

Lecture 6: Spectral methods for planted clique

In which we study a spectral algorithm that finds a planted clique in a random graph.

1 Finding a Planted Clique

We will analyze the following algorithm:

- Input: graph G , desired clique size k
- Let A be the adjacency matrix of G and define $M := A - \frac{1}{2}J$
- Let x be an eigenvector of the largest eigenvalue of M
- Let I be the set of k vertices v that have the largest values of $|x_v|$
- Let C be the set of vertices $v \in V$ that have at least $\frac{3}{4}k$ neighbors in I
- Return C

We will show that if $k \gg \sqrt{n}$, for example $k > 100\sqrt{n}$, then with high probability the above algorithm finds the planted clique. Here, “high probability” means one minus an error term that goes to zero faster than the inverse of any polynomial.

2 The eigenvector of the largest eigenvalue of $A - \frac{1}{2}J$

An important theme of this lecture is that when we construct a random object using independent choices, then various quantities related to the object concentrate around the average.

Before we see what we mean by this, let us study the expectation of A , because, whenever we have a random object, it is always helpful to start by understanding its average.

Let us fix the set K which is the planted clique of size k in our distribution. Then the matrix $\mathbb{E}A$ has zeroes on the diagonal, and the off-diagonal entries (u, v) are 1 if u and v are both in K and $\frac{1}{2}$ otherwise. So we can write

$$\mathbb{E}A = \frac{1}{2}J + \frac{1}{2}\mathbf{1}_K\mathbf{1}_K^T - D$$

where J is the matrix that is one everywhere, $\mathbf{1}_K$ is the *indicator vector* of K , whose v -th coordinate is 1 if $v \in K$ and is 0 otherwise, and D is a diagonal matrix such that $D_{v,v}$ is equal to 1 if $v \in K$ and equal to $\frac{1}{2}$ otherwise.

Let us ignore D for a moment. If we had the matrix $\frac{1}{2}J + \frac{1}{2}\mathbf{1}_K\mathbf{1}_K^T$, we could subtract $\frac{1}{2}J$ from it and we would be left with the rank-one matrix $\frac{1}{2}\mathbf{1}_K\mathbf{1}_K^T$, whose only non-trivial eigenvector is the indicator of K . Of course we do not have such a matrix, but we have the matrix A whose average (up to the diagonal correction) is $\frac{1}{2}J + \frac{1}{2}\mathbf{1}_K\mathbf{1}_K^T$. So what happens if we pretend that A and $\mathbb{E}A$ are the same, and we compute the eigenvector of the largest eigenvalue of $A - \frac{1}{2}J$? It turns out that we find a vector that is close to $\mathbf{1}_K$! We will now see why.

First, we need to establish that A and $\mathbb{E}A$ are close to each other in operator norm. We can think of the process of sampling the a planted clique graph G with clique K as first picking an Erdős-Renyi random graph G_{ER} , then defining G_K to be the graph whose edges are the edges within K that are missing from G_{ER} , and finally letting G be the graph whose edge set is the union of G_{ER} and G_K . If we let A , A_{ER} and A_K be, respectively the adjacency matrices of G , G_{ER} and G_K , we have $A = A_{ER} + A_K$. The important observation is that A_K is itself the adjacency matrix of an Erdős-Renyi random graph with edge probability $1/2$, and so we have that, with high probability,

$$\|A_{ER} - \mathbb{E}A_{ER}\| \leq (1 + o(1))\sqrt{n}$$

and

$$\|A_K - \mathbb{E}A_K\| \leq (1 + o(1))\sqrt{|K|}$$

which gives us the following bound.

Lemma 1 *Fix a set K of size $k = o(n)$. With high probability over the choice of a graph G from the k -planted clique distribution conditioned on K being the clique, if A is the adjacency matrix of G , we have*

$$\|A - \mathbb{E}A\| \leq (1 + o(1)) \cdot \sqrt{n}$$

where n is the number of vertices of G .

Note that the matrix D described above satisfies $\|D\| = 1$ and so whenever the conclusion of the above lemma holds we also have

$$\left\| A - \frac{1}{2}J - \frac{1}{2}\mathbf{1}_K\mathbf{1}_K^T \right\| \leq \|A - \mathbb{E}A\| + \|D\| \leq (1 + o(1))\sqrt{n}$$

which means that the matrix $A - \frac{1}{2}J$, which we know, and the matrix $\frac{1}{2}\mathbf{1}_K\mathbf{1}_K^T$, whose non-trivial eigenvector encodes the set we are looking for, are very close.

