Extremal graph theory

David Conlon

Lecture 1

The basic statement of extremal graph theory is Mantel's theorem, proved in 1907, which states that any graph on n vertices with no triangle contains at most $n^2/4$ edges. This is clearly best possible, as one may partition the set of n vertices into two sets of size $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ and form the complete bipartite graph between them. This graph has no triangles and $\lfloor n^2/4 \rfloor$ edges.

As a warm-up, we will give a number of different proofs of this simple and fundamental theorem.

Theorem 1 (Mantel's theorem) *If a graph G on n vertices contains no triangle then it contains at most $\frac{n^2}{4}$ edges.*

First proof Suppose that G has m edges. Let x and y be two vertices in G which are joined by an edge. If $d(v)$ is the degree of a vertex v, we see that $d(x) + d(y) \leq n$. This is because every vertex in the graph G is connected to at most one of x and y. Note now that

$$\sum_x d^2(x) = \sum_{xy \in E} (d(x) + d(y)) \leq mn.$$

On the other hand, since $\sum_x d(x) = 2m$, the Cauchy-Schwarz inequality implies that

$$\sum_x d^2(x) \geq \frac{(\sum_x d(x))^2}{n} \geq \frac{4m^2}{n}.$$

Therefore

$$\frac{4m^2}{n} \leq mn,$$

and the result follows. □

Second proof We proceed by induction on n. For $n = 1$ and $n = 2$, the result is trivial, so assume that we know it to be true for $n - 1$ and let G be a graph on n vertices. Let x and y be two adjacent vertices in G. As above, we know that $d(x)+d(y) \leq n$. The complement H of x and y has $n-2$ vertices and since it contains no triangles must, by induction, have at most $(n-2)^2/4$ edges. Therefore, the total number of edges in G is at most

$$e(H) + d(x) + d(y) - 1 \leq \frac{(n-2)^2}{4} + n - 1 = \frac{n^2}{4},$$

where the -1 comes from the fact that we count the edge between x and y twice. □

Third proof Let A be the largest independent set in the graph G. Since the neighborhood of every vertex x is an independent set, we must have $d(x) \leq |A|$. Let B be the complement of A. Every edge in G must meet a vertex of B. Therefore, the number of edges in G satisfies

$$e(G) \leq \sum_{x \in B} d(x) \leq |A||B| \leq \left(\frac{|A|+|B|}{2}\right)^2 = \frac{n^2}{4}.$$

Suppose that n is even. Then equality holds if and only if $|A| = |B| = n/2$, $d(x) = |A|$ for every $x \in B$ and B has no internal edges. This easily implies that the unique structure with $n^2/4$ edges is a bipartite graph with equal partite sets. For n odd, the number of edges is maximised when $|A| = \lceil n/2 \rceil$ and $|B| = \lfloor n/2 \rfloor$. Again, this yields a unique bipartite structure. □

The last proof tells us that not only is $\lfloor n^2/4 \rfloor$ the maximum number of edges in a triangle-free graph but also that any triangle-free graph with this number of edges is bipartite with partite sets of almost equal size.

The natural generalisation of this theorem to cliques of size r is the following, proved by Paul Turán in 1941.

Theorem 2 (Turán's theorem) *If a graph G on n vertices contains no copy of K_{r+1}, the complete graph on $r+1$ vertices, then it contains at most $\left(1 - \frac{1}{r}\right)\frac{n^2}{2}$ edges.*

First proof By induction on n. The theorem is trivially true for $n = 1, 2, \ldots, r$. We will therefore assume that it is true for all values less than n and prove it for n. Let G be a graph on n vertices which contains no K_{r+1} and has the maximum possible number of edges. Then G contains copies of K_r. Otherwise, we could add edges to G, contradicting maximality.

Let A be a clique of size r and let B be its complement. Since B has size $n - r$ and contains no K_{r+1}, there are at most $\left(1 - \frac{1}{r}\right)\frac{(n-r)^2}{2}$ edges in B. Moreover, since every vertex in B can have at most $r - 1$ neighbours in A, the number of edges between A and B is at most $(r-1)(n-r)$. Summing, we see that

$$e(G) = e_A + e_{A,B} + e_B \leq \binom{r}{2} + (r-1)(n-r) + \left(1 - \frac{1}{r}\right)\frac{(n-r)^2}{2} = \left(1 - \frac{1}{r}\right)\frac{n^2}{2},$$

where $e_A, e_{A,B}$ and e_B are the number of edges in A, between A and B and in B respectively. The theorem follows. □

Second proof We again assume that G contains no K_{r+1} and has the maximum possible number of edges. We will begin by proving that if $xy \notin E(G)$ and $yz \notin E(G)$, then $xz \notin E(G)$. This implies that the property of not being connected in G is an equivalence relation. This in turn will imply that the graph must be a complete multipartite graph.

Suppose, for contradiction, that $xy \notin E(G)$ and $yz \notin E(G)$, but $xz \in E(G)$. If $d(y) < d(x)$ then we may construct a new K_{r+1}-free graph G' by deleting y and creating a new copy of the vertex x, say x'. Since any clique in G' can contain at most one of x and x', we see that G' is K_{r+1}-free. Moreover,

$$|E(G')| = |E(G)| - d(y) + d(x) > |E(G)|,$$

contradicting the maximality of G. A similar conclusion holds if $d(y) < d(z)$. We may therefore assume that $d(y) \geq d(x)$ and $d(y) \geq d(z)$. We create a new graph G'' by deleting x and z and creating two extra copies of the vertex y. Again, this has no K_{r+1} and

$$|E(G'')| = |E(G)| - (d(x) + d(z) - 1) + 2d(y) > |E(G)|,$$

so again we have a contradiction.

We now know that the graph is a complete multipartite graph. Clearly, it can have at most r parts. We will show that the number of edges is maximised when all of these parts have sizes which differ by at most one. Indeed, if there were two parts A and B with $|A| > |B| + 1$, we could increase the number of edges in G by moving one vertex from A to B. We would lose $|B|$ edges by doing this, but gain $|A| - 1$. Overall, we would gain $|A| - 1 - |B| \geq 1$. □

This second proof also determines the structure of the extremal graph, that is, it must be r-partite with all parts having size as close as possible. So if $n = mr + q$, we get q sets of size $m + 1$ and $r - q$ sets of size m.