Summary of Lecture 6

In which we study the spectrum of random graphs.

In past lectures we used the fact that, for every $p(n) > \frac{\log n}{n}$, there is a high probability that if we sample $G \sim G_{n,p(n)}$ and let A be its adjacency matrix, we have $||A - pJ|| \leq O(\sqrt{p(n) \cdot n})$. In particular, there is a high probability that $||A - J/2|| \leq O(\sqrt{n})$ when A is the adjacency matrix of a graph sampled from $G_{n,\frac{1}{2}}$.

Today we will prove a slightly weaker result, and give an overview of the proofs of the stronger results.

We will use the following basic facts from linear algebra. Let M be a real symmetric matrix: then M has real eigenvalues, and let us call them $\lambda_1, \ldots, \lambda_n$ so that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. By definition, $||M|| = |\lambda_1|$. We have

1. $\sum_i M_{ii} = \sum_i \lambda_i$, and this quantity is called the *trace* of the matrix.
2. The eigenvalues of M^k are $\lambda_1^k, \cdots \lambda_n^k$

From the above two facts we have that, if k is even,

$$\sum_i (M^k)_{ii} = \sum_i \lambda_i^k = \sum_i |\lambda_i|^k$$

and so

$$||M||^k \leq \sum_i (M^k)_{ii} \leq n \cdot ||M||^k$$

If we choose $k = \ln n$, we have

$$||M|| \leq \left(\sum_i (M^k)_{ii} \right)^{\frac{1}{k}} \leq e \cdot ||M||$$

Thus we can approximate the spectral norm of a matrix by a computation of the trace of a power of the matrix. If M is a random matrix, then computing $\mathbb{E} \sum_i (M^k)_{ii}$ will give us high-probability upper bounds to the spectral norm.

Let A be the adjacency matrix of a $G_{n,1/2}$ random graph; we let M be the random matrix $M := 2 \cdot (A - \mathbb{E}\, A)$. We see that M is such that the diagonal entries are zero, all other entries are equally likely to be $+1$ or -1, and, other than satisfying the constraints $M_{ij} = M_{ji}$, all entries are mutually independent.

We first notice that, by symmetry

$$\mathbb{E} \sum_i (M^k)_{ii} = n \cdot \mathbb{E}(M^k)_{1,1}$$

and by definition of matrix product and by linearity of expectation we have

$$\mathbb{E}(M^k)_{1,1} = \sum_{a_1,\ldots,a_{k-1} \in V} \mathbb{E}\, M_{1,a_1} \cdot M_{a_1,a_2} \cdots M_{a_{k-2},a_{k-1}} \cdot M_{a_{k-1},1}$$

For each entry (a,b), we have that $\mathbb{E}\, M^t_{a,b} = 0$ if t is odd and $\mathbb{E}\, M^t_{a,b} = 1$ if t is even, so the expressions whose average we want to compute above have average zero is there is a $\{a,b\}$ that occurs an odd number of times in the sequence $\{1,a_1\},\ldots,\{a_{k-1},1\}$, and the average is one if all pairs occur an even number of times in the sequence.

Thus, $\mathbb{E}(M^k)_{1,1}$ is equal to the number of sequences $1, a_1, \ldots, a_{k-1}, 1$ such that, in the sequence $\{1,a_1\},\ldots,\{a_{k-1},1\}$, each pair appears an even number of times.

We prove that such number is at most $2^k \cdot k^{k/2} \cdot n^{k/2}$ by showing how to represent each sequence using at most $k + \frac{k}{2}\log k + \frac{k}{2}\log n$ in the following way: each element a_i is represented either as $(0, a_i)$, which takes $1 + \log n$ bits, if a_i appears for the first time in the sequence at location i, and as $(0, j)$ otherwise, where $j < i$ is such that $a_i = a_j$, which requires $1 + \log k$ bits. Using the fact that the sequence can contain at most $k/2$ distinct pairs and hence at most $k/2$ distinct vertices (not counting the start) we get the bound.

This means that

$$\mathbb{E} \sum_i (M^k)_{ii} \leq 2^k \cdot k^{k/2} \cdot n^{1+k/2}$$

If we take $k = \lg n$ we have

$$\mathbb{P}[||M|| > t] = \mathbb{P}[||M||^k > t^k] \leq \frac{\mathbb{E} \sum_i (M^k)_{ii}}{t^k} \leq \left(\frac{2e\sqrt{nk}}{t} \right)^k$$

And so with high probability $||M|| \leq 4e\sqrt{n \ln n}$.

