Szemerédi regularity lemma

We begin with some definitions.

Let A and B be disjoint sets of vertices in a graph. Let $e(A, B)$ denote the number of edges between A and B. The **density** (of the pair of sets), $d(A, B)$, is the ratio of the number of edges between the two sets to the total number of possible edges, i.e.,

$$d(A, B) = \frac{e(A, B)}{|A||B|}.$$

Fix $\epsilon > 0$. We call a pair of sets (A, B) **ϵ-regular** if for every pair $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| \geq \epsilon|A|$ and $|B'| \geq \epsilon|B|$ we have

$$|d(A, B) - d(A', B')| < \epsilon.$$

Roughly, a pair is ϵ-regular if the density of not-too-small subsets is close to the density of the sets themselves. Consequently, an ϵ-regular pairs have a "randomish" structure.

The goal of the regularity lemma is to show that certain graphs can be partitioned into (not too many) classes where most of the pairs of classes are ϵ-regular. This structure has many useful consequences. An **equipartition** (into r classes) is a partition of the vertices of a graph G where each class is of size $\lfloor \frac{n}{r} \rfloor$ or $\lceil \frac{n}{r} \rceil$ (i.e., classes are as close in size as possible).

Theorem 1 (Szemerédi regularity lemma, 1974)**.** *Given $\epsilon > 0$ and $m \geq 1$, there exists a constant $M = M(\epsilon, m)$ such that every graph on at least m vertices has an equipartition into r parts such that all but at most ϵr^2 pairs of classes are ϵ-regular and $m \leq r < M$.*

Often we call a partition that is given by the regularity lemma an **ϵ-regular partition** or a **Szemerédi partition**. The partition classes themselves are typically called **clusters**.

1: Show that the regularity lemma holds for graphs on $n < M$ vertices.

Hint: Partition into lots of classes.

Solution: Let G be a graph on $n \leq M$ vertices and consider the equipartition where every class is a single vertex. Every pair of classes has density either 0 or 1. Furthermore, for two classes A and B the only subsets A' and B' that satisfy $|A'| \geq \epsilon|A|$ and $|B'| \geq \epsilon|B|$ are A and B themselves, so every pair of classes is ϵ-regular, so the partition is ϵ-regular.

In this case, $n < M$, the lemma does not usually give useful information about the graph, so we are more concerned with the case when our graph has (many) more vertices than M.

It is not possible to avoid pairs that are not ϵ-regular. HW later on so called half graphs with ϵr bad pairs. Conlon and Fox gave construction that has at least r^2/r^* bad pairs, where r^* is the inverse Ackerman function. Bad news is that M is large. Known $O(\epsilon^{-5})$-level iterated exponential of m. Gowers found graphs that need M at least $\epsilon^{-1/16}$-level iterated exponential of m.

Furthermore, when G has $o(n^2)$ edges, then the regularity lemma also becomes trivial and similarly does not give useful information for large n. Therefore, we generally will only use the regularity lemma to examine graphs with $\Theta(n^2)$ edges and with n (very very) large.

The regularity lemma was developed to prove a famous conjecture of Erdős and Turán on the maximum possible size of a subset of $[n]$ that contains no k-term arithmetic progressions. Before the development of the regularity lemma, Roth proved the conjecture for $k = 3$. Szemerédi later proved the theorem for $k = 4$ and eventually for all values of k.

We will postpone the proof of the regularity lemma. Instead we begin with some applications. First we prove a lemma that uses only the definition of ϵ-regular pairs.

Lemma 2 (Triangle counting lemma)**.** *Let A, B, C be a partition of the vertices of a graph G. Suppose each pair of classes is ϵ-regular with densities $d(A,B) = x$, $d(B,C) = y$, and $d(C,A) = z$. If $x, y, z \geq 2\epsilon$, then the number of triangles with one vertex in each partition class is at least*

$$(1-2\epsilon)(x-\epsilon)(y-\epsilon)(z-\epsilon)|A||B||C|.$$

Proof. **2:** Observe that A contains at most $\epsilon|A|$ vertices each with less than $(x-\epsilon)|B|$ neighbors in B.

Solution: Indeed, if there are more, call them A' and note the density

$$d(A', B) = \frac{e(A', B)}{|A'||B|} < (x - \epsilon).$$

This implies that $|d(A,B) - d(A',B)| > \epsilon$; a contradiction.

By the same argument we can get that A contains at most $\epsilon|A|$ vertices each with less than $(z-\epsilon)|C|$ neighbors in C. Thus at least $(1-2\epsilon)|A|$ vertices in A have $\geq (x-\epsilon)|B| \geq \epsilon|B|$ neighbors in B and $\geq (z-\epsilon)|C| \geq \epsilon|C|$ neighbors in C. Let v be such a vertex and let B' and C' be the set of neighbors of v in B and C, respectively.