Now we would like to say that if two matrices are close in spectral norm, then their respective eigenvectors of their largest eigenvalues are also close. The Davis-Kahan theorem says exactly that. For our purposes, it is sufficient to work with a special case that is easier to state.

Theorem 2 (Davis and Kahan) *If M is a symmetric matrix, $\mathbf{y}\mathbf{y}^T$ is a rank-one symmetric matrix, and \mathbf{x} is an eigenvector of the largest eigenvalue of M , we have*

$$|\sin(\hat{\mathbf{x}}\mathbf{y})| \leq \frac{\|M - \mathbf{y}\mathbf{y}^T\|}{\|\mathbf{y}\mathbf{y}^T\| - \|M - \mathbf{y}\mathbf{y}^T\|}$$

where $\hat{\mathbf{x}}\mathbf{y}$ is the angle between the vectors \mathbf{x} and \mathbf{y} .

That is, if the distance between M and $\mathbf{y}\mathbf{y}^T$ is small compared to the “size” of $\mathbf{y}\mathbf{y}^T$, then the eigenvector of the largest eigenvalue of M is almost aligned with \mathbf{y} .

Corollary 3 *If $\|(A - J/2) - \mathbf{1}_K\mathbf{1}_K^T/2\| \leq (1 + o(1))\sqrt{n}$, $k > 100\sqrt{n}$, and \mathbf{x} is an eigenvector of the largest eigenvalue of $A - J/2$ scaled so that $\|\mathbf{x}\|^2 = k$, then*

$$\min\{\|\mathbf{x} - \mathbf{1}_K\|^2, \|\mathbf{x} + \mathbf{1}_K\|^2\} \leq 0.002k$$

for sufficiently large n .

PROOF: We apply the Davis-Kahan theorem and we see that

$$|\sin(\mathbf{x} \hat{\mathbf{1}}_K)| \leq \frac{(1 + o(1))\sqrt{n}}{\frac{k}{2} - (1 + o(1))\sqrt{n}} = \frac{1}{49} + o(1)$$

Now we have to reason about the distance between \mathbf{x} and $\mathbf{1}_K$. Recall that both vectors have length-squared equal to k .

$$\begin{aligned} \|\mathbf{x} - \mathbf{1}_K\|^2 &= \|\mathbf{x}\|^2 + \|\mathbf{1}_K\|^2 - 2\langle \mathbf{x}, \mathbf{1}_K \rangle = 2k - 2\langle \mathbf{x}, \mathbf{1}_K \rangle \\ \|\mathbf{x} + \mathbf{1}_K\|^2 &= 2k + 2\langle \mathbf{x}, \mathbf{1}_K \rangle \end{aligned}$$

so we have

$$\min\{\|\mathbf{x} - \mathbf{1}_K\|^2, \|\mathbf{x} + \mathbf{1}_K\|^2\} = 2k - 2 \cdot |\langle \mathbf{x}, \mathbf{1}_K \rangle|$$

We also have

$$\langle \mathbf{x}, \mathbf{1}_K \rangle = \|\mathbf{x}\| \cdot \|\mathbf{1}_K\| \cdot \cos(\mathbf{x} \hat{\mathbf{1}}_K) = k \cos(\mathbf{x} \hat{\mathbf{1}}_K)$$

So

$$|\langle \mathbf{x}, \mathbf{1}_K \rangle| = k |\cos(\mathbf{x} \hat{\mathbf{1}}_K)| = k \cdot \sqrt{1 - \sin^2(\mathbf{x} \hat{\mathbf{1}}_K)} \geq k \sqrt{\frac{49^2 - 1}{49^2} - o(1)} > 0.999 \cdot k$$

for sufficiently large n . \square

3 Cleaning up

The previous section analyzes what happens in the first two lines of the algorithm: with high probability we have a vector \mathbf{x} such that either \mathbf{x} or $-\mathbf{x}$ is close to $\mathbf{1}_K$. How do we use such a vector to find K ? The first observation, which analyzes the third line of the algorithm, is that, if we define the set I to be the k vertices v for which $|x_v|$ is largest, then the set I and the set K are almost the same.

