Lecture 14: The probabilistic method and the Erdős-Rényi random graph · 1MA020

Vilhelm Agdur[1]

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

28 November 2023

> We introduce a new tool to the course, the probabilistic method, and use it to prove some new results. We also introduce the Erdős-Rényi random graph, which is interesting in its own right as well as being a tool in the probabilistic method.

The minimum bisection problem

Suppose we are given a graph G, which is too large to fit on a single computer – so we wish to distribute the computation we want to do on the graph between two machines. We imagine the computation we want to do is in some sense possible to do "locally" – so if the graph had two equally sized connected components, the two computers wouldn't have to talk to each other at all.

Of course, most graphs do not split up that nicely, but we can still try to divide it into two equally sized parts that have as few edges between them as possible. How well can we do this?

Proposition 1. *For any graph $G = (V, E)$ on $n \in 2\mathbb{N}$ vertices with maximum degree $\Delta < \frac{n}{2}$, there exists a partition $A \coprod B$ of V, such that $|A| = |B| = \frac{n}{2}$, and*

$$e(A, B) \leq \frac{|E|}{2},$$

where $e(A, B)$ is the number of edges between A and B.

Proof. Let G^c be the complement graph of G – that is, G^c has the same vertices as G, and there is an edge $x \sim y$ in G^c if and only if $x \sim y$ is not an edge of G.

Since we assumed $\Delta < \frac{n}{2}$, it is clear that the *minimum* degree of G^c is at least $\frac{n}{2}$, and so by Dirac's theorem[2] there exists a Hamilton cycle in G^c. Picking every second edge of this Hamilton cycle, we get a perfect matching M on G^c.

[2] Which we proved earlier in the course.

Now, we apply the probabilistic method, picking a random partition $A \coprod B$ of V: For each edge of M, we randomly place one of its endpoints in A and one in B, independently between different edges. We can then compute that

$$\mathbb{E}\left[|e(A, B)|\right] = \mathbb{E}\left[\sum_{x \sim y \in E} \mathbb{1}_{\{x \sim y \in E(A,B)\}}\right]$$

$$= \sum_{x \sim y \in E} \mathbb{P}\left(x \text{ and } y \text{ are in different sets}\right).$$

Now, whenever x and y are not connected by an edge in M, their assignment to A or B is independent, and so the probability that they are in different sets is precisely $\frac{1}{2}$. However, we picked our matching M so that it is a perfect matching on G^c – in particular this means that $M \cap E = \varnothing$, and so all of the probabilities that we are summing must in fact be $\frac{1}{2}$ by our previous argument.

So the sum is just a sum of $|E|$ many $\frac{1}{2}$'s, and so what we have seen is that $\mathbb{E}\left[|e(A, B)|\right] = \frac{|E|}{2}$. Now, we know that the expectation is never greater than every specific outcome, so there must exist some choice of A and B with no more than $\frac{|E|}{2}$ edges between them, and we have shown the proposition. □

Girth and chromatic number

Already when we first introduced the chromatic number, we mentioned the obvious fact that a large clique forces a graph to have high chromatic number. One might be led by this to believe that being very sparse, without any cliques, would mean you have low chromatic number. This unfortunately turns out to be not at all true, in a pretty strong sense.

Definition 2. The *girth* of a graph G is the length of the shortest cycle in the graph.

So the notion of girth generalizes the notion of being triangle-free, which is just having girth greater than three. Of course any triangle-free graph also contains no larger cliques, so this is indeed a strong notion of having no cliques.

Theorem 3 (Erdős, 1959). *For all positive integers k there exists a graph G of girth and chromatic number at least k.*

The proof of this uses the probabilistic method, and proceeds by considering a random graph from a suitably chosen distribution, and showing that the probability that it has girth and chromatic number at least k is greater than 0. We postpone the proof of this until a later lecture, when we will have developed our probabilistic toolbox a bit more.

Exercises