Random and Quasirandom Graphs — Instructor: Padraic Bartlett

Lecture 2: Modeling a Random Graph

Week 4 — *Mathcamp 2012*

Yesterday, we introduced the idea of a **probability space**. Throughout this course, we will almost always study the space $G_{n,1/2}$:

- The set for $G_{n,1/2}$ is the collection of all graphs on n vertices.
- The probability measure for $G_{n,1/2}$ is the measure that says that all graphs are equally likely: i.e. for any $H \in G_{n,1/2}$, $\mathbb{P}(H) = \frac{1}{2^{\binom{n}{2}}}$.

This model corresponds to the following process for generating a random graph:

- Take n labeled vertices $\{1, \dots n\}$.
- For each unordered pair of vertices $\{a, b\}$, flip a fair coin. If it comes up heads, connect these vertices with an edge; otherwise, do not.

To see why, simply consider the probability of generating any given labeled graph in the model above; because it involves making $\binom{n}{2}$ choices on whether an edge exists or not, the probability of this graph occurring is $\left(\frac{1}{2}\right)^{\binom{n}{2}}$, i.e. the same probability that we gave above in our probability space.

Given this model, a natural sequence of questions to ask is "what properties is a random graph likely to have?" For example, consider counting the number of edges in a random graph, or triangle subgraphs, or connectivity, or other such properties: what should we expect these values to be?

We study these properties in this lecture:

1 Properties of the Random Graph

As a warmup, we start with the following question:

Question 1 *Let $e(H)$ denote the function that takes in a graph H and outputs the total number of edges in H. What is the expected value of e over $G_{n,1/2}$? In other words, if you take a random graph on n vertices under our model, how many edges would you expect to see on average?*

Answer. We calculate the expected value of this function e, using the definition we came up with on yesterday:

$$\mathbb{E}(e) = \sum_{H \in G_{n,1/2}} e(H) \cdot \mathbb{P}(H) = \sum_{H \in G_{n,1/2}} e(H) \cdot \frac{1}{2^{\binom{n}{2}}} = \left(2^{\binom{n}{2}} \cdot \frac{\binom{n}{2}}{2} \right) \cdot \frac{1}{2^{\binom{n}{2}}} = \frac{\binom{n}{2}}{2}.$$

Alternately, if you think of the model we have for our random graph, this is pretty clear: if you have n vertices and you're flipping a coin for each pair of them, you'd expect to see $\frac{\binom{n}{2}}{2}$ many heads, i.e. $\frac{\binom{n}{2}}{2}$ many edges.

In general, we will switch between using either of these arguments, depending on which is easier for us to use and calculate.

More interestingly, we can ask how likely our random graph is to contain a given **structure**. For example, instead of just asking how many edges our graph has, we could ask how many distinct labeled triangles occur as subgraphs of our graph:

Question 2 *Let $t(H)$ denote the function that takes in a graph H and outputs the total number of distinct labellings of triangles in H. What is the expected value of t over $G_{n,1/2}$? In other words, if you take a random graph on n vertices under our model, how many triangles would you expect to see on average?*

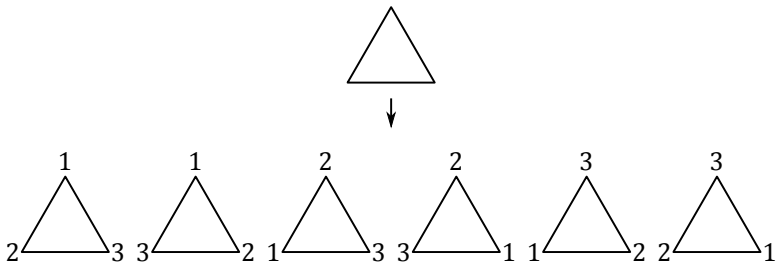
Answer. Let's use the second description of our random graph model, where we think of it as a process: i.e. where we construct a random graph by picking n vertices and flip a coin for each pair. Under this model, the total number of triangles we would expect to see is

