PRINCETON COS 521: ADVANCED ALGORITHM DESIGN

Average-Case Models: Planted Partition and Random Matrices

In this lecture, we study the problem of finding large communities in networks. For a natural example, consider a social network such as Facebook where one can naturally associated nodes with people and edge with friends/acquaintance relationships. Communities in such graphs may naturally correspond to sociological clusters – people with the same workplace, or geographical location, or studying at the same university and so on. One can thus imagine that given a graph representation of a social network finding such communities may be a useful primitive.

A natural abstraction of a community is a clique in a graph (a collection of vertices that are all connected). Can we find large cliques in graphs? Unfortunately, in the worst-case this problem turns out to be hard. Not just in the exact sense but also for approximation. More precisely, it turns out that for any $\epsilon > 0$, given a graph that is promised to contain a $n^{1-\epsilon}$ size clique, it is NP-hard to find a clique that contains more than n^{ϵ} vertices. Our best known algorithms in fact can only find a $\log^{O(1)}(n)$ size cliques even in graphs promised to contain a $\tilde{O}(n)$ size clique.

We may contest that clique is a rather stringent demand and for finding communities, we may be okay with finding "dense subgraphs". But it turns out that the strong hardness results continue to hold even for finding densest subgraphs (we will omit a formal definition here).

The strong hardness of approximation that we see for clique is not all that rare for expressive/interesting optimization problems. However, they do not preclude the search for good algorithms; they suggest that we model our input instances in further detail. And this effort to better model instances of worst-case hard optimization problems now forms a broad research area called *beyond the worst-case analysis of algorithms.*

In today's class, we will see an example of how to model input instances with community structure. This model might appear simplistic but leads to new spectral tricks for finding large communities that do find broad usage in applications.

Definition 1 (Stochastic Block Model)**.** *Let $G(V,E)$ be a random graph with vertices $V = 1,\ldots,n$. Let S,T form a bisection of V. I.e. $S,T \subset V$ with $S \cup T = V$, $S \cap T = \emptyset$ and $|S| = |T| = n/2$. For probabilities $p > q$, construct G by adding edge (i,j) independently with probability Y_{ij}, where:*

$$Y_{ij} = \begin{cases} p & \text{if both } i,j \in T \text{ or } i,j \in S \\ q & \text{if } i \in T, j \in S \text{ or } i \in S, j \in T. \end{cases}$$

We can think of S and T as disjoint "communities" in our graph. Nodes are connected randomly, but it is more likely that they are connected to members of their community than members outside their community.

Our goal is to design a spectral method to recover these underlying communities. Today, we will just give a sketch of an algorithm/proof.

Let's introduce another matrix $B \in \mathbb{R}^{n \times n}$ defined as follows:

$$B_{ij} = \begin{cases} p & \text{if } i, j \in T \text{ or } i, j \in S \\ q & \text{if } i \in T, j \in S \text{ or } i \in S, j \in T. \end{cases}$$

It is not hard to see that $B = \mathbb{E}[A] + pI$, where I is an $n \times n$ identity. Accordingly, at least in expectation, A has the eigenvectors as B. What are these eigenvectors?

B is rank two, so it only has two, u_1 and u_2, where:

$$u_1(i) = \frac{1}{\sqrt{n}} 1 \; \forall i,$$

$$u_2(i) = \begin{cases} \frac{1}{\sqrt{n}} 1 & \forall i \in S \\ \frac{1}{\sqrt{n}} - 1 & \forall i \in T. \end{cases}$$

$Bu_1 = \frac{n}{2}(p+q)u_1$ and $Bu_2 = \frac{n}{2}(p-q)u_2$. In this case, u_1 and u_2 are also B's singular vectors.

So, if we could compute B's eigenvectors, we could immediately recover our community by simply examining u_2. Of course, we don't have access to B, but we do have accesses to a perturbed version of the matrix via:

$$\hat{A} = A + pI.$$

Consider $R = B - \hat{A}$. Classic perturbation theory results in linear algebra tell us that if $\|R\|_2$ is small, then $\hat{A}$'s eigenvalues and eigenvectors will be close to those of B.

