Exercise session 15: Extremal graph theory and Szemerédi's regularity lemma · 1MA170

Vilhelm Agdur[1]

[1] vilhelm.agdur@math.uu.se

29 November 2023

We introduce the concept of extremal graph theory, starting with Turan's theorem. Then we introduce Szemerédi's regularity lemma as a tool for extremal graph theory.

Extremal graphs

We start with the central definition of extremal graph theory, and then we explain what it actually means through some exercises.

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 is 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 $\mathrm{ex}(n; H)$, to be the number of edges of a maximum H-free graph on n vertices.

Exercise 1. As a warm-up exercise, if H is the path on three vertices, what is $\mathrm{ex}(n; H)$? What are the extremal graphs for this problem?

Letting H_k be a star graph with k leaves,[2] what is $\mathrm{ex}(n; H_k)$? Which are the extremal graphs here?

[2] That is, a tree with one root with k children, and no other vertices or edges.

Having done this warmup, we can move on to the original question that motivated the start of extremal graph theory: How many edges can a graph have if it does not contain any triangles? This requirement clearly imposes *some* bound on the number of edges – a complete graph certainly contains a triangle – but what is the bound?

Exercise 2. Letting $H = K_3$, the triangle graph, what is $\mathrm{ex}(n; K_3)$? What do the extremal graphs look like?[3]

[3] Side exercise: Can you find a graph that is maximal triangle-free but not extremal?

Exercise 3. Can you generalize what you just did to finding $\mathrm{ex}(n; K_k)$ for $k > 3$?

In the lecture, we will see several very elegant proofs of Turán's theorem, which is the theorem that gives the answer to this question. Let us close out with an exercise that is definitely rather tough, but has a very elegant solution[4] which points towards more advanced problems.

[4] Which uses a probabilistic method style of argument.

Exercise 4. Let us generalize the definition of $\mathrm{ex}(n; H)$ to saying that for every graph G,

$$\mathrm{ex}(G; H) = \max\{|E(F)| \mid H \not\subseteq F \subseteq G\},$$

that is, $\mathrm{ex}(G; H)$ is the largest number of edges of an H-free subgraph of G. So $\mathrm{ex}(n; H) = \mathrm{ex}(K_n; H)$.

Prove that for all H and all n-vertex graphs G,

$$\mathrm{ex}(G; H) \geq \mathrm{ex}(n; H) \frac{|E(G)|}{\binom{n}{2}}.$$

The Szemerédi regularity lemma

Let G be a very large graph with lots of vertices and edges. Is there a way to summarize roughly what this graph looks like? At first glance, the question seems absurd: If we are considering *any* graph, surely it can look like anything, and so we can't compress the information any further?

The regularity lemma[5] gives us a way of doing this, including error bounds on how wrong our summary is.

Before we give the statement of the regularity lemma with all its quantifiers, let us introduce the following random graph model:

Definition 2. Take a weighted graph $R = ([k], E, w)$ with weights in the interval $(0, 1]$, which we call a *blueprint*. Then, for any n the *random multipartite graph with blueprint R*, denoted $G(n, R)$, is a k-partite graph whose parts $A_1, A_2, \ldots, A_k$ all have n vertices, and where an edge between $a \in A_i$ and $b \in A_j$ is present with probability $w(i \sim j)$.

Exercise 5. Convince yourself that if you take R to be the complete graph on n vertices with all edge weights equal to p, then the graph $G(1, R)$ is just an Erdős-Rényi graph $G(n, p)$.

What the regularity lemma says is essentially that every large enough graph looks mostly like a random multipartite graph for some blueprint, if you squint a bit at it and ignore a few edges. So this blueprint is the summary of the graph which we were looking for.

To see the sense in which this is true, we need first to introduce two more concepts.

