

# Lecture 18: Extremal graphs and Szemerédi's regularity lemma · 1MA170

Vilhelm Agdur<sup>1</sup>

<sup>1</sup> vilhelm.agdur@math.uu.se

11 December 2023

We discuss some basic notions of extremal graph theory, and give some nice proofs of Turán's theorem and related ideas. Then we introduce the Szemerédi regularity lemma, and use it to prove the triangle removal lemma, which we can use to prove Roth's theorem on arithmetic progressions of length three.

## Turán's theorem

As usual, we start by repeating a definition from the exercise session.

**Definition 1.** Given any graph  $H$ , we say that a graph  $G$  is  $H$ -free if it has no subgraph isomorphic to  $H$ . We say that it is *maximal*  $H$ -free if adding any edge to it would create a subgraph isomorphic to  $H$ , and we say that it is *maximum*  $H$ -free (or *extremal* among  $H$ -free graphs) if additionally no other  $H$ -free graph has more edges than  $G$ .

For each integer  $n$ , we define the *extremal function* for  $H$ , denoted  $\text{ex}(n; H)$ , to be the number of edges of a maximum  $H$ -free graph on  $n$  vertices.

In the exercises, you were asked to investigate what an extremal  $r$ -clique-free graph might look like. The answer to this problem is a famous theorem due to Turán.<sup>2</sup>

**Definition 2.** The *Turán graph*  $T(n, r)$  is a complete multipartite graph divided into  $r$  parts, with  $n$  vertices divided as equally as possible between the parts.<sup>3</sup>

This graph has

$$\left(1 - \frac{1}{r} + o(1)\right) \frac{n^2}{2}$$

edges.

**Theorem 3** (Turán, 1941). *It holds for every  $r$  that*

$$\text{ex}(n; K_{r+1}) \leq \left(1 - \frac{1}{r}\right) \frac{n^2}{2},$$

*and in particular the Turán graphs are the extremal  $(r + 1)$ -clique-free graphs.*

We will see two different proofs of this theorem. The first will only show the upper bound, not the extremality of the Turán graphs, but

<sup>2</sup> The special case of triangle-free graphs is due to Mantel.

<sup>3</sup> Concretely, if  $n = qr + s$  for natural numbers  $q$  and  $s < r$ , it has  $s$  parts of size  $q + 1$  and  $r - s$  parts of size  $q$ .

the second will prove that the Turán graphs are extremal in order to show the bound.

We start with one that uses a result we have seen before, the Caro-Wei result about independent sets.<sup>4</sup>

*Proof of Theorem 3.* We begin by noting that a clique in a graph is precisely an independent set in its complement graph. Now  $d_{G^c}(v) = n - 1 - d_G(v)$ , so by Caro-Wei we get that

$$\omega(G) = \alpha(G^c) \geq \sum_{v \in V} \frac{1}{d_{G^c}(v) + 1} = \sum_{v \in V} \frac{1}{n - d_v}.$$

Next, we are going to use the Cauchy-Schwarz inequality, which states that<sup>5</sup> for any  $(a_1, a_2, \dots, a_n), (b_1, b_2, \dots, b_n) \in \mathbb{R}^n$ , we have

$$\left( \sum_{i=1}^n a_i b_i \right)^2 \leq \left( \sum_{i=1}^n a_i^2 \right) \left( \sum_{i=1}^n b_i^2 \right).$$

So let  $a_v = \sqrt{n - d_v}$  and  $b_v = \frac{1}{\sqrt{n - d_v}}$ , so that  $a_v b_v = 1$  for all  $v$ , and Cauchy-Schwarz states that

$$\begin{aligned} n^2 &\leq \left( \sum_{v \in V} n - d_v \right) \left( \sum_{v \in V} \frac{1}{n - d_v} \right) \\ &\leq \omega(G) \left( \sum_{v \in V} n - d_v \right) \\ &= \omega(G) (n^2 - 2|E|), \end{aligned}$$

where in the first step we used our bound on  $\omega(G)$  we have just shown, and in the second the fact that  $\sum_v d_v = 2|E|$  that we showed near the very beginning of the course.