Let B have eigenvectors $u_1, \ldots, u_n$ and eigenvalues $\lambda_1, \ldots, \lambda_n$. Let $\hat{A}$ have eigenvectors $\hat{u}_1, \ldots \hat{u}_n$ and eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$.

Claim 1. *If B and $\hat{A}$ are real symmetric matrices with $\|B - \hat{A}\|_2 \le \epsilon$, $\forall i$,*

$$|\lambda_i - \hat{\lambda}_i| \le \epsilon.$$

In words, if $\hat{A}$ and B are close in spectral norm, their eigenvalues are close.

The proof of this claim is based on the following Courant-Fisher theorem, which is very useful in general and shows that eigenvalues of a real symmetric matrix arise as solutions to a natural optimization problem. We recall that for any non-zero vector x, $\frac{x^\top M x}{\|x\|_2^2}$ is called the *Rayleigh quotient* of x with respect to M.

Theorem 1 (Courant-Fisher Theorem). *Let M be a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$. Then,*

$$\lambda_k = \max_{S:\ subspace\ , \dim(S)=k} \min_{x \in S, x \neq 0} \frac{x^\top M x}{\|x\|_2^2} = \min_{S:\ subspace\ \dim(S)=n-k+1} \max_{x \in S, x \neq 0} \frac{x^\top M x}{\|x\|_2^2}$$

Proof. We will only sketch the proof of the first equality since the second is similar. Let $u_1, u_2, \ldots, u_n$ be the unit length eigenvectors of M corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

To prove equality, we will prove that the LHS is at least the RHS and is at most the RHS separately.

For the first, it is enough to show that there is a subspace S of dimension k such that every vector in x makes the Rayleigh quotient at least λ_k. For this, choose S to be the span of $\{u_1, u_2, u_k\}$. Any vector x in the span can be written as $\sum_{i \leq k} \alpha_i u_i$ for $\sum_{i \leq k} \alpha_i^2 = 1$. Finally, observe that by direct computation $x^\top M x = \langle x, Mx \rangle = \sum_i \alpha_i^2 \lambda_i \geq \lambda_k \sum_{i \leq k} \alpha_i^2 = \lambda_k$.

For the second, consider any subspace S of dimension $n - k + 1$. The key observation is that S must intersect the k-dimensional subspace spanned by $\{u_1, u_2, \ldots, u_k\}$ in a non-zero vector v. Because if not, then the direct sum of S and the span of $\{u_1, u_2, \ldots, u_k\}$ will be of dimension $n - k + 1 + k > n$. Then v can be written as $\sum_{i \leq k} \alpha_i u_i$ as above and $v^\top M v \geq \lambda_k$ by the same argument as above.

□

We will state without proof (but this will be a homework problem) the following consequence of Courant-Fisher theorem that immediately implies the above claim.

Corollary 1. *Let M, M' be two symmetric, real $n \times n$ matrices. Then, for any $1 \leq k \leq n$,*

$$\lambda_k(M) + \lambda_n(M') \leq \lambda_k(M + M') \leq \lambda_k(M) + \lambda_1(M')$$

Exercise 1. *Prove the Corollary from the theorem above and apply it to complete the proof of the eigenvalue closeness claim about $\hat{A}$ and B above.*

For our application, we further need that the matrices *eigenvectors are close*. Below is a classic result quantifying this – you can find a simple proof of a slightly weaker version in [1].

[1]

Theorem 2 (Davis-Kahan, 1970). *Suppose B and $\hat{A}$ are real symmetric matrices with $\|B - \hat{A}\|_2 \leq \epsilon$. Let θ_i denote the angle between u_i and $\hat{u}_i$. For all i,*

$$\sin \theta_i \leq \frac{\epsilon}{\min_{j \neq i} |\lambda_i - \lambda_j|}.$$

Let's unpack this claim. It says that if B and $\hat{A}$ are close in spectral norm, their corresponding eigenvectors are close. However, the distance is effected by a factor of $1/|\lambda_i - \lambda_j|$. This makes sense – suppose $\lambda_i < \lambda_{i+1} + \epsilon$. Then a pertubation with spectral norm ϵ could cause the u_i and u_{i+1} to "swap" order – specifically just add $\epsilon u_{i+1} u_{i+1}^T$ to B to cause such a change. In the perturbed matrix, $\hat{u}_i = u_{i+1}$, which is orthogonal to u_i.