3: Use ϵ-regularity between B' and C' to finish the proof.

Solution: As $x, z \geq 2\epsilon$ we get $|B'| \geq (x-\epsilon)|B| \geq \epsilon|B|$ and $|C'| \geq (z-\epsilon)|C| \geq \epsilon|C|$. Therefore, by regularity, we have

$$|d(B,C) - d(B',C')| < \epsilon$$

so, $d(B',C') > y - \epsilon$. Thus, the total number of such edges between B' and C' is at least

$$d(B',C')|B'||C'| \geq (y-\epsilon)(x-\epsilon)(z-\epsilon)|B||C|.$$

Each of these edges forms a desired triangle with v and there are at least $(1-2\epsilon)|A|$ choices for v, so the total number of desired triangles is at least

$$(1-2\epsilon)(x-\epsilon)(y-\epsilon)(z-\epsilon)|A||B||C|.$$

□

Theorem 3 (Triangle removal lemma, Ruzsa-Szemerédi, 1978)**.** *For $\alpha > 0$, there exists $\beta > 0$ such that if G is an n-vertex graph that requires the removal of αn^2 edges to be triangle-free, then G has at least βn^3 triangles.*

The triangle removal lemma essentially says that a graph with few triangles can always be made triangle-free with the removal of relatively few edges. This straightforward idea is confirmed by application of the regularity lemma.

Proof. Let G be a graph as in the statement of the triangle removal lemma and note that $\alpha < \frac{1}{2}$. Let us apply the regularity lemma to G with $\epsilon = \frac{1}{6}\alpha$ and $m > \frac{3}{\alpha}$. That is, there exists a constant M (depending on α) such that for n large enough, G has an $\frac{1}{6}\alpha$-regular equipartition $V_1, \dots, V_r$ with $\frac{3}{\alpha} < r < M$.

We begin by removing edges from G. To avoid ceiling functions we use the inequality $\lceil \frac{n}{r} \rceil^2 \leq 2(\frac{n}{r})^2$.

4: Count the number of edges, that are 1. inside each cluster, 2. between pairs that are not ϵ-regular, 3. between pairs of clusters V_i and V_j if $d(V_i, V_j) \leq \frac{1}{3}\alpha$. Then sum 1.+2.+3. and conclude if removing these edges kills all triangles.

Solution:

1. Remove the edges inside of each cluster V_i. There are at most $r\binom{\lceil n/r \rceil}{2} \leq \frac{n^2}{r} < \frac{1}{3}\alpha n^2$ such edges.
2. Remove the edges between all pairs V_i, V_j that are not $\frac{1}{6}\alpha$-regular. There are at most $\frac{1}{6}\alpha r^2$ such pairs and each has at most $\lceil \frac{n}{r} \rceil^2 \leq 2(\frac{n}{r})^2$ edges. So we remove at most $\frac{1}{3}\alpha n^2$ such edges.
3. Remove the edges between all pairs V_i, V_j if the density of the pair $d(V_i, V_j) < \frac{1}{3}\alpha$. There are less than $\frac{1}{3}\alpha\binom{r}{2}2(\frac{n}{r})^2 < \frac{1}{3}\alpha n^2$ such edges.

In total we have removed less than αn^2 edges, thus the graph still contains a triangle xyz.

5: Find any triangle that remains after removing the edges from the previous exercise and show that counting lemma can be used to get β.

Solution: There are no edges inside any clusters, so the triangle must be between three classes; call them V_1, V_2, V_3. All three pairs of clusters must be $\frac{1}{6}\alpha$-regular and have density at least $\frac{1}{3}\alpha$. Furthermore, each class V_i has $|V_i| \geq \lfloor \frac{n}{r} \rfloor > \frac{n}{M}$. Therefore, by the triangle counting lemma, the number of triangles in V_1, V_2, V_3 and thus in G is at least

$$\left(1 - \frac{1}{3}\alpha\right)\left(\frac{\alpha}{6}\right)^3\left(\frac{n}{M}\right)^3 > \frac{5\alpha^3}{6^4 M^3}n^3.$$

Hence we can use $\beta = \frac{5\alpha^3}{6^4 M^3}$.

□

Extra note - when building the proof, one starts with constants like ϵ, m and in the end can pick find the proper dependence on α. The dependence is not really as important here anyway. Also note, the edges between irregular pairs, inside parts and in non-dense pairs were all negligible and hidden in some small error in the end.