So if we assume that  $\omega(G) \leq r$ , we get that

$$n^2 \leq r (n^2 - 2|E|),$$

which if you solve for  $|E|$  becomes precisely the inequality of Turán's theorem.  $\square$

The second proof we present of this theorem is due to Erdős.

*Proof of Theorem 3.* Our proof will proceed by showing that we can always increase the edge-count of a graph, while retaining its  $r$ -clique-freeness, in a way that makes it closer to a complete multipartite graph. This will show that the extremal graph must be complete multipartite, and thus in particular must be the Turán graphs.<sup>6</sup>

Let  $G = (V, E)$  be some  $r$ -clique-free graph on  $n$  vertices, and let  $v$  be a vertex of maximum degree in  $G$ . Let  $S = N(v)$  and  $T = V \setminus N(v)$ .

<sup>4</sup> Which was originally invented to be used in this proof of Turán's theorem.

<sup>5</sup> Or in the more compact linear algebra formulation,

$$|\langle a, b \rangle| \leq \|a\| \|b\|,$$

but it's more convenient for us to just write out what it means termwise.

<sup>6</sup> That this is so follows from the following fact:

**Exercise 1.** Show that the Turán graphs are the  $r$ -partite graphs on  $n$  vertices with the most edges.

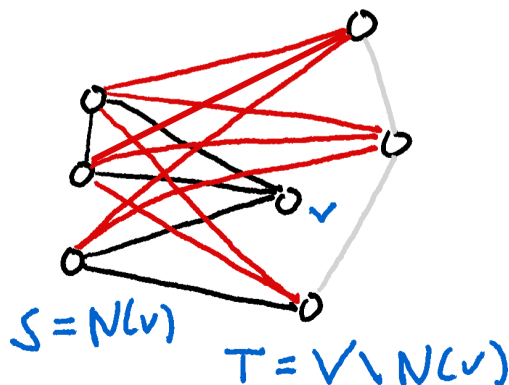


Figure 1: A figure of the construction of  $H$  from  $G$  in the Erdős proof of Theorem 3. The black edges are edges that are kept, grey edges are removed, and red edges are added. The vertex  $v$  and sets  $S$  and  $T$  are labelled in blue.

We note that since  $G$  is  $r$ -clique-free and  $v$  is adjacent to every vertex of  $S$ ,  $G[S]$  must in fact be  $(r-1)$ -clique-free.

Now we construct a graph  $H$  by taking  $G$ , deleting every edge inside of  $T$ , and adding every edge between  $S$  and  $T$ . This construction is illustrated in Figure 1.

Since  $G[S]$  is  $(r-1)$ -clique-free, adding a new neighbour to all of them cannot create an  $r$ -clique, so  $H$  is still  $r$ -clique-free. It remains to see that it has at least as many edges as  $G$ .

So consider the degree of any vertex  $w$ . If  $w = v$ , its degree didn't change, if  $w \in S$ , its degree can only have increased, and if  $w \in T$  its new degree is  $d_v$ , which by assumption was maximum, so it again has not decreased. So since the degree of every vertex has increased or remained unchanged, we conclude that  $|E(H)| \geq |E(G)|$ .

So we conclude that the maximal  $r$ -clique-free graphs must be ones which are unchanged by this process, and a moment's thought reveals that these must be precisely the complete multipartite graphs.<sup>7</sup>  $\square$

We can actually employ the idea in this proof to answer a related question: If a triangle-free graph has very nearly  $\text{ex}(n; K_3)$  edges, will it be very close to being bipartite?<sup>8</sup> Results of this kind are called *stability* results, since they answer questions of the type "if you nudge the extremal object a bit, will it still look similar".

**Theorem 4** (Simonovits 1968, Füredi 2015). *Every triangle free graph with  $\text{ex}(n, K_3) - t$  edges may be made bipartite by removing  $t$  edges.*

*Proof.* Take  $G = (V, E)$  to be such a graph, and as before, pick  $v \in V$  to be a vertex of maximum degree. We set

$$A = V \setminus N(v), \quad B = N(v)$$

and notice that due to triangle-freeness, there can be no edges internal to  $B$ .

