MIT 6.854/18.415: Advanced Algorithms Spring 2016

Lecture 21 – April 25, 2016

Prof. Ankur Moitra *Scribe: Andrew Spielberg, Tom Yan, Kevin Li, Matthew Brennan*

1 Introduction

Last Time: Grothendieck's inequality and the Lovasz Theta function. The Lovasz Theta function is constructed with vectors from Taylor expansion of $\sin(x)$. It satisfies $\alpha \leq \theta \leq \overline{\chi}$ where χ is the chromatic number and α is the independence number of the graph.

In this lecture we'll explore Random Matrix Theory by exploring an application in network analysis. In particular, we'll examine how Random Matrix Theory can be employed to develop an efficient randomized algorithm for finding planted cliques. In doing so, we will explore a general method for analyzing random graph problems with "planted" structure by treating them as a signal reconstruction problem in the presence of random noise.

2 Erdős-Renyi Model

The Erdős-Renyi model $G(n, p)$ is a model for generating random graphs. It defines a distribution over all simple graphs on n vertices.

Definition 1 (Erdős-Renyi Model)**.** *The Erdős-Renyi model $G(n, p)$, defines a random graph of n nodes $v_1, \cdots, v_n$. For each $i, j \in [n]$ where $i \neq j$, $v_i v_j$ is included in G independently at random with probability p.*

This model is undirected and does not consider self-edges. Note that $G(n, p)$ defines a *uniform* distribution over all possible $2^{\binom{n}{2}}$ simple graphs on n vertices.

As an example, consider $G(n, \frac{1}{2})$. Let $\omega(G)$ denote the largest clique in a graph G. How large is $\omega(G(n, \frac{1}{2}))$? We first derive an initial upper bound on $\omega(G(n, p))$.

$$\begin{aligned}
\mathbb{E}\left[\#k\text{-cliques in } G(n, 1/2)\right] &= \binom{n}{k} 2^{-\binom{k}{2}} \\
&\leq n^k 2^{-\binom{k}{2}} \\
&= 2^{k\left(\log_2 n - \frac{k-1}{2}\right)}
\end{aligned}$$

Note that that because the graphs are uniformly distributed, each unique graph occurs with probability $2^{-\binom{k}{2}}$; meanwhile there are $\binom{n}{k}$ cliques graphs with a k-clique, hence the first line. Therefore if $k = (2 + \delta) \log_2 n$ then the expected number of k-cliques behaves like

$$\mathbb{E}\left[\#k\text{-cliques in } G(n, 1/2)\right] \lesssim 2^{-2\delta \log n}$$

A stronger fact that we will not prove here, which implies that this analysis is generally tight, is the following.

Theorem 2. *With high probability, the size of the largest clique of $G(n, 1/2)$ is*

$$\omega(G) = (2 \pm o(1)) \log_2 n$$

3 Planted Clique Model (PC)

The planted clique model is another model for generating random graphs that defines a distribution over all simple graphs on n vertices.

Definition 3 (Planted Clique Model)**.** *A graph G in the PC model $G(n, p, k)$ is generated in two steps:*

1. *Choose a random graph H from the distribution $G(n, p)$; and*
2. *Choose k vertices from H uniformly at random and plant a k-clique on them (that is, fully connect them) to obtain G.*

Once again, we will take a particular look at the model with $p = \frac{1}{2}$ *e.g.* $G(n, \frac{1}{2}, k)$.

A natural question is: can we recover the k vertices in the planted clique in a given instance G of PC? Intuitively, this should only be possible when the cliques which appear in H are all smaller than k; if there were multiple cliques of size k in G, there would be natural ambiguity about which clique is the planted one! In fact, it is information theoretically impossible to recover the planted clique when it has size at most $k \leq 2 \log_2 n$; high probability $G(n, 1/2)$ already has cliques of this size. However we can find the k-clique in quasi-polynomial time if it is slightly larger.

Theorem 4. *There is an $n^{O(\log n)}$-time algorithm for recovering the planted clique that succeeds with high probability if $k \geq (2 + \delta) \log_2 n$, for some small δ.*

Proof Idea. With high probability, the only $k-$clique in the graph will be the planted clique (note that this is the only variable which determines success or failure). Given that the planted clique is the unique $k-$clique, we simply need to find it. This can be accomplished by brute-force searching for a $(2 + \delta) \log_2 n$ sized clique by searching over all sets of vertices of size $(2 + \delta) \log_2 n$. Then find all common neighbors of the vertices which form a $(2 + \delta)$-clique and output these vertices. □

Unfortunately, this algorithm is quasi-polynomial in time. What can we say if we wish to restrict ourselves to only polynomial time algorithms?