Lemma 4 *If \mathbf{x} is a vector such that $\|\mathbf{x}\|^2 = |K| = k$, and*

$$\{\|\mathbf{x} - \mathbf{1}_K\|^2, \|\mathbf{x} + \mathbf{1}_K\|^2\} \leq \epsilon k$$

and if I is the set of k vertices v with largest $|x_v|$, then I and K have at least $k \cdot (1 - 4\epsilon)$ vertices in common.

PROOF: Let us call a vertex v *bad* if it is in K and we have $|x_v| \leq \frac{1}{2}$ or if it is not in K and we have $|x_v| \geq \frac{1}{2}$. Let B be the set of bad vertices.

The first observation is that $|B| \leq 4\epsilon$, because each vertex in B contributes at least $1/4$ both to $\|\mathbf{x} - \mathbf{1}_K\|^2$ and to $\|\mathbf{x} + \mathbf{1}_K\|^2$.

The second observation is that I and K must have at least $k - |B|$ vertices in common. To see why, consider two cases. Let t be the smallest value of $|x_v|$ among the vertices v in I . That is, if we sorted the vertices of G in decreasing order of $|x_v|$, then t would be the value that we would see in the k -th position of the sorted order. If $t > 1/2$, then all the vertices $v \notin K$ that are included in I have $|x_v| > 1/2$, and so are bad vertices, which means that I contains at most $|B|$ vertices not in K , and hence contains at least $k - |B|$ vertices from K . If $t \leq 1/2$, then every vertex of K that we do not include in I has $|x_v| \leq 1/2$, and so it is a bad vertex. This means that I contains at least $k - |B|$ of the k vertices of K . \square

We now understand what happens after the first three lines of our algorithm are executed: with high probability, the set I contains at least $1 - 4 \cdot 0.002 \cdot k = .992 \cdot k$ of the k elements of K .

In conclusion, we need to understand what happens in the fourth line. Note that, if the above happens, then C will include all the elements of K , since all the elements of K have at least $.992 \cdot k > \frac{3}{4}k$ neighbors in I . The last thing to prove is that, with high probability, C will not contain any vertex outside of K .

We will use the following result, which, again, relates the probable value of a random variable to the value of its average.

Theorem 5 (Chernoff bounds) *If X_1, \dots, X_k are independent random variables taking value either 0 or 1, and $X := \sum_{i=1}^k X_i$, then, for every $0 < \epsilon < 1$*

$$\mathbb{P}[X - \mathbb{E}X > \epsilon k] \leq e^{-2\epsilon^2 k}$$

$$\mathbb{P}[X - \mathbb{E}X < -\epsilon k] \leq e^{-2\epsilon^2 k}$$

Now, let us go back to our random graph: if $v \notin K$, then the number of neighbors of v in K is a random value of average $k/2$ that can be written as a sum of k independent 0/1 random variables each of average $1/2$. This means that each vertex $v \notin K$ has a very low probability of having much more than $k/2$ neighbors in K .

Corollary 6 *Fix a vertex $v \notin K$. With probability at least $1 - e^{-0.2k}$ over the choice of G , the vertex v has at most $.6k$ neighbors in K .*

PROOF: Apply the Chernoff bound with $\epsilon = 0.1$. \square

Now we apply the *union bound* to argue that the same happens for all vertices not in K .

Corollary 7 *With probability at least $1 - (n - k) \cdot e^{-0.2k}$ over the choice of G , every vertex $v \notin K$ has at most $.6k$ neighbors in K .*

PROOF: Consider the negation of the event that we are interested in, that is, the event that there is a vertex $v \notin K$ with more than $.6k$ neighbors in K . This is the OR of $n - k$ events, each of them having probability at most $e^{-0.2k}$, and so its probability is at most $(n - k) \cdot e^{-0.2k}$. \square

So, with $1 - e^{-\Omega(\sqrt{n})}$ probability, each vertex $v \notin K$ has at most $.6k$ neighbors in K . But the set I contains a subset of K plus at most $.008$ other vertices, so each vertex $v \notin K$ has at most $.608k$ neighbors in I , which is less than $3/4$. This means that C will (with high probability) not include any vertex $v \notin K$.