. Then $Z = d(U', W')$.

The expectation of Z is

$$\sum_{U' \in \mathcal{U}, W' \in \mathcal{W}} \frac{|U'||W'|}{|U||W|} d(U', W') = \sum_{U' \in \mathcal{U}, W' \in \mathcal{W}} \frac{|U'||W'|}{|U||W|} \frac{e(U', W')}{|U'||W'|} = d(U, W)\,.$$

By Jensen's Inequality, $\mathrm{E}\left[Z^2\right] \geq (\mathrm{E}\left[Z\right])^2$, and the desired result follows, as $\mathrm{E}\left[Z^2\right] = \frac{n^2}{|U||W|} q(\mathcal{U}, \mathcal{W})$ and $(\mathrm{E}\left[Z\right])^2 = d^2(U, W) = \frac{n^2}{|U||W|} q(U, W)$.

(ii) This is an immediate consequence of (i).

(iii) Since the pair (U, W) is not ε-regular, there are subsets $U_1 \subset U, W_1 \subset W$ so that $|U_1| \geq \varepsilon |U|$, $|W_1| \geq \varepsilon |W|$ and $|d(U_1, W_1) - d(U, W)| > \varepsilon$. Put $U_2 =$

$U - U_1$, $W_2 = W - W_1$ and define the partitions $\mathcal{U} = \{U_1, U_2\}$, $\mathcal{W} = \{W_1, W_2\}$. Let Z be the random variable defined in the proof of part (i). Then, as shown in that proof

$$\mathrm{Var}[Z] = \mathrm{E}\left[Z^2\right] - (\mathrm{E}\left[Z\right])^2 = \frac{n^2}{|U||W|}\left(q(\mathcal{U}, \mathcal{W}) - q(U, W)\right).$$

However, as $\mathrm{E}\left[Z\right] = d(U, W)$ it follows that with probability $\frac{|U_1||W_1|}{|U||W|}$, Z deviates from $\mathrm{E}\left[Z\right]$ by more than ε, implying that

$$\mathrm{Var}(Z) > \frac{|U_1||W_1|}{|U||W|}\varepsilon^2 \geq \varepsilon^4.$$

This provides the desired result. ∎

Proposition 9.4.3 *Suppose* $0 < \varepsilon \leq 1/4$, *let* $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ *be an equipartition of* V *where* V_0 *is the exceptional set,* $|V_0| \leq \varepsilon n$, *and* $|V_i| = c$ *for all* $1 \leq i \leq k$. *If* $\mathcal{P}$ *is not* ε*-regular then there exists a refinement* $\mathcal{P}' = \{V_0', V_1', \ldots, V_\ell'\}$ *of* $\mathcal{P}$, *in which* $k \leq \ell \leq k4^k$, $|V_0'| \leq |V_0| + \frac{n}{2^k}$ *all other sets* V_i *are of the same size, and* $q(\mathcal{P}') \geq q(\mathcal{P}) + \frac{\varepsilon^5}{2}$.

Proof. For every pair $1 \leq i < j \leq k$ define a partition $\mathcal{V}_{ij}$ of V_i and $\mathcal{V}_{ji}$ of V_j as follows. If the pair (V_i, V_j) is ε-regular, then the two partitions are trivial. Else, each partition consists of two parts, chosen according to Lemma 9.4.2, part (iii). For each $1 \leq i \leq k$, let $\mathcal{V}_i$ be the partition of V_i obtained by the Venn Diagram of all $(k-1)$-partitions $\mathcal{V}_{ij}$. Thus each $\mathcal{V}_i$ has at most 2^{k-1} parts. Let $\mathcal{Q}$ be the partition of V consisting of all parts of the partitions $\mathcal{V}_i$ together with the original exceptional set V_0. By Lemma 9.4.2 parts (ii), (iii), and since $\mathcal{P}$ is not ε-regular, we conclude that the index of $\mathcal{Q}$ satisfies

$$q(\mathcal{Q}) \geq q(\mathcal{P}) + \varepsilon k^2 \varepsilon^4 \frac{c^2}{n^2} = q(\mathcal{P}) + \varepsilon^5 \frac{(kc)^2}{n^2} > q(\mathcal{P}) + \frac{\varepsilon^5}{2},$$

