# Lecture 6

*In which we study the spectrum of random graphs.*

## 1 Overview

When attempting to find in polynomial time an upper bound certificate on the max cut and maximum independent set of a graph, we have used the following property.

**Proposition 1** *If $G \sim G_{n,\frac{1}{2}}$, then with high probability $\|A - \mathbb{E}(A)\| \leq O(\sqrt{n})$, where $\|\cdot\|$ is the spectral norm. Generally, if $G \sim G_{n,p}$ and $p > \frac{\log n}{n}$ then w.h.p.*

$$\|A - \mathbb{E}(A)\| \leq O(\sqrt{np}).$$

Today we will prove how to obtain the bound in Proposition 1 with an extra term of $\log n$, as well as show an outline of the method of finding the bound in Proposition 1. We will also show how when $p$ is small this bound breaks down, namely how when $p = \Theta(\frac{1}{n})$,

$$\|A - \mathbb{E}(A)\| \geq \Omega\left(\sqrt{\frac{\log n}{\log \log n}}\right).$$

## 2 Introducing the Trace

Henceforth $M_{ij}^k$ signifies $(M^k)_{ij}$. Take $M$ symmetric and real. All eigenvalues of this matrix are real, and we can enumerate them $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$.

**Definition 1** *The trace $\mathrm{Tr}(A)$ is defined to be $\mathrm{Tr}(A) = \sum_{i=1}^n A_{ii}$ where $A$ is an $n \times n$ matrix.*

Moreover we know that

**Theorem 2** $\mathrm{Tr}(A) = \sum_{1=1}^n \lambda_i.$

If we take $k$ large and even, the eigenvalues of $M^k$ are $|\lambda_1|^k \geq |\lambda_2|^k \geq \ldots \geq |\lambda_n|^k$. Therefore we have

$$\sum_i M_{ii}^k = \text{Tr}(M^k) = \sum_i |\lambda_i|^k \geq |\lambda_1|^k = \|M\|^k.$$

Moreover we have

$$\text{Tr}(M^k)^{\frac{1}{k}} = \left( \sum_i |\lambda_i|^k \right)^{\frac{1}{k}} \leq n^{\frac{1}{k}} |\lambda_1| = n^{\frac{1}{k}} \|M\|.$$

This gives us an estimation of the norm, $\|M\| \leq (\sum_i M_{ii}^k)^{\frac{1}{k}} \leq n^{\frac{1}{k}} \|M\|$, which for $k > \log n$ gives a constant factor approximation of $\|M\|$.

# 3  Using the Trace to Bound the Spectral Norm

Assume that $G \sim G_{n,\frac{1}{2}}$ and $A$ is the adjacency matrix of $G$. We will prove the following.

**Theorem 3** $\underset{G \sim G_{n,\frac{1}{2}}}{\mathbb{E}} (\text{Tr}((A - \mathbb{E}(A))^k))$ *is bounded above by* $2^{O(k)} n^{1+k/2} k^{k/2}$. *If* $k > \log n$, *by taking the kth root we achieve a bound of* $O(\sqrt{n \log n})$ *on* $\|A - \mathbb{E}(A)\|$.

## 3.1  Expected Value of Matrix Entries

First, we examine the matrix $M = 2A - 2\mathbb{E}(A)$. We have $M_{ii} = 0$ and $M_{ij} \in \{\pm 1\}$ with equal probability of each when $i \neq j$. Moreover $M_{ij} = M_{ji}$. If $i \neq j, \mathbb{E}(M_{ij}^k) = 0$ if $k$ is odd and $\mathbb{E}(M_{ij}^k) = 1$ for $k$ even.

$\mathbb{E}(\sum_i M_{ii}^k) = n \mathbb{E}(M_{11}^k)$ by the linearity of expectation and symmetry between the entries. We evalute $M_{11}^k$.

$$M_{11}^k = \sum_{\{i_1, \ldots, i_{k-1}\}} M_{1i_1} M_{i_1 i_2} \cdots M_{i_{k-1},1}$$

where $i_1, \ldots i_{k-1}$ represents the intermediate steps on a "path" between vertices that starts at 1 and returns to 1. For example, $M_{11}^2 = \sum M_{1i} M_{i1}$. Note that we can repeat edges in these paths. By the linearity of expectation

$$\mathbb{E}(M_{11}^k) = \sum_{\{i_1, \ldots, i_{k-1}\}} \mathbb{E}(M_{1i_1} M_{i_1 i_2} \cdots M_{i_{k-1},1}).$$