7

**Exercise 2.** Think for a moment.

<sup>8</sup> I first saw this proof in a summer school in Prague in 2022, and when the lecturer there introduced this proof he said that it was extremely nice, and if we ever taught a course on graph theory, we had to include it. So now, a year and a half later, I am teaching a course on graph theory, and so I dug it up, found it is indeed very nice, and included it.

Let us consider the sum of the degrees of the vertices in  $A$ . On the one hand, we can compute

$$\begin{aligned}\sum_{w \in A} d_w &= 2e(A) + e(A, B) \\ &= e(A) + (e(A) + e(A, B) + e(B)) = e(A) + e(G) \\ &= \text{ex}(n; K_3) - t + e(A).\end{aligned}$$

On the other hand, we have since  $d_v$  is maximum, that

$$\sum_{w \in A} d_w \leq |A|d_v = (n - d_v)d_v \leq \text{ex}(n; K_3).$$

Here, the final inequality follows from that  $(n - d_v)d_v$  is precisely the number of edges of a complete bipartite graph with parts of size  $d_v$  and  $n - d_v$ , which is certainly triangle free on  $n$  vertices and thus has at most  $\text{ex}(n; K_3)$  edges.

So we have seen that  $\text{ex}(n; K_3) \geq \text{ex}(n; K_3) - t + e(A)$ , so  $e(A) \leq t$ . So if we remove these at most  $t$  edges, there will be no edges internal to either  $A$  or  $B$ , and so  $G$  will be bipartite.  $\square$

### *Szemerédi's regularity lemma and the triangle removal lemma*

As we explored in the exercises, the Szemerédi regularity lemma tells us that all large enough graphs look roughly like a random multipartite graph, so their structure can be summarized by a “blueprint” graph. Let us now see how we can use this result to see a sort of converse result to our previous result about nearly maximum triangle-free graphs.

In particular, the result we are going to show states that if you need to remove a large amount of edges to destroy every triangle in the graph, then the graph must in fact have contained a lot of triangles.

**Theorem 5** (Erdős-Simonovits, 1983). *For every  $\varepsilon > 0$  there is a  $\delta > 0$  such that if  $G$  is an  $n$ -vertex graph such that at least  $\varepsilon n^2$  edges have to be removed from  $G$  in order to make it triangle-free, then  $G$  has at least  $\delta n^3$  triangles.*

Since the proof of this uses Szemerédi regularity, let us restate that theorem here as well, along with a crucial definition used in its statement:

**Definition 6.** Let  $G = (V, E)$  be a graph on  $n$  vertices, and  $A$  and  $B$  two disjoint subsets of  $V$ . For any  $\varepsilon > 0$ , we say that the pair  $(A, B)$  is  $\varepsilon$ -regular if it holds for all  $X \subseteq A, Y \subseteq B$  with  $|X| \geq \varepsilon|A|$  and  $|Y| \geq \varepsilon|B|$  that

$$|d(X, Y) - d(A, B)| \leq \varepsilon.$$

**Theorem 7** (Szemerédi's regularity lemma). *For every  $\varepsilon > 0$  and  $m \in \mathbb{N}$  there exists an  $M \in \mathbb{N}$  such that for every graph  $G = (V, E)$  on at least  $M$  vertices and every  $\delta \in [0, 1]$ , there exists*

- a) a blueprint  $R = ([k], E_R, w)$  whose minimum weight is at least  $\delta$ ,
- b) a partition  $V = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$ ,
- c) and a spanning subgraph  $G'$  of  $G$ ,

such that

- 1.  $m \leq k \leq M$ ,
- 2.  $0 \leq |V_0| \leq \varepsilon|V|$ ,<sup>9</sup>
- 3.  $|V_1| = |V_2| = \dots = |V_k|$ ,
- 4. for every  $v \in V$

$$d_{G'}(v) > d_G(v) - (\delta + \varepsilon)|V|,$$

where  $d_{G'}$  and  $d_G$  denote degrees in  $G'$  and  $G$  respectively,

- 5. the graph  $G'' = G'[V \setminus V_0]$  is multipartite with the sets  $V_i$  as parts,<sup>10</sup>
- 6. and all of the pairs  $(V_i, V_j)$  for  $1 \leq i, j \leq k$  are  $\varepsilon$ -regular, and their density is  $w(i \sim j)$  if  $i \sim j$  is an edge of  $R$ , and otherwise there are no edges between them.