where here we used the fact that $kc \geq (1 - \varepsilon)n \geq 3n/4$. Note that $\mathcal{Q}$ has at most $k2^{k-1}$ parts (besides the exceptional set), but those are not necessarily of equal sizes. Define $b = \left\lfloor c/4^k \right\rfloor$ and split every part of $\mathcal{Q}$ arbitrarily into disjoint sets of size b, throwing the remaining vertices in each part, if any, to the exceptional set. This process creates a partition $\mathcal{P}'$ with at most $k4^k$ non-exceptional parts of equal size, and a new exceptional set V_0' of size smaller than $|V_0| + k2^{k-1}b < |V_0| + kc/2^k \leq |V_0| + \frac{n}{2^k}$. Moreover, by Lemma 9.4.2, part (ii), the index $q(\mathcal{P}')$ of $\mathcal{P}'$ is at least $q(\mathcal{Q}) > q(\mathcal{P}) + \frac{\varepsilon^5}{2}$, completing the proof. ∎

Proof of Theorem 9.4.1. It suffices to prove the lemma for $\varepsilon \leq 1/4$ and t satisfying $2^{t-2} > \frac{1}{\varepsilon^6}$, hence we assume that these inequalities hold. Put $s = \left\lceil \frac{1}{\varepsilon^5} \right\rceil$, and note that for this choice $\frac{1}{2^k} \leq \frac{\varepsilon}{2s}$ for all $k \geq t$. Define $k_0 = t$ and $k_{i+1} = k_i 4^{k_i}$ for all $i \geq 0$. We prove the lemma with $T = k_s$.

Let $G = (V, E)$ be a graph with $|V| = n \geq T$ vertices. Start with an arbitrary partition $\mathcal{P} = \mathcal{P}_0$ of its vertices into $k = k_0 = t$ pairwise disjoint parts, each of size $\lfloor n/t \rfloor$, and let the exceptional set consist of the remaining vertices, if any. Note that their number is less than t, which is (much) smaller than $\varepsilon n/2$. As long as the partition $\mathcal{P}$ we have already defined is not ε-regular, apply Proposition 9.4.3 to refine it to a new equipartition $\mathcal{P}'$ with at most $k4^k$ non-exceptional parts, whose index exceeds that of $\mathcal{P}$ by at least $\frac{\varepsilon^5}{2}$, while the size of the exceptional set increases by at most $\frac{n}{2^k} < \frac{\varepsilon n}{2s}$. As the initial index is non-negative, and the index never exceeds $1/2$, the process must terminate in at most s steps, yielding an ε-regular partition with at most T non-exceptional parts, and an exceptional set of size smaller than εn. ■

Remark. The proof shows that $T(\varepsilon, \frac{1}{\varepsilon})$ is bounded by a tower of exponents of height roughly $1/\varepsilon^5$. Surprisingly, as shown by Gowers (1997), this tower-type behavior is indeed necessary.

Our next result both gives a good illustration of the near random nature of ϵ-regularity and shall play a role in the next section. $N(x)$, as usual, denotes the set of neighbors of x in the graph G. Let H be a graph on vertex set $1, \ldots, s$. Let G be a graph on vertex set V. Let $A_1, \ldots, A_s$ be disjoint subsets of V, each of size m. Let N denote the number of choices of $x_1 \in A_1, \ldots, x_s \in A_s$ such that x_i, x_j are adjacent in G whenever i, j are adjacent in H. (Note: Other x_i, x_j may or may not be adjacent.) Set $p_{ij} = d(A_i, A_j)$.

Theorem 9.4.4 *For all ε there exists $\gamma = \gamma_H(\varepsilon)$ with the following property: Assume, using the above notation, that (A_i, A_j) is ε-regular for all i, j adjacent in H. Then*

$$|Nm^{-s} - \prod_{\{i,j\} \in H} p_{ij}| \leq \gamma \tag{9.6}$$

Further, and critically, we may take γ such that

$$\lim_{\varepsilon \to 0^+} \gamma_H(\varepsilon) = 0 \tag{9.7}$$

The proof has some technicalities and the reader may take H as a triangle and $p_{12} = p_{13} = p_{23} = \frac{1}{2}$ to get the gist of the argument.

Proof. Set, with foresight, κ such that $(\kappa - \varepsilon)^s \geq \varepsilon$. We say we are in Case 1 if some $p_{ij} \leq \kappa$ (with i, j adjacent in H) and otherwise we are in Case 2. We will have $\gamma = \max(\gamma_1, \gamma_2)$ where γ_1, γ_2 handle two cases.