If any pair $\{i, j\}$ occurs $\ell$ times in the sequence of pairs $\{1, i_1\}, \{i_1, i_2\}, \ldots, \{i_{k-1}, 1\}$, where $\ell$ is odd, then as the value of this term is independent from all other terms and $\mathbb{E} M_{ij}^\ell = 0$ for odd $\ell$, then $\mathbb{E}(M_{1i_1} M_{i_1 i_2} \cdots M_{i_{k-1} 1}) = 0$. If all pairs occur an even number of times, their product's expectation is 1. Therefore $\mathbb{E}(M_{11}^k)$ is the number of sequences $i_1, \ldots, i_{k-1} \in V^{k-1}$ such that, in the sequence of pairs $\{1, i_1\}, \{i_1, i_2\}, \ldots, \{i_{k-1}, 1\}$, each pair occurs an even number of times.

## 3.2 Encoding argument

In order to give an upper bound on the number of such sequences, we will show how to encode a sequence where there are $m$ distinct edges. In the sequence $i_1, \ldots i_{k-1}$, the element $i_j$ is represented either as $(0, i_j)$, which takes $1 + \log n$ bits, if $i_j i$ appears for the first time in the sequence at location $j$, and as $(0, \ell)$ otherwise, where $\ell < j$ is such that $i_\ell = i_j$, which requires $1 + \log k$ bits. Notice that, if $i_j$ occurs for the first time at location $j$, then the pair $\{i_{j-1}, i_j\}$ also occurs for the first time at the location $j - 1$ and $j$. Thus the number of times that we encounter a vertex for the first time is at most the number of distinct edges. If we have $t$ distinct vertices (other than vertex 1), then we are using $k + t \log n + (k - t) \log k$; for $k < n$, this value increases with $t$, but we have $t \leq m \leq k/2$ (because every edge has to appear an even number of times and so there can be at most $k/2$ distinct edges. This means that we use at most $k + \frac{k}{2} \log n + \frac{k}{2} \log k$ bits in the encoding.

The number of strings that can be encoded using at most $L$ bits is $2^{L+1}$. If we assume $k < n$, we have the bound $\mathbb{E}(M_{11}^k) \leq k^{\frac{k}{2}} n^{\frac{k}{2}} 2^{k+1}$, meaning

$$\mathrm{Tr}(M) = n \, \mathbb{E}(M_{11}^k) \leq n^{1 + \frac{k}{2}} 2^{k+1} k^{\frac{k}{2}}.$$

Therefore using suitable $k$ and $t$ we achieve our bound on $\|M\|$. For example, choose $k = \log n$ and $t = 10\sqrt{n}\sqrt{\log n}$. We use Markov's inequality to obtain

$$\mathbf{P}(\|M\| > t) = \mathbf{P}(\|M\|^k > t^k) \leq \frac{\mathbb{E}\|M\|^k}{t^k} \leq \left( \frac{2n^{\frac{1}{k}}\sqrt{n}\sqrt{k}}{t} \right)^k \leq e^{-\Omega(\log n)} \to 0.$$

# 4 Tightening the Bound

To obtain the sharper bound of $O(\sqrt{n})$, we need to count the number of pairs more sharply and remove the $k^{\frac{k}{2}}$ term, namely improve the way we talk about repetitions. Here we give an outline for how to find a tighter bound.

The worst case in the above analysis is when the number of distinct vertices (not counting vertex 1) is maximal, which is $k/2$. In that case, the number of distinct "edges" $\{i_j, i_{j+1}\}$ is $k/2$, and they must form a connected graph over $1 + k/2$ vertices, that is, they have to form a tree. Furthermore, each edges is repeated exactly twice in the closed walk, otherwise we would not have enough distinct edges to connect $1 + k/2$ distinct vertices.

If the pairs form a tree, then the only way we can have closed walk in which every edge is repeated twice is that the closed walk is a depth-first visit of the tree. In this case, we can improve our encoding in the following way. In a depth-first visit of a tree only two events are possible at each step: either we discover a new vertex, or we backtrack on the edge between the current node and the parent node. Thus we only need to pay $1 + \log n$ bits to encode a new node in the sequence and 1 bit to encode an already seen node, and we obtain a bound of $2^{\frac{k}{2} + \frac{k}{2} \log n + \frac{k}{2}} = 2^k n^{\frac{k}{2}}$. By taking the $k$th root we obtain a bound on $\|M\|$ of $O(\sqrt{n})$.