$$\begin{aligned}
&\sum_{a,b,c \text{ distinct labeled vertices}} \mathbb{P}(abc \text{ is a triangle}) \\
=&\sum_{a,b,c \text{ distinct labeled vertices}} \mathbb{P}(ab \text{ is an edge}) \cdot \mathbb{P}(bc \text{ is an edge}) \cdot \mathbb{P}(ac \text{ is an edge}) \\
=&\sum_{a,b,c \text{ distinct labeled vertices}} \frac{1}{8} \\
=&n \cdot (n-1) \cdot (n-2) \cdot \frac{1}{8},
\end{aligned}$$

because there are $n(n-1)(n-2)$ many ways to choose three vertices along with a labeling, and the probability that that individual triple of edges forms a triangle is $\frac{1}{8}$.

For $n = 3$, this formula predicts that we'll see $\frac{3!}{0!} \cdot \frac{1}{8} = \frac{3}{4}$ triangles on average. However, if we look at the collection of all graphs on three vertices, we see that there are eight possible graphs, only one of which is a triangle; so we'd expect the probability here to be $\frac{1}{8}$. Why is there this incongruity between our answers?

Well, because in our problem we're counting the number of **distinct labellings of triangles**, not just triangles! To be precise, we are counting all of the ways of mapping the three numbers $1, 2, 3$ onto a set of three vertices in a random graph, so that the resulting image under this mapping is a triangle. Therefore, whenever there is a triangle in our random graph, we actually wind up counting it six times:



This accounts for the discrepancy between our two counts: while we might have expected $\frac{1}{8}$, we will really get $\frac{6}{8}$ if we want to count distinct labeled occurrences of our triangle.

If we want to ignore this question of distinct labellings, and just count the number of different triangles, we can just divide our formula by the number of symmetries of a triangle, because this is the same thing as the number of ways of labeling a triangle and having it remain a triangle. In other words, a formula for counting distinct triangles without caring about labellings is

$$n \cdot (n-1) \cdot (n-2) \cdot \frac{1}{6} \cdot \frac{1}{8}.$$

For $n = 4$, this predicts that your typical graph will contain about half a triangle. If you enumerate all of the 64 graphs on 4 vertices and count, you can see that there are 32 triangles, and therefore that this prediction aligns with reality there.

We can generalize the proof method above dramatically, from triangles to *any* given subgraph:

Question 3 *Take any labeled graph L on s vertices. Let $l(H)$ denote the function that takes in a graph H and outputs the total number of distinct labeled induced subgraphs isomorphic to L in H. What is the expected value of l over $G_{n,1/2}$? Again, in other words,if you take a random graph on n vertices, how many copies of L would you expect to find in it?*

Answer. Again, let's use the "pick n vertices and flip a coin for each pair" model. Under this model, we have

$$\begin{aligned}
&\sum_{\text{ways to pick } s \text{ distinct labeled vertices}} \mathbb{P}(\text{these labeled } s \text{ vertices form a copy of } L) \\
=&\sum_{\text{ways to pick } s \text{ distinct labeled vertices}} \frac{1}{2^{\binom{s}{2}}} \\
=&\frac{n \cdot (n-1) \cdot \ldots (n-s+1)}{2^{\binom{s}{2}}},
\end{aligned}$$

because there are $n \cdot (n-1) \cdot \ldots (n-s+1)$ many ways to pick out s vertices and label them from a set of n vertices, and the probability that these vertices in this order correspond to a copy of L as written is $\frac{1}{2^{\binom{s}{2}}}$.

Again, this process counts labeled occurrences of L, as opposed to occurrences of L. If you want to just count the occurrences of L, you want the formula

$$\frac{n \cdot (n-1) \cdot \ldots (n-s+1)}{2^{\binom{s}{2}}} \cdot \frac{1}{\text{Aut}(L)}.$$

Here, $\text{Aut}(L)$ is the number of "symmetries" possessed by the graph L. More specifically, it is the **group of automorphisms** of L, where a **graph automorphism** is the total

number of ways of permuting the vertices of a graph without changing the graph itself: i.e. after an automorphism, two vertices i, j should have an edge between them if and only if they had an edge between them before the automorphism. For example, the total number of automorphisms of a triangle is 6; no matter how you permute the three vertices of a triangle, you get a triangle. In general, the total number of automorphisms of a K_n is $n!$; any permutation of their vertices will not change whether this graph looks like a K_n. For a less trivial example, the graph on three vertices $\{1, 2, 3\}$ consisting of one edge from 1 to 2 has just one automorphism, namely switching vertices 1 and 2.