Definition 3. Let $G = (V, E)$ be some graph on n vertices. The *density* of G is

$$d(G) = \frac{|E|}{\binom{n}{2}}.$$

For any two disjoint sets of vertices X and Y, let $E(X, Y)$ denote the set of edges with one endpoint in X and one in Y, and let $e(X, Y) =$

[5] Which is what we will call it throughout, omitting Szemerédi's name in the interest of not having to pronounce too many Hungarian names.

$|E(X, Y)|$. Then the *density* of the pair (X, Y) is

$$d(X, Y) = \frac{e(X, Y)}{|X||Y|}.$$

Exercise 6. Prove that in $G(n, R)$, the random multipartite graph with blueprint R, if $X = A_i$ and $Y = A_j$ are two parts of the partition of the graph, the density $d(X, Y)$ converges to $w(i \sim j)$ in probability and almost surely.[6]

So as the result of the previous exercise indicates, the density of a pair of vertex sets is sort of the probability of an edge between them – except it is a deterministic thing, so it makes sense to talk about for arbitrary graphs.

In fact, our random bipartite graph will have a much stronger property than just having the "right" densities between parts – even if you zoom in on just a subset of the parts, they will still have the right density, as long as they aren't too tiny. No particular subset of a part is "special". The way to make this formal is through the notion of ε-regularity.

Definition 4. 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 ε-regular if it holds for all $X \subseteq A, Y \subseteq B$ with $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ that

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

Exercise 7. Can you find a graph G with a pair of vertex sets that fails to be $\frac{1}{4}$-regular?

Exercise 8. Convince yourself that it is plausible that for every blueprint R and every $\varepsilon > 0$, it holds with high probability that every pair of parts of a random multipartite graph $G(n, R)$ is ε-regular.[7]

If you feel like it, you could even try to prove this, though that would be more of an exercise in probability than in graph theory.[8]

So the concept of ε-regularity captures a sense in which the edges between two sets of vertices can look like they were chosen at random, which is made plausible by the fact that an actually random graph will have its pairs be ε-regular.

Now, we can finally state the Szemerédi regularity lemma, and it will hopefully make sense why we say it morally means all graphs look roughly like random multipartite graphs.

Theorem 5 (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 δ,

[6] If you don't recall this from probability theory, what you want to use to prove this is the weak and the strong laws of large numbers, respectively.

[7] The order of quantifiers matters here, so spend a moment to think about it. We first pick an ε, and then we claim that the probability that *every* pair of parts is ε-regular tends to one as n goes to infinity.

[8] To show this it actually suffices to show the version of the statement with the weaker quantifier ordering: that for every ϵ and every pair of parts, that pair is ϵ-regular with high probability, since there are only a fixed and finite number of pairs of parts.

b) a partition $V = V_0 \coprod V_1 \coprod \ldots \coprod V_k$,

c) and a spanning subgraph G' of G,

such that

1. *$m \leq k \leq M$,*

2. *$0 \leq |V_0| \leq \varepsilon |V|$,[9]*

3. *$|V_1| = |V_2| = \ldots = |V_k|$,*

4. *for every $v \in V$*
$$d_{G'}(v) > d_G(v) - (\delta + \epsilon)|V|,$$
   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,[10]*

6. *and all of the pairs (V_i, V_j) for $1 \leq i, j \leq k$ are ε-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.*

> [9] So notice in particular that V_0 may be empty.

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

Exercise 9. That theorem was quite the mouthful. Take a moment to convince yourself that it really is saying that there exists a subgraph $G'' = G'[V \setminus V_0]$ that looks like it was sampled from $G(m, R)$ for the given blueprint R and an m that is approximately $\frac{n}{k}$.

It is obvious that we have lost at most $\varepsilon |V|$ vertices of G by passing to G''. How many edges can we have lost?

In the lecture, we will see how we can use this machinery we have set up to prove Roth's theorem on the existence of arithmetic progressions of length three, and also mention that it can also be used to prove the Erdős-Stone-Simonovits theorem, which is a vast generalization of what we did with the K_k-free graphs.