2MMD30: Graphs and Algorithms **Lecture 8** **Date: 9/3/2016**
Instructor: Nikhil Bansal

1 Applications of the probabilistic method

As will shall see repeatedly in the rest of the course, the probabilistic method is a very powerful tool for proving theorems. This method uses the fact that if $E(X)$ is the expected value of the random variable X, then there must be some event x with $x \geq E(X)$ and some event y with $y \leq E(X)$. Hence, if you can find out the expected value of some random variable, then that proves the existence of such events. Note that lower or upper bounds on the expected value work as well, but yield only an event $x \geq LB$ if $E(X) \geq LB$, or an event $y \leq UB$ or $E(X) \leq UB$.

2 Basic probability theory

We recall some basic probability theory. Let X be a random variable, and let $\{v_1, v_2, \ldots, v_n\}$ be the values that X can take (the realizations of X). Let $P(X = v_i) = p_i$ be the chance that X takes the value v_i. Let $E(X) = \sum_i v_i p_i$ be the expected value of X. Two random variables X and Y are independent if and only if $P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y)$.

An important property of the expected value is its linearity: if X and Y are two (not necessarily independent) random variables, then $E(X + Y) = E(X) + E(Y)$. This is called *linearity of expectation* and is extremely useful.

There are a few important elementary estimates that we often use and we record them here.

$$\begin{array}{rrcll} & 1-x & \leq & e^{-x} & x \in (0,1) \\ \text{so} & (1-p)^n & \leq & e^{-pn} & p \in (0,1) \\ & \binom{n}{k} & \leq & \left(\frac{en}{k}\right)^k & \\ & \binom{n}{k} & \leq & 2^n & \end{array}$$

3 Mutual Friends or Mutual Strangers (Ramsey Numbers)

This is a classic puzzle. Suppose there are six people at party. Show that there are always either three people who mutually know each other, or some three that are complete strangers.

The proof is as follows. Consider a graph with six vertices, one for each person, and put a blue edge between two vertices, if the corresponding people know each other, and a red edge otherwise. We need to show that there is always either a all blue or all red triangle.

Consider person 1, she has 5 out-going edges, so without loss generality, say three or more of them are white. Let $2, 3, 4$ denote the endpoint of any three of these edges. Now if any edge among

$2, 3, 4$ is white, this makes a white triangle together with 1. But if not, then this means that $2, 3, 4$ form a red triangle. Hence we are done.

The study of these type of phenomena is known as Ramsey Theory, and has several surprising applications to many areas of mathematics. Morally, these results say that perfect randomness cannot exist: If you have a large enough graph (or structure), you can never avoid certain patterns.

Let us extend the puzzle beyond 3. For an integer k, let $R(k, k)$ denote the smallest integer such that any complete red-blue graph of $R(k, k)$ vertices, contains either a red K_k or a blue K_k. We just saw that $R(3, 3) \leq 6$. Can you also show that $R(3, 3) \geq 6$ (and hence $R(3, 3) = 6$). (For this you need to show a two-coloring of edges of K_5 that has no monochromatic K_3.)

What is known about $R(k, k)$? We know that $R(4, 4) = 18$. For $R(5, 5)$ we only know that it lies somewhere between 43 and 49. The gap in our knowledge gets larger as k increases.

It is known that $R(k, k) \leq \binom{2k-2}{k-1}$ which is approximately $4^{s-1}/\sqrt{\pi s}$ (we shall prove this later in the exercise). This bound has only been improved very slightly over the years. On the lower bound side, Erdos showed that $R(k, k) \geq 2^{k/2}$ or more precisely, $k2^{s/2}/(\sqrt{2}e)$. Interestingly, he did not construct an explicit counter-example to show this (as we did in the class for $R(3, 3)$). He did this using a counting/probabilistic argument!

In fact, the best explicit counter-examples we know for $R(k, k)$ only give a bound of about $k^{\Omega(\log k)}$ which is about $2^{\Omega(\log^2 k)}$ and hence extremely weak compared to the one obtained by the probabilistic method. Giving a better explicit construction is a major open problem.

3.1 The Probabilistic Proof

Let us now see the probabilistic proof that $R(k, k) > 2^{k/2}$. To show that $R(k, k) > n$ (for some suitably chosen n), the idea is to show via a counting argument that not all edge bi-colorings of K_n can have a monochromatic K_k. This implies that there must exist some graph which suffices as our counter-example, even though we have no clue how this looks like! In fact, 0.999 fraction of the colorings are counter-examples, which means that if take any random coloring, it is almost always a counter example. And yet, we do not know to explicit find one!

Consider the graph K_n and independently color each edge blue or red with probability 1/2. Call a coloring bad, if there is a monochromatic K_k. Now, consider all possible $\ell := \binom{n}{k}$ subsets of vertices. Call these sets $S_1, \ldots, S_\ell$. If a coloring is bad, clearly one (or more) of the subsets is monochromatic. Now, what is the probability that the coloring on S_1 is bad? For this to happen all the $\binom{k}{2}$ edges of S_1 are colored either all red or all blue. The chance of this happening is $2 \cdot 2^{-k(k-1)/2}$.

So, by linearity of expectation the expected number of bad sets will be

$$\binom{n}{k} 2 2^{-k(k-1)/2} \leq \frac{n^k}{k!} 2^{-k^2/2} 2^{1+k/2}$$

If $n = 2^{-k/2}$ then the right hand side is $2^{1+k/2}/k!$ which is strictly less than 1 for $k \geq 3$ (easy to check as if k increases by 1, the denomination increases by factor k and numerator by $\sqrt{2}$).

This implies that there must exist a coloring of edges of K_n with 0 monochromatic K_k, and hence we are done.

4 Independent sets

We introduce the probabilistic method by strengthening a bound on independent sets that was part of the first set of homework exercises. This bound was that for any graph G with maximum degree d, it has some independent set of size at least $\frac{n}{d+1}$. We sharpen this bound as follows:

Theorem 1 *For any graph $G = (V, E)$, G contains some independent set I with $|I| \geq \sum_{v \in V} \frac{1}{d_v+1}$.*

Proof: Pick a random permutation of V, say $\{v_1, v_2, \ldots, v_n\}$ uniformly among all the $n!$ possible ones. We use this permutation to compute some independent set I as follows. We first initialize $S = V$. We pick the vertex with the lowest index v_i in S (which is v_1 in the first iteration), add it to I and then remove v_i and all its neighbors from S. We repeat this until S is empty. Obviously, the resulting I is an independent set, but we don't know exactly how large it is.

We will find the expected size of I. To this end, we introduce new random variables $I_1, I_2, \ldots I_n$ such that $I_i = 0$ iff v_i is not in I, and $I_i = 1$ iff v_i is in I. These kinds of random variables (that are either 0 or 1) are called 'indicator' random variables. Indicator variables are useful because $E(I_i) = P(X_i = 1)$ (which follows immediately from the definitions).

Obviously, the size of I is exactly the sum over the indicator variables we introduced, so $|I| = \sum_i I_i$. We therefore have $E(|I|) = E(\sum_i I_i) = \sum_i E(I_i) = \sum_i P(I_i = 1)$ by linearity of expectation. Now pick some vertex v_i. It is included in I exactly if none of its neighbors appear earlier in the random permutation. The chance that this happens is therefore $\frac{1}{d_{v_i}+1}$, where d_{v_i} is the degree of v_i. This gives us that $E|I| = \sum_{v \in V} \frac{1}{d_v+1}$, which in turn implies there exists some independent set with size at least $\sum_{v \in V} \frac{1}{d_v+1}$ by the probabilistic method, which proves the theorem. $\square$

5 Crossing numbers

For a graph G, we define the crossing number $C(G)$ as the minimum of the number of edge crossings that occur in a drawing of the graph, over all such drawings. A crossing means that two edges intersect, so an edge may partake in more than one crossing. For example, any planar graph has $C(G) = 0$, and in fact, a graph is planar if and only if $C(G) = 0$.

If we look at the complete graph K_n, not much is known about its crossing number. We know it for K_{11}, but not for K_{13} for example. We do know that $\Omega(n^4) \leq C(K_n) \leq \frac{3}{8}\binom{n}{4}$.

We will use the probabilistic method to prove the following theorem:

Theorem 2 *For any graph $G = (V, E)$ with $n = |V|$ and $m = |E|$ where $m \geq 4n$, we have $C(G) \geq \frac{m^3}{64n}$.*

Note that this theorem immediately gives the $\Omega(n^4) \leq C(K_n)$ bound as $m = O(n^2)$.

We first prove a lemma we will use to prove the above theorem. Note that we assume $m \geq 4n$ throughout this section.

Lemma 3 $C(G) \geq m - (3n - 6)$

Proof: If $m \leq 3n - 6$, then the claim holds trivially.

If $m > 3n - 6$, the graph is non-planar, so there is some edge crossing. We remove an edge involved in some crossing, decreasing m by one and $C(G)$ by at least 1. We repeat this until we have removed at least $m-(3n-6)$ edges. This means we have found at least $m-(3n-6)$ crossings, which proves the lemma. □

Proof:(Theorem 2:) We now use the probabilistic method to amplify the bound given by the lemma to the one in the theorem. To this end, we create a new graph G' which includes a vertex from V with probability p, which we determine later (we remove all edges that were connected to vertices we didn't include). Let n' be the number of vertices in G' and m' the number of edges. The lemma gives us the following:

$$\begin{aligned} C(G') &\geq m' - 3n' + 6 \geq m' - 3n' \\ E(C(G')) &\geq E(m') - 3E(n') \\ &= mp^2 - 3pn \qquad (1) \end{aligned}$$

Note that G' is a random graph, but the first statement above holds for every G'. The second step follows by taking expectation (over the probability distribution of G'). The last step follows as $E(n') = pn$ as every vertex was included with probability p, and $E(m') = p^2m$ as an edge was included if both its endpoints were included.

Now we look at $E(C(G'))$. Consider the 'optimal' drawing of G (the drawing with the least number of crossings). A crossing there survives as a crossing of G' (in the drawing of G' induced from that of G) only if both the edges defining this crossing survive. But this happens with probability exactly p^4. So the expected number of crossings that survive is $p^4C(G)$. Now, the optimal way of drawing G', so we get $E(C(G')) \leq p^4C(G)$, or equivalently $p^4C(G) \geq E(C(G'))$. Together with (1), we therefore have

$$p^4C(G) \geq mp^2 - 3pn.$$

We now pick $p = \frac{4n}{m}$. Note that $p \leq 1$, as $m \geq 4n$ was assumed at the beginning. We fill in p and obtain that

$$C(G) \geq \frac{m}{p^2} - \frac{3n}{n^3} = \frac{m^3}{16n^2} - \frac{3nm^3}{64n^3} = \frac{m^3}{64n^2}$$

□

Remark: The choice of $p = 4n/m$ is not optimum and be improved some more by a more careful analysis. However, this does not make a qualitative difference and only improves the lower bound by a small constant factor.

6 Tournament graphs

We define a tournament graph as a directed graph in which every pair of vertices has exactly one edge between them. The direction of this edge then determines what the graph looks like. An example of such a graph would be the results of a tournament: an edge (v, u) then signifies that team v has defeated team u.

We're interested in choosing k teams (vertices) that we declare 'the winners' of the tournament. We say that some subset of vertices is a reasonable set of winners if there is no other team that beats all the winners (indeed, if some team beats all the winners, and yet is labelled a loser, the fans of this team will not be happy). We'd like to know if we can always choose k winners from every tournament.

Theorem 4 *For any k, there exists some tournament graph such that no subset of k teams is a reasonable set of k winners.*

Proof: We will use the probabilistic method to find such a tournament. We take a random tournament graph on n edges, where we will specify n later. To do this, we flip a coin per edge to determine its direction.

We pick a set of vertices S of size k. The chance that some vertex u not in S does not beat all of S is $1 - 2^{-k}$. The chance that all vertices not in S do not beat all of S is $\left(1 - 2^{-k}\right)^{n-k}$, as these chances are independent. This is therefore also the chance that S is a reasonable set of k winners. There are $\binom{n}{k}$ subsets of size k, so we have:

$$\begin{aligned}
P(G \text{ has a reasonable set of winners}) &\leq \binom{n}{k} P(S \text{ is a reasonable set of winners}) \\
&\leq \binom{n}{k} \left(1 - 2^{-k}\right)^{n-k} \\
&\leq n^k e^{-2^{-k}(n-k)} \\
&= e^{k \log n - 2^{-k}(n-k)}
\end{aligned}$$

If we make n large enough, we can make this last expression as small as we want (say $\leq 1/2$) for fixed k. This is because $n - k$ grows much faster the $\log n$ as n increases). This proves that there exists some tournament that does not have a reasonable set of k winners. □

7 Graphs with large girth and chromatic number

If a graph contains a k clique (a complete subgraph), then it needs a coloring of at least k colors. It is natural to wonder that if we exclude such subgraphs, whether the graph in question then becomes colorable by $O(1)$ colors, in particular, if the graph is triangle-free. In one of the first applications of the probabilistic method, Erdös proved that there exists graphs with both large girth and chromatic number, which answers the question with a 'no'. In a sense, it means that 'local considerations are not useful for $\chi(G)$', where $\chi(G)$ is the chromatic number of the graph, which is the least number of colors needed to color the graph.