Case 1: Some $p_{ij} \leq \kappa$. Set $\gamma_1 = \kappa$. The product of the p_{ij} over edges $\{i, j\}$ is itself at most κ. There are $p_{ij}m^2 \leq \kappa m^2$ choices of adjacent x_i, x_j and therefore $N \leq \kappa m^2 m^{s-2} = \kappa m^s$. Also $N \geq 0$. Thus (9.6) is satisfied.

Case 2: $p_{ij} \geq \kappa$ for all adjacent i, j. For $1 \leq r \leq s$ we call a choice $x_i \in A_i$, $1 \leq i \leq r$ a partial copy if x_i, x_j are adjacent in G whenever i, j are adjacent in H. Further we call the choice normal (else abnormal) if the following holds for all $r < l \leq s$: Let U be the set of $u \leq r$ which are adjacent to l in H. Let Y be the intersection of the $N(x_u)$, $u \in U$ and A_l. Then

$$\prod_{u \in U} (p_{ul} - \varepsilon) \le |Y| m^{-1} \le \prod_{u \in U} (p_{ul} + \varepsilon) \tag{9.8}$$

Let $x_1, \ldots, x_r$ be a normal partial copy. We say it is destroyed by $x_{r+1} \in A_{r+1}$ if $x_1, \ldots, x_{r+1}$ is a partial copy but is not normal. We claim at most $2s\varepsilon m$ vertices x_{r+1} can destroy $x_1, \ldots, x_r$. How can this occur? Let $l > r+1$ be adjacent to $r+1$ and let U, Y be as above (looking only at $x_1, \ldots, x_r$). Then $Y \cap N(x_{r+1})$ would need to be either too big or too small. If more than $2s\varepsilon m$ vertices x_{r+1} destroyed $x_1, \ldots, x_r$ then there would be a set $X \subset A_{r+1}$ of size at least $m\varepsilon$ of such x_{r+1}, all with the same l and with either all $Y \cap N(x_{r+1})$ too big or all too small. Assume the former, the latter being similar. Then $d(X, Y) > p_{r+1,l} + \varepsilon$. But $|Y| \ge m\varepsilon$ by our choice of κ. From ε-regularity $|X| \le m\varepsilon$, as claimed.

The N choices of $x_i \in A_i$, $1 \le i \le s$ for which x_i, x_j are adjacent in G whenever i, j are adjacent in H fall into two categories. There are at most $2s^2\varepsilon m^s$ choices such that x_{r+1} destroys $x_1, \ldots, x_r$ for some r. The other choices are bounded in number between $m^s \prod(p_{ij} - \varepsilon)$ and $m^s \prod(p_{ij} + \varepsilon)$, the products over i, j adjacent in H. Let $f(\varepsilon)$ denote the maximum distance between either of these products and $\prod p_{ij}$. We can then set $\gamma_2 = 2s^2\varepsilon + f(\varepsilon)$.

■

9.5 GRAPHONS

As in Section 9.3 we set $N_G(H)$ denote the number of labelled copies of H as a (not necessarily induced) subgraph of G. We set $t(H, G) = N_G(H)n^{-a}$ where H, G have a, n vertices respectively. This may naturally be interpreted as the proportion of H in G, $0 \le t(H, G) \le 1$ tautologically.

Definition 5 *A sequence of graphs G_n is called a limit sequence, if $\lim_{n\to\infty} t(H, G_n)$ exists for all finite graphs H.*

Definition 6 *Two limit sequences G_n, G'_n are called equivalent if $\lim_{n\to\infty} t(H, G_n) = \lim_{n\to\infty} t(H, G'_n)$ for all finite graphs H. A* graphon *is an equivalence class of limit sequences.*

A graphon is a subtle object, an abstract limit of a convergent (by Definition 5) sequences of graphs. (We call a limit sequence G_n a graphon even though, technically, the graphon is the equivalence class.) It is *not* itself an infinite graph, though it may seem like one. It reflects the properties of very large graphs (formally, in a limit sense) of similar nature. The excellent book (Lovász 2012) serves as a general reference to graphons.

Surprisingly, and integral to the strength of this concept, there is a good characterization of graphons. Let $W : [0,1] \times [0,1] \to [0,1]$ be a Lebesgue measurable function with $W(x, y) = W(y, x)$ for all $x, y \in [0,1]$. For each positive integer n we define a random graph, denoted $G(n, W)$ on vertex set $1, \ldots, n$ as follows: