LTCC Course: Graph Theory January-February 2012
§4 Probabilistic Methods and Random Graphs

Jozef Skokan*

February 6, 2012

1 References

Chapter 11 of Diestel is a good source for most of what we will cover in this course. Rather more material is to be found in Chapter VII of Bollobás.

For those interested in exploring this topic further, there are some excellent textbooks devoted entirely to this subject.

The classic text on random graphs is:

- B. Bollobás, *Random Graphs*, Cambridge Univerity Press, 2nd Edition (2001). See `http://www.cup.cam.ac.uk/us/catalogue/catalogue.asp?isbn=9780521797221`

The first edition, from 1985, is also a very good source, but the 2nd edition is naturally more up-to-date.

Another excellent book with the same title is:

- S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley (2000). See `http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471175412,descCd-authorInfo.html`

Finally, another classic text with a rather broader scope is:

- N. Alon and J. Spencer, *The Probabilistic Method*, Wiley, 3rd Edition (2008). See `http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470170204.html`

Earlier editions are also excellent.

2 Random Methods

Is there, for all large values of n, a graph on n vertices with no clique of size $5 \log n$, and no independent set of size $5 \log n$?

Is there a graph with some number n of vertices, no cycles of length less than 100, and no independent sets of size greater than $n/100$? Note that such a graph has chromatic number at least 100.

Is there, for all large values of n, a graph on n vertices with maximum degree at most $n^{2/3}$, and, for every pair (U, V) of sets of vertices with $|U|, |V| \geq n^{1/2}$, there is an edge from U to V?

The answer to each of these questions is yes, but how might one go about proving that?

*These notes were written by Graham Brightwell for LTCC Graph Theory Courses in 2009-11.

The natural first reaction on being confronted by questions like these is to try and *construct* graphs with the required combination of properties. In each case, one of the two required properties demands that the graph have rather few edges, and the other demands rather many. It's hard to strike a balance, and any attempt to base a construction around some nice structure seems doomed to failure.

For each of the problems above, by far the best way to solve the problem is not to give an explicit construction at all, but instead to "construct" the graph "at random", and show that, with positive probability, the random graph constructed has the required combination of properties.

As a first, striking, example, let's consider the first question. Here, what we do is to build our random graph in the most naive possible way. We take some (large) set of vertices, say $V = [n] = \{1, \dots, n\}$. Then, for each pair of vertices, we toss a fair coin, and put an edge between the pair if we get a head.

Let's calculate the probability that a particular set C of c vertices forms a clique. This means that each of the $\binom{c}{2}$ coin-tosses corresponding to the pairs in the clique was a head, an event with probability $2^{-\binom{c}{2}}$.

Now, the *expected* number of cliques of size c is given by:

$$\binom{n}{c} 2^{-\binom{c}{2}} \le \frac{n^c}{c!} 2^{-\binom{c}{2}} \le \left(\frac{ne}{c2^{(c-1)/2}} \right)^c .$$

If $c \ge 2\log_2 n$, then $n \le 2^{c/2}$, and so the expected number of c-cliques is at most $(e\sqrt{2}/c)^c$. For $c \ge 5$, this expectation is at most $1/3$. This implies that the probability that the random graph has a clique of size c is at most $1/3$. The same calculation shows that the probability that the graph has an independent set of size c is also at most $1/3$.

So, with probability at least $1/3$, the random graph on $n \ge 5$ vertices has neither a clique nor an independent set of size as large as $2\log_2 n$.

This delightfully simple argument was first given by Paul Erdős in 1947, and it is one of the first instances of the successful use of the "probabilistic" method in combinatorics.

One can, and indeed Erdős did, recast this entire proof as a "counting argument". Of the $2^{\binom{n}{2}}$ graphs with vertex set $[n]$, the number of them in the class A_C with a clique or independent set on the set C of size c is ..., and then effectively the same calculation shows that there are some graphs in none of the sets A_C.

The result we've proved is usually stated in terms of *Ramsey numbers*, a topic we'll return to next week. It says that, for all $c \in \mathbb{N}$, there is a graph on $n = \frac{c}{e} 2^{(c-1)/2}(1 + o(1))$ vertices containing neither a clique nor an independent set of size c. (Here the $o(1)$ refers to a term that tends to zero as $c \to \infty$.) This is exactly what it means to say that the *Ramsey number* $R(c, c)$ is at least $\frac{c}{e} 2^{(c-1)/2}(1 + o(1))$.