Theorem 5 *For any k, there exists some graph such that the girth of the graph is at least k and $\chi(G) \geq k$.*

Before we prove this theorem, we look at the Kneser graphs. A Kneser graph $KG_{n,k}$ is a graph with $\binom{n}{k}$ vertices each corresponding to some k subset of a set with n elements. It has an edge between two vertices if their subsets are disjoint. If we pick $k = \frac{n}{3} + 1$, then the Kneser graph is triangle-free - a triangle would correspond to three pairwise disjoint sets of $\frac{n}{3} + 1$ elements each, which would imply there are $n + 3$ elements in total - but as we will see later in the course, $\chi(KG_{n,k}) \geq n - 2k + 2 = n/3$ for this graph. This is therefore a counterexample to the idea that triangle-free graphs might have at most a constant chromatic number (but note that theorem 5 is say something much stronger than just triangle-free).

Proof: We will now prove the theorem. We create a random graph on n vertices by including every possible edge with probability p, where we fix

$$p = \frac{n^{1/2k}}{n}$$

We'd first like to know how many cycles of length at most k this random graph contains. Note that any graph contains at most $\binom{n}{l}l!$ cycles of length l, as there are $\binom{n}{l}$ subsets of size l and $l!$ ways to cycle through such a subset (we will double-count like crazy in this proof - we only need an upper bound).

$$E(\text{number of cycles of length} \leq k) \leq \sum_{l=1}^{k} \binom{n}{l} l! p^l \leq \sum_{l=1}^{k} \frac{n^l}{l!} l! p^l$$

$$= \sum_{l=1}^{k} (np)^l = \sum_{l=1}^{k} (n^{1/2k})^l \leq 2\sqrt{n} \qquad \text{(assuming } n \text{ is large enough s.t. } n^{1/(2k)} \leq 1/2)$$

Note that we haven't proven anything yet about the girth of this graph. We just need this property later on. We first look at how many independent sets of size at least $\frac{n}{2k}$ there are in this graph.

Let $r = \frac{n}{2k}$. Fix some subset S with r vertices. The chance that S is an independent set is $(1-p)^{\binom{r}{2}}$. There are $\binom{n}{r}$ subsets of size r. We can therefore upper-bound the chance that the graph contains an independent set of size at least r:

$$\begin{aligned} P(\text{there is some independent set of size} \geq r) &\leq \binom{n}{r}(1-p)^{\binom{r}{2}} \\ &\leq 2^n e^{-pr^2/4} \\ &= 2^n e^{-\frac{n^{1/2k}}{n}\frac{n^2}{4k^2}} \\ &= e^n e^{-\frac{n^{1+1/2k}}{k^2}} \end{aligned}$$

We can make this last expression as small as we want (say $\leq 1/50$) by making n large enough.

Now, we combine the two results we have so far. By Markov's inequality, P(there are more than $50 \cdot 2\sqrt{n}$ cycles of length at most k) $\leq \frac{1}{50}$. As we also have that there is a very small chance that there is an independent set of size at least r for large enough n, we conclude that there must exist a graph with both properties, that is, it doesn't have an independent set of size at least r and it has at most $100\sqrt{n}$ cycles of length at most k.

We are not quite done yet, as we still have some short cycles in our graph. For every cycle of length $\leq k$, we remove one vertex in the cycle, which makes the graph free of such cycles. We need to remove at most $100\sqrt{n}$ vertices for this. This gives us a graph G' with girth at least k. There was no independent set of size at least $r = \frac{n}{2k}$ in the original graph G, the maximum independent set in G' is also at most $n/2k$ (note that any independent set of G' is also an independent set of G). But, choosing n large enough, $\frac{n}{2k} \leq (n - 100\sqrt{n})/k = \frac{n'}{k}$, where n' is the number of vertices in G'. Thus $\chi(G') \geq k$. □

This last method is called 'the probabilistic method with alterations': we prove the existence of some graph that is quite nearly what we want, and then we change it a bit so it does have the properties we want.

References

[1] `http://en.wikipedia.org/wiki/Miller-Rabin_primality_test`

[2] `http://en.wikipedia.org/wiki/AKS_primality_test`