Theorem 5. *There is a polynomial time algorithm for finding a planted $k-$clique that succeeds with high probability if $k \geq C\sqrt{n \log_2 n}$, $C \geq 1$.*

Proof. A vertex u has degree $\deg(u)$ which has one of two possible distributions.

$$\deg(u) = \begin{cases} \text{Bin}(n-1, 1/2) & \text{if } u \text{ is not in the planted clique} \\ k - 1 + \text{Bin}(n-k, 1/2) & \text{if } u \text{ is in the planted clique} \end{cases}$$

With high probability (by the Chernoff and by the union bound over all nodes), the following holds:

1. For all u not in the planted clique,

$$\deg(u) \leq \frac{n}{2} + \frac{C}{4}\sqrt{n \log_2 n}$$

2. For all u in the planted clique,

$$\deg(u) \geq \frac{n-k}{2} + k - \frac{C}{4}\sqrt{n \log_2 n}$$

Thus, a successful algorithm can now find the nodes in the planted clique by selecting all nodes with degree $\geq \frac{n-k}{2} + k - \frac{C}{4}\sqrt{n \log_2 n}$. □

This algorithm is efficient, but not tight. By using tools from random matrix theory (RMT), we can remove the factor of $\sqrt{\log_2 n}$. The main goal of the remainder of the lecture will be to prove the following theorem.

Theorem 6 (AKS)**.** *There is a polynomial time algorithm for finding the k-clique of graph generated from PC with high probability if $k \geq C\sqrt{n}$.*

4 Random Matrix Theory

In this section, let $A \in \mathbb{R}^{n \times n}$ be a symmetric random matrix given by

$$A_{ij} = \begin{cases} \text{random } \pm 1 & \text{if } i \leq j \\ A_{ji} & \text{otherwise} \end{cases}$$

Each entry A_{ij} where $i \leq j$ is equal to each of ± 1 with probability $\frac{1}{2}$.

We will now turn our attention to operator norms. A key component of the proof will be to compare the operator norm of planted cliques vs. random graphs. We will design an algorithm which takes advantage of the fact that the operator norm of the planted clique is much larger than the operator norm of the "noise" generated by the random graph.

The operator norm $\|A\|_{op}$ of a symmetric random matrix can be bounded as follows.

Theorem 7. *A random symmetric matrix A satisfies with high probability that*

$$\|A\|_{op} \leq (2 + o(1))\sqrt{n}.$$

While this is true, we instead prove a weaker theorem relaxing the constant $2 + o(1)$ to $C > 0$ with elementary tools. To prove this result, we first require the notion of a maximal q-net and a bound on the size of any maximal $\frac{1}{4}$-net.

We define a Maximal q-Net as follows.

Definition 8 (Maximal q-Net)**.** *A set of points $\Sigma \subseteq S^{n-1}$ where S^{n-1} is the unit sphere in n dimensions is a maximal q-net if it satisfies the properties:*

1. *For all $x, y \in \Sigma$ where $x \neq y$, $\|x - y\|_2 \geq q$.*

2. For all $z \in S^{n-1}$, there is some $x \in \Sigma$ with $\|x - z\|_2 < q$.

The maximal net is a set of points on the sphere. Intuitively, the first condition ensures that each point in the net is sufficeintly far away from each other (in other words, the points must be relatively sparse). The second condition ensure that the net sufficiently covers the area of the sphere (which in turn lower bounds the total number of points).

Now we bound the size of a maximal $\frac{1}{4}$-net in n dimensions from above.

Theorem 9. *An maximal $\frac{1}{4}$-net $\Sigma \subseteq S^{n-1}$ in n dimensions has size at most $|\Sigma| \leq 11^n$.*

Proof. Note that the balls of radius 1/10 centered each point in the maximal $\frac{1}{4}$-net in $\{B(x, 1/10)\}_{x\in\Sigma}$ are disjoint (from 1. in the definition). Moreover, the union of all of these balls is contained in $B(0, 11/10)$. This implies that the size of Σ can be bounded by

$$|\Sigma| \leq \frac{\text{Vol}(B(0, 11/10))}{\text{Vol}(B(0, 1/10))} = 11^n$$

□

This bound and the fact that maximal $\frac{1}{4}$-nets exist in n dimensions allows us to now prove the weaker theorem.