This process lets us count the number of occurrences of tons of different kinds of graphs! However, it only works on finding induced copies of some given graph as a subgraph: i.e. subgraphs where we've completely determined which edges we want to exist and which others we don't want to exist within the vertices we've picked out.

We often don't want to be this picky. For example, a quantity we could want to count in a random graph is the number of k-cycles (not necessarily induced!) we'd expect to see:

Question 4 *Take any k-cycle, and let $c(H)$ denote the function that takes in a graph H and outputs the total number of distinct labeled copies of C_k (not necessarily induced) in H. What is the expected value of c over $G_{n,1/2}$?*

Answer. Again, we have

$$\begin{aligned}
&\sum_{\text{ways to pick } k \text{ distinct labeled vertices}} \mathbb{P}(\text{these labeled } s \text{ vertices form a copy of } C_k) \\
=&\sum_{\text{ways to pick } k \text{ distinct labeled vertices}} \left(\text{given } v_1 \dots v_k,\ \prod_{i=1}^{k} \mathbb{P}((v_i, v_{i+1}) \text{ is an edge})\right) \\
=&\sum_{\text{ways to pick } k \text{ distinct labeled vertices}} \frac{1}{2^k} \\
=&n \cdot (n-1) \cdot \ldots (n-k+1)\frac{1}{2^k}.
\end{aligned}$$

Like the above examples, this process counts labeled cycles; if you want to ignore the labeling, simply divide the formula above by the number of symmetries of a k-cycle, $|D_{2k}| = 2k$.

Using the methods we've illustrated above, we can easily count the existence and expected number of occurrences of many different kinds of structure within a random graph. Moving on from this, another quantity that seems worthwhile to study is the collection of **eigenvalues** for a random graph: as I mentioned rather briefly in our last class, they can often give you a large amount of information about your graph.

Question 5 *Given a graph H on n vertices, let $\lambda_1, \dots \lambda_n$ be the n eigenvalues corresponding to the distinct orthogonal eigenvectors of $A(H)$, the adjacency matrix of H. Order them such that $|\lambda_1| \geq \ldots \geq |\lambda_n|$. Suppose that H is a random element of $G_{n,1/2}$. What are the expected values of these eigenvalues?*

Answer. First, let's study λ_1, the largest eigenvalue in terms of magnitude of our graph H. Because $A = A(H)$ is symmetric and therefore we have n orthogonal eigenvectors for H, we know that for any vector $\mathbf{v} \in \mathbb{R}^n$, we have

$$||A\mathbf{v}|| \leq |\lambda_1| \cdot ||\mathbf{v}||.$$

This is because (in a sense) λ_1 is the "largest" direction in which A can stretch space; therefore, if we multiply any vector $\mathbf{v}$ by our matrix, it cannot be stretched as much as it would be if it were pointing in the direction given by the eigenvector corresponding to λ_1.

In particular, if we let $\mathbf{v} = (1, 1, \ldots 1)$, we have $||A(1, \ldots 1)|| = |\lambda_1| \cdot ||(1, \ldots 1)||$. But

$$A(1, 1, \ldots 1) = (\deg(v_1), \deg(v_2), \ldots \deg(v_n)),$$

where the v_i's are the vertices of H, because the dot product of $(1, 1 \ldots 1)$ with any row of A just returns the total number of 1's in that row: i.e. the number of edges leaving that corresponding vertex.

We can easily calculate the expected value of any $\deg(v_i)$: it's just $\frac{n-1}{2}$, because we flip a coin for every edge. Therefore, if we use this observation, we have the observation that we will typically expect to see

$$A(1, 1, \ldots 1) = (\deg(v_1), \deg(v_2), \ldots \deg(v_n)) = \frac{n-1}{2} \cdot (1, \ldots 1),$$

and therefore that $|\lambda_1| \geq \frac{n-1}{2}$.