A more careful counting shows that the number of sequences is actually at most $2^{O(k)} n^{k/2}$, leading to the conclusion that $||M|| \leq O(\sqrt{n})$ with high probability. We can understand the set of sequences that we want to count in the following way: a sequence $1, a_1, \ldots, a_{k-1}, 1$ defines a graph over the vertices $\{1, a_1, \ldots, a_{k-1}, 1\}$ with edges $\{\{1,a_1\},\ldots,\{a_{k-1},1\}\}$. Because of repetitions, we have at most $k/2$ edges, and hence at most $1 + k/2$ vertices. In our previous counting argument, we upper bounded the number of sequences with ℓ distinct vertices (not counting the start vertex 1) as $\binom{k}{\ell} \cdot k^{k-\ell} \cdot n^\ell$, which increases with ℓ and maxes out at $\ell = k/2$, thus it is interesting to see, as a first step toward a better proof, if the count can be improved in the $\ell = k/2$. In this case, the graph is a tree, which is visited in a depth-first search order. This means that, at every step in the sequence, we either see a new vertex or we *backtrack* in our depth-first search, and we only need to spend one bit to represent the latter even, leading to a much better $2^k \cdot n^{k/2}$ bound. Obtaining a $2^{O(k)} n^{k/2}$ bound for all sequences is much harder and we will not attempt to do so.

Let us briefly review how to generalize the analysis to $G_{n,p}$ for $p < 1/2$. In this case we will look at

$$M := A - \mathbb{E}\, A = A - pJ + pI$$

and we see that M is symmetric, it has zeroes on the diagonals, and the off-diagonal entries satisfy

$$\mathbb{E}\, M_{i,j} = 0$$

$$\mathbb{E}\, M_{i,j}^2 = p - p^2 \leq p$$

$$EM_{i,j}^k \leq p \quad \forall k > 2$$

As before

$$\mathbb{E}\, trace(M^k) = n\, \mathbb{E}(M^k)_{1,1}$$

and

$$\mathbb{E}(M^k)_{1,1} = \sum_{a_1,\ldots,a_{k-1} \in V} \mathbb{E}\, M_{1,a_1} \cdot M_{a_1,a_2} \cdots M_{a_{k-2},a_{k-1}} \cdot M_{a_{k-1},1}$$

Now the expectation on the right-hand side is zero if there is a pair that occurs only once in the sequence. Otherwise, it is p^ℓ where ℓ is the number of distinct pairs in the sequence. The contribution to the weighted sum of sequences with $k/2$ distinct "edges" is thus $2^{O(k)} p^{k/2} n^{k/2}$. When $p > \frac{\log n}{n}$, this is the dominant term, and the whole weighted sum is at most $2^{O(k)} p^{k/2} n^{k/2}$, leading to a high probability bound $||M|| \leq O(\sqrt{pn})$.

Unfortunately, for $p = o\left(\frac{\log n}{n}\right)$ we do not have the high probability bound $||M|| \leq O(\sqrt{pn})$. To see why not, first observe that, for every $p > 1/n$ there is a high probability that the *largest* degree in the graph is $\Omega\left(\frac{\log n}{\log \log n}\right)$. Furthermore, if a graph has a vertex of large degree, then $||A - pJ||$ will have large spectral norm.

Lemma 1 *Let G be a graph, A its adjacency matrix, and suppose that G has a vertex v of degree $D \geq 1$. Then, for every p, we have*

$$||A - pJ|| \geq \sqrt{D} - 2pD$$

PROOF: Let $u_1, \ldots, u_D$ be the D neighbors of v. Consider the vector $\mathbf{x}$ defined as $\mathbf{x}_v = 1$, $\mathbf{x}_{u_i} = 1/\sqrt{D}$ for $i = 1, \ldots, D$, and $\mathbf{x}_z = 0$ for all other vertices z.

Then

$$||\mathbf{x}||^2 = 2$$
$$\mathbf{x}^T A\mathbf{x} \geq 2\sqrt{D}$$
$$\mathbf{x}^T pJ\mathbf{x} = p \cdot (\sum_z \mathbf{x}_z)^2 \leq p \cdot (1+\sqrt{D})^2 \leq 4pD$$

and

$$||A - pJ|| \geq \frac{\mathbf{x}^T(A-pJ)\mathbf{x}}{||\mathbf{x}||^2} \geq \sqrt{D} - 2pD$$

□

Thus, for $p < \frac{\log n}{n}$, we have with high probability $||A - pJ|| \geq \Omega\left(\sqrt{\frac{\log n}{\log\log n}}\right)$