3 Simple Random Arguments

This section is devoted to a couple of famous combinatorial results, and slick proofs via random arguments. (Both of these examples can be found in Alon and Spencer.)

3.1 Sperner's Theorem

The first result isn't about graphs at all. (Apologies.)

Consider the Boolean cube Q^n. This is the collection of all 2^n subsets of the set $[n]$, ordered by inclusion. An *antichain* or *Sperner family* in Q^n is a family of subsets of $[n]$, no one of which is contained in another.

One way to construct a large antichain is to take the family of all subsets of $[n]$ of size exactly $\lfloor n/2 \rfloor$. Is this the largest? A famous theorem of Sperner from 1927 says that it is.

A stronger result is the so-called *LYM Inequality*, which says that, if $\mathcal{A}$ is any antichain in Q^n, then

$$\sum_{A \in \mathcal{A}} \binom{n}{|A|}^{-1} \leq 1.$$

This implies Sperner's Theorem, as the largest binomial coefficient is $\binom{n}{\lfloor n/2 \rfloor}$.

Here's a snappy proof of the LYM Inequality.

Take any antichain $\mathcal{A}$ in Q^n. Now take a random ordering $(x_1, x_2, \dots, x_n)$ of $[n]$, and consider the *chain* $\mathcal{C}$ of sets $C_0, C_1, \dots, C_n$, where $C_j = \{x_1, \dots, x_j\}$. Notice that, for each pair of elements of $\mathcal{C}$, one is contained in the other; thus $\mathcal{C}$ intersects $\mathcal{A}$ in at most one element. Hence it is certainly the case that $\mathbb{E}|\mathcal{C} \cap \mathcal{A}| \leq 1$.

Now, for each $A \in \mathcal{A}$, the probability that A is in the random chain $\mathcal{C}$ is exactly $1/\binom{n}{|A|}$, since $\mathcal{C}$ includes exactly one set with $|A|$ elements, and it's equally likely to be any of them.

So we have

$$1 \geq \mathbb{E}|\mathcal{A} \cap \mathcal{C}| = \sum_{A \in \mathcal{A}} \mathbb{P}(A \in \mathcal{C}) = \sum_{A \in \mathcal{A}} \frac{1}{\binom{n}{|A|}},$$

and we are done.

3.2 Turán's Theorem

Turán's Theorem is one of the most celebrated theorems of graph theory. It's the cornerstone of what is called "Extremal Graph Theory". This consists of problems and results of the following type: if we restrict some parameter of a graph, how large or small can some other parameter be?

Turán's Theorem is usually stated as follows. Given parameters n and k define the *Turán Graph* $T_{n,k}$ to be a graph on n vertices, divided into k classes as evenly as possible, with edges between two vertices if and only if they are in different classes. Notice that this graph contains no clique of size $k+1$. Let $t_{n,k}$ denote the number of edges of $T_{n,k}$.

Theorem 3.1. *If G is a graph on n vertices with no clique of size $k+1$, then $|E(G)| \leq t_{n,k}$, with equality if and only if G is the Turán graph $T_{n,k}$.*

Any decent textbook on graph theory contains at least one straightforward proof of Turán's Theorem.

We'll start by re-phrasing the result. It's easy to see that Turán's Theorem is equivalent to the following. For any n and k, let $r_{n,k} = \binom{n}{2} - t_{n,k}$, the number of edges in the complement $R_{n,k}$ of $T_{n,k}$. (The *complement* of a graph G is the graph G^c obtained by swapping edges and non-edges.) Note that $R_{n,k}$ is a disjoint union of cliques, the sizes of which differ by at most 1.

Theorem 3.2. *Suppose G has n vertices and $r_{n,k}$ edges. Then G has an independent set of size k, and G has an independent set of size $k+1$ unless G is $R_{n,k}$.*

We'll prove this via the following lemma, which has a nice probabilistic proof. Here, $\alpha(G)$ denotes the *independence number* of G, i.e., the size of the largest independent set in G.

Lemma 3.3. *For every graph $G = (V, E)$,*

$$\alpha(G) \geq \sum_{v \in V} \frac{1}{d(v)+1},$$

with equality if and only if G is a disjoint union of cliques.

Proof. We construct an independent set I in G using the "random greedy" algorithm, as follows. We put the vertices of G in a random order, and consider them in that order: when we consider vertex v, we put it into our independent set I whenever we can, i.e., if we have not so far taken any neighbour of v into the set.

The probability that vertex v is taken into I is *at least* the probability that, among v and its $d(v)$ neighbours, v comes first in the order. So this probability is at least $1/(d(v)+1)$. Moreover, if there are vertices u, w such that uv and uw are edges, but vw is not, then we have strict inequality, since, with positive probability, the first three vertices in the order are w, u, v, in which case v is taken into the independent set even though its neighbour u appears before it in the order.

So the expected size of the randomly generated independent set is at least $\sum_{v\in V} 1/(d(v)+1)$, with strict inequality whenever there are three vertices u, v, w with uv and vw edges but uw not. In other words, we have strict inequality unless the graph is a disjoint union of cliques.

The following trivial observation is very often the driving force behind a probabilistic argument: if the *expected* size of the random independent set is at least this large, then *there is* at least one order of the vertices that generates an independent set of that size. In particular, G has an independent set of size at least the given sum.

It remains to prove that, if G *is* a disjoint union of cliques, then $\alpha(G)$ is equal to the given sum. This is straightforward: basically note that every clique contributes one to the sum, and $\alpha(G)$ is indeed the number of cliques. □

I've set this argument out in detail. To an audience adept at the use of random methods, "take a random order of the vertices, and construct an independent set using the greedy algorithm" is pretty much a complete proof.

Now to deduce Turán's Theorem, in the form stated in Theorem 3.2.

Proof. Suppose that G has n vertices and $r_{n,k}$ edges. We note that $\sum_{v\in V} d(v) = 2r_{n,k}$. (If you haven't seen this before, treat it as an – easy – exercise.)

We now seek to minimise $Q = \sum_{v\in V} 1/(d(v)+1)$ subject to the constraint on $\sum_{v\in V} d(v)$. The way to do this is to take the $d(v)$ to be as close together as possible, subject to the constraint. (The most elementary and convincing way to see this is to suppose that two of the $d(v)$ differ by 2 or more, and show that Q is decreased – strictly – on increasing the smaller value and decreasing the larger by 1.)

Taking the $d(v)$ as close together as possible gives exactly the degrees of the vertices in $R_{n,k}$. For $R_{n,k}$, each of the k components is a clique, whose vertices contribute exactly 1 to the sum Q. Therefore the minimum value of Q, for a graph on n vertices with $r_{n,k}$ edges, is exactly k.

Now it's just a matter of tidying up. By Lemma 3.3, and the above calculation, we have:

$$\alpha(G) \geq \sum_{v\in V} \frac{1}{d(v)+1} \geq k.$$

We can only have equality if both (a) the graph G is a disjoint union of k cliques, and (b) the vertex degrees are as equal as possible. This means that $R_{n,k}$ is the only graph achieving equality. □

4 Girth and Chromatic Number

This section concerns another classic application of the probabilistic method. It's covered very well in Diestel, for instance, and I refer the reader there for full details.

The *girth* of a graph G is the length of the shortest cycle in G. A graph with high girth, i.e., no short cycles, is "locally" like a tree. Specifically, suppose the girth is at least g, and we take

$2k < g$ and let $N_k(v)$ be the number of vertices at distance at most k from vertex v, then all the sets $N_k(v)$, for $v \in V(G)$, induce trees. In particular, all the induced subgraphs on the sets $N_k(v)$ are 2-colourable. Is it possible for the whole graph to have a large chromatic number?

How can we prove that the chromatic number of a graph is large? More often than not, we prove that the graph has no large independent sets. (The set of vertices receiving any given colour is an independent set.)

We shall prove the following result.

Theorem 4.1. *For each $k \in \mathbb{N}$, there exists a graph on some number n_0 of vertices with girth greater than k and no independent set of size $\lceil n_0/k \rceil$.*

The proof of this result (also due to Erdős, this time from 1959) is slightly more complicated than the one about Ramsey numbers. For a start, our random process involves the tossing of *biassed* coins. Again we fix a vertex set $[n]$: now we put an edge between each pair of vertices with probability p, all choices made independently. Here $p = p(n)$ is a function of n that we can choose to suit our needs. What we have defined here is the standard model $\mathcal{G}(n,p)$ of *random graphs*. We use $G(n,p)$ to denote a random graph chosen according to this method.

Proof. (Sketch)

For this application, we fix a natural number k and take $p(n) = n^{-1+\varepsilon}$, where $0 < \varepsilon < 1/k$.

It is easy to show that, for sufficiently large n, the probability that $G(n,p)$ has an independent set of size $n/2k$ is at most $1/2$. Indeed, this is true with plenty to spare – see Diestel, Lemma 11.2.1. – the calculation is essentially the same as we did in §2.

Now we consider the expectation of $N = N(G)$, the number of cycles in G of length at most k.

$$\mathbb{E}N = \sum_{g=3}^{k} \frac{n(n-1)\cdots(n-g+1)}{2g} p^g \leq \frac{1}{2}(k-2)(np)^k.$$

Notice that $(np)^k = n^{\varepsilon k}$, so $\mathbb{E}N = o(n)$. Now, N is a non-negative random variable, and we can use *Markov's Inequality*: $\mathbb{E}N \geq q\mathbb{P}(N \geq q)$, so $\mathbb{P}(N \geq q) \leq \mathbb{E}N/q$, for any q. Applying this with $q = n/2$, we find that $\mathbb{P}(N \geq n/2) < 1/2$ for sufficiently large n.

This means that, with positive probability, $G(n,p)$ has the following two properties: $N(G) < n/2$, and G has no independent sets of size $n/2k$.

This means that there is some graph H with both these properties.

Take such an H, and delete $n/2$ of its vertices, including at least one vertex from each cycle of length at most k. This leaves us with a graph H' on $n_0 = n/2$ vertices, with girth greater than k. Moreover, as H has no independent sets of size n_0/k, neither does H'. □

5 Tools from Probability

In the proofs we've given, we've used no sophisticated probabilistic techniques. If you like, we've used nothing more than the linearity of the expectation operator.

The theory of random graphs developed in this way. To begin with, surprising results could be proved using just the basic ideas of probability: little more than use of the expectation and variance of random variables, together with (sometimes) intricate combinatorial arguments. Then, increasingly, more advanced tools and techniques from probability, often developed with combinatorial applications in mind, were used to great effect. The interested reader is referred to the textbooks mentioned at the top of these notes.

We'll give one example in this direction.

Let's go back and think about what we did in the section on Ramsey numbers. We wanted to construct a graph with vertex set $[n]$ containing no cliques or independent sets of size c. We chose the parameters so that the expected number of "bad" sets was less than 1, and concluded that, with positive probability, there were no bad sets.

Now suppose we choose the parameters so that the expected number of bad sets is small: small enough that most vertices are not in a bad set, say. In this case, should we not expect to be able to "patch up" the graph, adjusting an edge here or there to kill all the bad sets without introducing any new ones? Or perhaps we might be able to prove that the number of bad sets is asymptotically a Poisson random variable, and so is zero with positive probability.

It turns out not to be possible to carry out either of these plans exactly as stated, but these ideas gave rise to a number of tools that can be used in circumstances like the ones we have here.

We'll state a simple version of the *Lovász Local Lemma.* [See Alon and Spencer for a proof, and a number of more general versions.]

Theorem 5.1. *Let $A_1, \dots, A_N$ be events in a probability space. Suppose, for each i, that there is a subset D_i of the index set $[n]$, so that A_i is independent of $\{A_j : j \notin D_i\}$. Suppose that $\mathbb{P}(A_i) \le p$ and $|D_i| \le d$ for each i, and that $ep(d+1) \le 1$. Then the probability that none of the A_i occur is positive.*

To apply this to improve the bound on Ramsey numbers, take the probability space to be the space $\mathcal{G}(n, 1/2)$, take the events A_C to be the events that the set $C \subset [n]$, with $|C| = c$, induces either a clique or an independent set, and calculate.

For a given set C of size c, let D_C be the family of all c-sets intersecting C in at least 2 elements. The event A_C depends on the edges inside C; the events $\{A_B : B \notin D_C\}$ depend on edges not inside C, so the hypotheses are satisfied.

Calculation now gives:

Theorem 5.2. *If $e\left(\binom{c}{2}\binom{n}{c-2}+1\right)2^{1-\binom{c}{2}} < 1$, then there is a graph on n vertices with neither a clique nor an independent set of size c – in other words, $R(c,c) > n$.*

Thus $R(c,c) > \frac{\sqrt{2}}{e}(1+o(1))c2^{c/2}$.

This application to Ramsey numbers, improving the simple bound from Section 2 by a factor of 2, is due to Spencer. What is astonishing is that no-one has managed to do better: this is the best known lower bound on $R(c,c)$, for all but very small values of c. The best known upper bound is of the form $(4-o(1))^c$.

6 The shape of random graphs

In order to understand when and how random graphs can be useful, it is necessary to understand roughly the typical "shape" of $G(n,p)$ for various ranges of p.

Here are just a few samples of the types of results that have been proved. All the results are stated loosely. Precise versions would make statements about the probability that certain properties hold tending to 0 or 1 as $n \to \infty$.

For $p < c/n$, $c < 1$, the random graph $G(n,p)$ has no component of size greater than $A(c)\log n$; all components are trees except possibly a few components with one cycle.

For $p = c/n$, $c > 1$, there is one "giant component" containing about $t(c)n$ vertices; all other components have size at most $A(c)\log n$, and have at most one cycle.

For $p \geq \frac{1}{n}(\log n + k \log\log n + \omega(1))$, the random graph has minimum degree at least k, and is k-connected. For $k = 1$, it has a perfect matching; for $k = 2$, it has a Hamilton cycle.

For $p = n^{-t}$, $0 < t < 1$, it is possible to determine which small graphs appear as subgraphs of $G(n, p)$ with probability close to 1, and which do not. (See Exercise 4.)

7 Exercises

There are plenty more exercises in the textbooks!

1. For $k \in \mathbb{N}$, a graph $G = (V, E)$ has *Property* S_k if, for every pair (A, B) of disjoint k-element subsets of V, there is a vertex x of the graph that is adjacent to every vertex of A and no vertex of B.
(a) Find a graph with property S_1.
(b) Show that, for each $k \in \mathbb{N}$, there is a graph with property S_k.

2. Fill in the details in the proof of Theorem 5.2.

3. A k-uniform *hypergraph* is a pair $H = (V, E)$, where V is a set of *vertices*, and E is a family of k-element subsets of V. (So a 2-uniform hypergraph is just a graph.) A hypergraph $H = (V, E)$ has *Property B* if V can be partitioned into two subsets V_1 and V_2 in such a way that no edge is entirely contained within one of the two sets.
(a) Show that, if $H = (V, E)$ is a k-uniform hypergraph with $|E| < 2^{k-1}$, then H has property B.
(b) Show that, if $H = (V, E)$ is a k-uniform hypergraph such that each edge in E intersects at most d others, and $e(d + 1) \leq 2^{k-1}$, then H has property B.

4. (a) Let $p = n^{-t}$, for $0 < t < 1$, and let k be a fixed natural number. Write down an expression for the expected number of k-cliques in $G(n, p)$. Hence show that, if $t > 2/(k - 1)$, the probability that $G(n, p)$ contains a k-clique tends to zero as $n \to \infty$.

It is also true that, if $t < 2/(k - 1)$, then the probability that $G(n, p)$ contains a k-clique tends to one as $n \to \infty$: to prove this, one needs to work with the variance of the number of k-cliques.
(b) Let H denote the graph on five vertices a, b, c, d, e with seven edges: a, b, c, d form a clique, and de is also an edge. For $p = n^{-7/10}$, find the expected number of copies of H in $G(n, p)$. What is

$$\lim_{n\to\infty} \mathbb{P}(G(n, p) \text{ contains a copy of } H)?$$

(c) There is a parameter $b(H)$ of graphs such that, if $p = n^{-t}$ and $t > b(H)$, then the probability that $G(n, p)$ contains a copy of H as a subgraph tends to zero, while if $p = n^{-t}$ and $t < b(H)$, then this probability tends to 1. Based on the calculations in this question, what do you think this parameter $b(H)$ might be?

5. Set $p = n^{-2/5}$, and consider a random graph $G = G(n, p)$.
(a) Show that the degree of any fixed vertex v has a Binomial distribution, and find an upper bound on the probability that this degree is greater than or equal to $n^{2/3}$. [You may need to look up some estimates on the tails of the distribution of a Binomial random variable.]
(b) Show that the probability that the maximum degree of G is at most $n^{2/3}$ is at least 2/3.
(c) Show that, with probability at least 2/3, for every pair (U, V) of subsets of $V(G)$, with $|U|, |V| \geq n^{1/2}$, there is an edge from U to V.
(d) What can you deduce from (b) and (c)?