*Remark 8.* The graph  $G''$  will be multipartite, and have at most  $\frac{\delta+3\varepsilon}{2}n^2$  edges fewer than the graph  $G$  we started with.<sup>11</sup>

Let us now give an idea of a proof of the triangle removal lemma – except we avoid the details of what the exact values of our  $\varepsilon$ s and  $\delta$ s will be. We don't need to know that, and it would be painful to keep track of in a lecture.<sup>12</sup>

*Proof sketch of Theorem 5.* Suppose  $G$  is a graph satisfying the hypotheses of the theorem, that is, it has  $n$  vertices and you need to remove at least  $\varepsilon n^2$  edges to make it triangle-free.

Then we can pick some  $\varepsilon_0$  and  $\delta$  such that we, when we insert those into the regularity lemma, get a blueprint  $R$  and a multipartite subgraph  $G''$  of  $G$  which has had less than  $\varepsilon n^2$  edges removed.

Then, by our hypothesis, this  $G''$  must contain a triangle, and since it is multipartite the vertices of this triangle must belong to three different parts  $V_i, V_j, V_k$ .

Thus there has to be a triangle  $\{i, j, k\}$  in the blueprint  $R$ , and the weight of each of the edges of this triangle has to be at least  $\delta$ . The intuition now is that  $G''$  looks like a random multipartite graph with

<sup>9</sup> So notice in particular that  $V_0$  may be empty.

<sup>10</sup> Which concretely just means that for every  $i = 1, 2, \dots, k$ , there are no edges internal to  $V_i$  in  $G'$ , which you could alternatively phrase as that these sets are independent in  $G'$ .

<sup>11</sup> Showing this was one of the exercises in the exercise session on this topic.

<sup>12</sup> If you wish, try to flesh out the proof with all those details – or just use Google to find a fully detailed proof.

blueprint  $R$ , and in such a random multipartite graph we would expect to see

$$w(i \sim j)w(i \sim k)w(k \sim j)n^3 \geq \delta^3 n^3$$

triangles between the three parts.

Now, we can of course not reason like that, because the graph  $G$  is not actually random – however, all the three pairs  $(V_i, V_j)$ ,  $(V_i, V_k)$  and  $(V_j, V_k)$  are  $\varepsilon_0$ -regular, which essentially means that they can't behave too differently from the behaviour of a random graph. So applying the regularity property a few times,<sup>13</sup> we see that there is indeed a  $\delta'$  such that  $G''$ , and thus also  $G$ , has at least  $\delta'n^3$  triangles.

□

### Roth's theorem

For what may be the first time in this course, let us apply graph theory to prove a result about some other part of mathematics. Particularly, we are going to look at part of the result that Szémerédi originally proved the regularity lemma for, namely arithmetic progressions in positive-density sets.

The original theorem he proved is the following:<sup>14</sup>

**Theorem 10** (Szemerédi, 1975). *For every  $k \in \mathbb{N}$  and  $\alpha > 0$ , there exists some  $n_0 \in \mathbb{N}$  such that if  $n \geq n_0$  and  $A \subseteq [n]$  with  $|A| \geq \alpha n$ , then  $A$  contains a  $k$ -term arithmetic progression.*<sup>15</sup>

The theorem we shall be proving is just the case of  $k = 3$ , which is known as Roth's theorem, and was originally proved using techniques from analytic number theory.