# 5  Generalizing to any $p$

Now assume $G \sim G_{n,p}$ and $A$ is the adjacency matrix of $G$. We also assume $p < \frac{1}{2}$. We define
$$M = A - \mathbb{E}(A).$$
In this matrix $M_{ii} = 0$ and if $i \neq j$, $M_{i,j} = 1 - p$ with probability $p$ and $-p$ with probability $1 - p$. Therefore $\mathbb{E}(M_{ij}) = 0, \mathbb{E}(M_{ij}^2) = p - p^2 \leq p$. In fact, $\mathbb{E}(M_{ij}^k) \leq p$ for all $k \geq 1$.

From this we see we need to sum over sequences such that the multiset has each pair occuring at least two times, as if any pair occurs once, the expectation is 0.

Therefore the bound is
$$\mathbb{E}(M_{11}^k) \leq \sum_{i_1, \dots i_{k-1}} p^{\ell}$$

where $\ell$ is the number of distinct pairs and the sum is taken over multisets where each pair occurs at least twice. For large $\ell$, the number of sequences where each pair occurs at least twice with $\ell$ distinct pairs is approximately $2^{O(\ell)} n^{\ell}$. This would give us

$$\sum_{i_1, \dots i_{k-1}} p^{\ell} = \sum_{\ell} p^{\ell} 2^{O(\ell)} n^{\ell} \leq O(p^{\frac{k}{2}} 2^{O(k)} n^{\frac{k}{2}})$$

so the bound on $\|M\|$ is $O(\sqrt{np})$. However, the bound on the number of sequences with $\ell$ distict pairs breaks down when $\ell$ is much smaller than $k$. In a full proof much more complicated calculations must be done.

# 6  Problems with sparse graphs

**Theorem 4** *If $p = \Theta(\frac{1}{n})$ , then $\|A - \mathbb{E}(A)\| \geq \Omega\left(\sqrt{\frac{\log n}{\log \log n}}\right)$ w.h.p.*

This breaks down the nice bound we obtained in section 5. This follows from the irregularity of sparse graphs. There will be isolated vertices and vertices with degree much higher than average.

**Lemma 1** *If $p = \Theta(\frac{1}{n})$ then w.h.p. the highest degree vertex of $G$ is of order $\Theta\left(\frac{\log n}{\log \log n}\right)$.*

**Proposition 5** *If $G$ has a node of degree $\geq d$, then, for every $p < \frac{1}{4\sqrt{d}}$, $\lambda_{\max}(A - pJ) \geq \Omega(\sqrt{d})$. This implies that $\forall 0 < p < .\frac{1}{4\sqrt{d}}$, $\|A - pJ\| \geq \Omega(\sqrt{d})$.*

PROOF: We have
$$\lambda_{\max}(A - pJ) = \max_{\mathbf{x}} \frac{\mathbf{x}^T(A - pJ)\mathbf{x}}{\|\mathbf{x}\|^2}$$

where the maximum is taken over all nonzero vectors $\mathbf{x}$. Call $v$ a node of degree $\geq d$ and call $d$ of its neighbors $u_1, \dots, u_d$.

Consider the vector $\mathbf{x}$ such that $x_v = 1, x_{u_i} = \frac{1}{\sqrt{d}}$ and $x_w = 0$ for other vertices $w$. We have

$$\mathbf{x}^T A \mathbf{x} \geq 2\sqrt{d}$$

$$\mathbf{x}^T p J \mathbf{x} = p \cdot \left( \sum x_i \right)^2 = p \cdot (1 + \sqrt{d})^2 \leq 4pd$$

$$||\mathbf{x}||^2 = 2$$

Therefore if $p \leq \frac{1}{4\sqrt{d}}$,

$$\frac{\mathbf{x}^T (A - pJ) \mathbf{x}}{||\mathbf{x}||^2} \geq \sqrt{d} - \frac{1}{2}\sqrt{d} = \Omega(\sqrt{d})$$

yielding the desired bound.

□

Theorem 4 proceeds immediately from Proposition 5 and Lemma 1.