How about the other eigenvalues? Also, this is just a lower bound: can we improve this to an upper bound as well? To answer both of these questions, recall the theorem on walks we proved yesterday:

Theorem 6 *Suppose G is a graph with vertex set $\{1, \ldots n\}$ with adjacency matrix A. Then the (i, j)-th entry of A^k denotes the number of distinct walks of length k from i to j.*

In particular, this tells us that for our random graph H, the sum of the entries on the diagonal of A^4 correspond to all of the walks of length 4 that start and end at the same point! There are three kinds of these walks:

1. The walks that start at some vertex v, go to another vertex w, return to v, go to another vertex w', and return to v again. These correspond to the number of (potentially not distinct) pairs of edges that you can pick leaving a vertex v. There are at most $(deg(v))^2 \leq n^2$ many such walks, starting at any vertex v, and therefore less than n^3 many such walks in total.

2. The walks that start at some vertex v, go to another vertex w, travel to a third vertex x, and then return to w and then return from there to v. In other words, these are all of the paths of length 2 starting from v, that then return along that same path to v. There are at most $\deg(v) \cdot (\max_{w \in n(v)}(\deg(w) - 1) \leq n^2$ many such walks starting at v, and therefore $\leq n^3$ many walks in total.

3. The actual 4-cycles starting and ending at v. By our earlier calculations, there are

$$n(n-1)(n-2)(n-3) \cdot \frac{1}{16}$$

many such labeled 4-cycles in our graph in total.

Therefore, the sum of the entries along our diagonal is at most

$$n(n-1)(n-2)(n-3)\cdot\frac{1}{16}+2n^3=\frac{n^4+26n^3+11n^2-6n}{16}.$$

However, notice that the eigenvalues of A^4 are simply the fourth powers of the eigenvalues $\lambda_1\dots\lambda_n$, and that the trace of a matrix is the sum of its eigenvalues! Therefore, we can combine our upper bound with these observations to get

$$\sum_{i=1}^{n}\lambda_i^4\leq\frac{n^4+26n^3+11n^2-6n}{16}.$$

Because $|\lambda_1|\geq\frac{n-1}{2}$, we have $\lambda_1^4\geq\frac{n^4-4n^3+4n^2-4n+1}{16}$, and therefore that

$$\begin{aligned}\sum_{i=2}^{n}\lambda_i^4\leq&\frac{n^4+26n^3+11n^2-6n}{16}-\frac{n^4-4n^3+4n^2-4n+1}{16}\\=&\frac{30n^3+7n^2-2n+1}{16}.\end{aligned}$$

In particular, this tells us that

$$\begin{aligned}\lambda_2^4\leq&\frac{30n^3+7n^2-2n+1}{16}\\\Rightarrow|\lambda_2|\leq&\left(\frac{30n^3+7n^2-2n+1}{16}\right)^{1/4}\leq 2^{1/4}n^{3/4}.\end{aligned}$$

In particular, this is a bound that grows much slower than n (i.e. is $o(n)$.) Therefore, because $|\lambda_2|\geq|\lambda_i|, i\geq 2$, we have that all of the eigenvalues $\lambda_i, i\neq 1$ are $o(n)$ in growth, and in particular have the bound we derived above.

Furthermore, we can also use this bound to get that

$$\begin{aligned}\sum_{i=1}^{n}\lambda_i^4\leq&\frac{n^4+26n^3+11n^2-6n}{16}\\\Rightarrow\lambda_1^4\leq&\frac{n^4+26n^3+11n^2-6n}{16}\\\Rightarrow\lambda_1\leq&\left(\frac{n^4+26n^3+11n^2-6n}{16}\right)^{1/4}.\end{aligned}$$

As n gets increasingly large, this bound becomes $\leq(1+\epsilon)\frac{n}{2}$, for any given constant $\epsilon>0$ that you care to pick, because ϵn^4 grows much much faster than $26n^3+11n^2-6n$. Therefore, we have that λ_1 is sandwiched in between $\frac{n-1}{2}$ and $\frac{n}{2}$, for large values of n.