**Theorem 11** (Roth, 1953). *For every  $\alpha > 0$  there is some  $n_0$  such that if  $n \geq n_0$  and  $A \subseteq [n]$  with  $|A| \geq \alpha n$ , then  $A$  contains a three-term arithmetic progression.*

*Proof.* Suppose we are given some set  $A \subseteq [n]$  with  $|A| = \alpha n$ . We create a tripartite graph  $G$  on the vertices  $V_1 \sqcup V_2 \sqcup V_3$ , where we let

$$V_1 = \{(i, 1) \mid i \in [n]\}, \quad V_2 = \{(i, 2) \mid i \in [2n]\}, \quad V_3 = \{(i, 3) \mid i \in [3n]\},$$

so there are  $6n$  vertices in total. We add edges by declaring that for every  $x \in [n]$  and  $a \in A$ , the triangle  $\{(x, 1), (x + a, 2), (x + 2a, 3)\}$  is present.

Now we observe that all these triangles we have added are edge-disjoint, so we need to remove one edge from each triangle to destroy them all – and there are a total of  $3n|A| = 3\alpha n^2$  edges. So by the triangle removal lemma, there are at least  $\delta n^3$  triangles in  $G$ , for some  $\delta$  depending only on  $\alpha$ .

<sup>13</sup> In particular, the following lemma comes in handy:

**Lemma 9.** *Let  $(A, B)$  be an  $\varepsilon$ -regular pair with density  $d$  in a graph  $G$ . For any  $a \in A$  and  $Y \subseteq B$ , let*

$$d_Y(a) = d_{G[Y \cup \{a\}]}(a),$$

*that is,  $d_Y(a)$  is the number of neighbours of  $a$  in  $Y$ .*

*It holds that, if  $|Y| \geq \varepsilon|B|$ , we have*

$$|\{a \in A \mid d_Y(a) < (d - \varepsilon)|Y|\}| < \varepsilon|A|.$$

*That is, for any large enough subset  $Y$  of  $B$ , nearly all vertices of  $A$  have the expected number of neighbours in  $Y$ .*

<sup>14</sup> His proof of this result is notoriously long and difficult to understand, motivating other proofs using methods from different areas of mathematics of the same result, most famously by Furstenberg using ergodic theory, and Gowers, using Fourier analysis.

<sup>15</sup> Recall that a  $k$ -term arithmetic progression is a set of integers  $\{a, a + d, a + 2d, \dots, a + (k - 1)d\}$ .

We now claim that these triangles in  $G$  correspond to three-term arithmetic progressions in  $A$ . So suppose we have a triangle on the vertices  $(x, 1)$ ,  $(y, 2)$  and  $(z, 3)$ . Since  $(x, 1) \sim (y, 2)$  is an edge, we must have  $y = x + a_0$  for some  $a_0 \in A$ . Similarly, since  $(x, 1) \sim (z, 3)$ , we must have  $z = x + 2a_1$  for some  $a_1 \in A$ , and likewise we must have  $z = y + a_2$  for some  $a_2 \in A$ .

If we do some elementary algebra on this, what we have seen is that

$$a_2 - a_1 = a_1 - a_0$$

so if  $a_1 \neq a_0$ , the three numbers  $a_0, a_1, a_2$  will form a three-term arithmetic progression in  $A$  with common difference  $a_1 - a_0$ .<sup>16</sup>

The only case in which this triangle did not correspond to a three-term arithmetic progression is when  $a_0 = a_1$ . However, these triangles are precisely the edge-disjoint triangles we started with, and we know there are only  $4\alpha n^2 < \delta n^3$  of those, so there has to be at least one triangle that did arise from an arithmetic progression.<sup>17</sup>  $\square$

## Exercises

<sup>16</sup> This may of course be negative – if you prefer your arithmetic sequences not to have negative difference, the sequence  $a_2, a_1, a_0$  may be more to your liking.

<sup>17</sup> In fact, this of course shows that there are  $\Omega(n^3)$  such triangles, which a bit of thought will reveal correspond to  $\Omega(n^2)$  arithmetic progressions – so we could have written a much stronger conclusion to our theorem: It not only has *one* arithmetic progression, it has quite many.