Theorem 10. *There is a constant $C > 0$, such that with high probability*

$$\|A\|_{op} \leq C\sqrt{n}.$$

Proof. For any fixed unit vector x with $\|x\| = 1$, the value $x^t Ax$ is a random variable with variance

$$\text{Var}[x^t Ax] = \sum_{i=1}^{n} x_i^4 + 2\sum_{i<j}(x_i x_j)^2 = \left(\sum_{i=1}^{n} x_i^2\right)^2 = \|x\|^4 = 1$$

Each term comes from the definition $Var(x) = \mathbb{E}[x^2] - \mathbb{E}[x]^2$ noting that the mean $\mathbb{E}[x] = 0$. The factor of two comes from symmetry.

Therefore it follows that $x^t Ax$ has variance 1 and satisfies, via Chernoff,

$$|x^t Ax| \leq c_1\sqrt{n}$$

with high probability $1 - e^{-10n}$ for some constant $c_1 > 0$ (the 10 was just chosen as some sufficiently large number). □

Theorem 11. *For some $d > 0$, $\mathbb{P}\left[\|A\|_{op} > C\sqrt{n}\right] < e^{-dn}$*

Proof. Now let z be any unit vector which maximizes $|z^tAz|$ and therefore satisfies $|z^tAz| = \|A\|_{op}$. By the second property of maximal $\frac{1}{4}$-nets, there is some x such that $\|x-z\|_2 < \frac{1}{4}$. Now by the triangle inequality,

$$\begin{aligned}|x^tAx| &\geq |z^tAz| - |(z-x)^tAx| - |x^tA(z-x)| - |(z-x)^tA(z-x)| \\ &\geq \|A\|_{op} - \frac{\|A\|_{op}}{4} - \frac{\|A\|_{op}}{4} - \frac{\|A\|_{op}}{16} \\ &\geq \frac{\|A\|_{op}}{4}\end{aligned}$$

Note that this inequality follows from Cauchy-Schwarz and the definition of the operator norm applied as follows

$$|(z-x)^tAx| \leq \|(z-x)^tA\|_2 \cdot \|x\|_2 \leq \|A\|_{op}\|x-z\|_2 < \frac{\|A\|_{op}}{4}$$

$$|x^tA(x-z)| \leq \|x\|_2 \cdot \|A(x-z)\|_2 \leq \|A\|_{op}\|x-z\|_2 < \frac{\|A\|_{op}}{4}$$

$$|(x-z)^tA(x-z)| \leq \|x-z\|_2 \cdot \|A(x-z)\|_2 \leq \|A\|_{op}\|x-z\|_2^2 < \frac{\|A\|_{op}}{16}$$

The inequality $|x^tAx| \geq \frac{\|A\|_{op}}{4}$ implies we just need to optimize over the maximal $\frac{1}{4}$-net. Therefore

$$\mathbb{P}\left[\|A\|_{op} > C\sqrt{n}\right] \leq \mathbb{P}\left[\exists x \in \Sigma \,:\, |x^tAx| > \frac{C}{4}\sqrt{n}\right] \leq |\Sigma|e^{-10n} < e^{-dn}$$

using $|\Sigma| \leq 11^n$ and the union bound, choosing $C = 4c_1$. Here d is some positive constant. □

As desired, the operator norm of the matrix corresponding to a random Erdős-Renyi graph is small. We can now exploit the fact that the operator norm of a matrix corresponding to a planted clique is large in order to recover the planted clique with high probability.

5 An Algorithm For Planted Clique

We now introduce a "spectral algorithm" for finding the planted clique.

Given a G, our graph with a planted clique, construct a corresponding matrix A as follows:

$$A_{ij} = \begin{cases} +1 & \text{if } i = j \\ +1 & \text{if } i \neq j, (i,j) \in G \\ -1 & \text{otherwise} \end{cases}$$

Let u' be the top eigenvector of A (the eigenvector with maximal corresponding eigenvalue). Let T be the top k coordinates of u'.

$$C \triangleq \{u | u \text{ has at least } \frac{4}{5}k \text{ neighbors in T}\}$$

Theorem 12. *If $k \geq C\sqrt{n}$ for sufficiently large C, T contains at least $\frac{4}{5}k$ vertices of the planted clique with high probability. Further, with high probability, every u not in the planted clique is adjacent to at most $\frac{4}{7}k$ of the vertices in the planted clique.*

With this theorem, since $\frac{4}{7}k + \frac{1}{5}k < \frac{4}{5}k$, with high probability C will not contain any vertices not in the planted clique. Vertices in the clique clearly have at least $\frac{4}{5}k$ neighbors in T, so with high probability, C is the planted clique.

Proof. We will simply sketch the proof here. We will decompose our A matrix into a sum of two matrices, *i.e.* $A = M + E$.

Define M and E as:

$$M = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & 0 \end{array}\right]$$

$$E = \left[\begin{array}{c|c} 0 & \pm 1 \\ \hline \pm 1 & \pm 1 \end{array}\right]$$

Where the partition into submatrices is into segments of dimension k and $n-k$ (in both row and column, *e.g.* the top-left quadrant is $k \times k$, the top-right quadrant is $k \times n-k$, etc.). Note that the $k \times k$ submatrix in M corresponds to the planted-clique, E corresponds to the remainder of the random graph.

It follows then from theorem 10 (and cheating a bit due to asymmetry in submatrices, but it doesn't affect the proof) that $|||E|||_{op} \leq 3C\sqrt{n}$. In other words, the "noise" from matrix E behaves nicely, *i.e.* the norm is not too large.

It is trivial to compute that $|||M|||_{op} = k$, and its top eigenvector is:

$$u = [\frac{1}{\sqrt{k}}, \frac{1}{\sqrt{k}}, \ldots, \frac{1}{\sqrt{k}}, 0, 0, \ldots 0]$$

i.e. the first k elements are $\frac{1}{\sqrt{k}}$ and the last $n-k$ elements are 0.

The intuition is that if $k \geq 800C\sqrt{n}$, then $|||A|||_{op} \approx |||M|||_{op}$ (by the triangle inequality). Further, the top eigenvectors are close, since $sin\theta(u, u') \leq \frac{2|||E|||_{op}}{k}$, *i.e.* the angular distance between u and u' is bounded roughly by the small norm of the error.

Roughly, this implies that $< u, u' >= 1 - \frac{1}{100}$. When this holds, the following statements are true:

1. At least $\frac{4k}{5}$ of the first k coordinates in u' have value at least $\frac{1}{2\sqrt{k}}$.
2. At most $\frac{k}{5}$ of the last $n-k$ coordinates of u' have value at least $\frac{1}{2\sqrt{k}}$.

The first fact is all we need here, since it implies that among the k largest coordinates in u', at least $\frac{4}{5}$ of them are from the planted clique.

□

So to recap, the steps of the proof were:

- The graph's corresponding matrix can be decomposed into into a matrix M corresponding to the planted clique, and a matrix E corresponding to the rest of the Erdős-Renyi matrix. $A = M + E$.
- Since the operator norm of the E matrix is small (from RMT), the top eigenvectors of A and M are close, with high probability.
- When they *are* close, the top k coordinates of the top eigenvector of M correspond to the planted clique and are $\frac{1}{\sqrt{k}}$. Thus, we expect the top k coordinates of the top eigenvector of A will correspond to the planted clique and have value roughly $\frac{1}{k}$. In fact, at least $\frac{4}{5} * k$ of them which *do* correspond to the planted clique be above a certain threshold, and at most $\frac{1}{5} * k$ of them *don't* correspond to the planted clique and be above a certain threshold.
- Thus, thresholding on the top eigenvector of A, we can collect most of the vertices of our clique with few false-positives.

From this starting vertex set, we can then efficiently collect the remaining vertices to recover the planted-clique.

One final question - why did we only have to look at the top eigenvalue? Couldn't we analyze other eigenvalues for further information? The matrix M corresponding to the true planted clique has rank 1. Because of that, it has exactly one nonzero eigenvalue, and its corresponding eigenvector provides information about the planted model. As we will soon see, this is not the case for all planted problems.

6 Planted Bisection

As a final note, in order to demonstrate that RMT can be applied in this way to develop and analyze algorithms for a number of graph problems, we consider another (similar) problem, planted bisection (PB).

Definition 13 (Planted Bisection)**.** *A graph G in the PB model $G(n, p, q)$ is generated in three steps:*

1. *Partition the node set into two equal sets of size $\frac{n}{2}$.*
2. *If two nodes are within the same node set, connect them with probability p.*
3. *Otherwise connect them with probability q.*

Again, this model is undirected and does not consider self-edges.

Theorem 14. *There exists a polynomial time algorithm to recover the partition when $\frac{p-q}{p} \geq C\frac{\log(n)}{np}$.*

We will not prove this theorem here, but merely sketch the intuition for why this is possible.

Let A be the adjacency matrix of G. A can be again decomposed into two matrices M and E. In this case, $E = A - M$, where:

$$M = \left[\begin{array}{c|c} p & q \\ \hline q & p \end{array}\right]$$

Where the p block corresponds to edges within a single node set, and the q block corresponds to nodes in separate node sets. Note that M is now a rank 2 matrix instead of rank 1.

Analysis of this problem would follow as such:

- Bound the operator norm of E.
- use this to say the top two eigenvectors of A are close to the top two eigenvectors of M with high probability.
- Establish what the top two eigenvectors of M would look like.
- Show that when the eigenvectors are close, there is an efficient scheme to exactly determine the node partition using values of those eigenvectors.