Fortunately, in our case, we have a gap between B's eigenvalues – in particular, $|\lambda_2 - \lambda_1| \geq nq$ and $|\lambda_2 - 0| = \frac{n}{2}(p-q)$. Let's assume a challenging regime where q is close to p and thus $\frac{n}{2}(p-q) \leq nq$).

A simple corollary of Claim 2 is that $\|u_i - \hat{u}_i\|_2 \leq \frac{\sqrt{2}\epsilon}{\min_{j \neq i} |\lambda_i - \lambda_j|}$.

As an estimate for our community indicator vector u_2, let's consider $\mathsf{sign}(\hat{u}_2)$. Suppose this estimate differs from u_2 on k entries. Then it must be that:

$$\|\hat{u}_2 - \mu_2\|_2 \geq \sqrt{\frac{k}{n}}$$

So, by the eigenvector perturbation argument, we can bound

$$k \leq O\left(\frac{\epsilon^2}{n(p-q)^2}\right)$$

Eigenvalues of Random matrices

So we are left to bound $\|R\|_2$. $R = B - \hat{A}$ is a random matrix with half of its entries equal to p with probability $(1-p)$ and $(p-1)$ with probability p, and the other half equal to q with probability $(1-q)$ and $(q-1)$ with probability q.

It is possible to prove:

Theorem 3 (Van Vu, 2007)**.** *If $p \geq O(\log^4 n/n)$, then with high probability,*

$$\|R\|_2 \leq O(\sqrt{pn})$$

You will prove a very related (but slightly looser statement on the problem set).

With this bound in place, we immediately have that our spectral algorithm recovers the hidden partition with a number of mistakes bounded by:

$$k = O\left(\frac{p}{(p-q)^2}\right).$$

This is very good. Even when $q = p - O(1/\sqrt{n})$ (e.g. our probabilities are very close, so the communities should be hard to distinguish) we only make $O(n)$ mistakes – i.e. we can guess a large constant fraction of the community identities correctly.

In the next lecture, we will discuss the spectral norm upper bound above in more detail. Right now, let us see another example application of it to yet another average-case model.

The Planted Clique Model

Back in 1976, only < 5 years after the discovery of the proof of the Cook-Levin theorem, Karp proposed studying the task of finding the maximum clique in a random graph. Precisely speaking, let $G(n, 1/2)$ be the distribution on graphs on n vertices where every edge is included independently in the graph with probability $1/2$. Such a graph is dense — it has $\sim n^2/4$ edges in expectation.

Properties of such random graphs are extremely well-studied. In particular, we know that the maximum clique in such graphs is of size $\sim 2\log_2 n$ with all but a negligible probability as $n \to \infty$. Obtaining an estimate of $\Theta(\log n)$ on the size of the maximum clique is not so hard and is an application of the first and second moment method. First, calculate the expected number of k-cliques for a parameter k by linearity of expectation. Notice that this turns out to be $2^{-\binom{k}{2}}\binom{n}{k}$ and is $\ll 1$ if $k \gg 2\log_2 n$ which already implies by Markov's inequality that the maximum clique cannot be more than $O(\log n)$ in size. To show that there is a clique of size $\sim 2\log_2 n$, we need to compute the variance of the random variable that counts the number of k-cliques. This takes some more effort but is still elementary.

It turns out that the concentration of the maximum clique in $G(n, 1/2)$ is much stronger than the above argument might suggest. Indeed, this random variable exhibits what is called a *two-point* concentration inequality that says that the maximum clique takes one of two possible values $\lfloor 2\log_2 n \lfloor$ and $\lfloor 2\log_2 n \lfloor +1$ with $1 - o_n(1)$ probability.

So we know that $G \sim G(n, 1/2)$ has a maximum clique of size $2\log_2 n$. Can we find it?

There's a simple greedy algorithm that finds a clique of size $\log_2 n$ (i.e. about half the size of the maximum clique): repeat until no longer possible: 1) Let $S = \emptyset$, 2) take any vertex in the common neighborhood of S and add it to S.

Exercise 2. *Analyze the above algorithm and argue that it finds a clique of size* $\geq log_2 n - o(\log_2 n)$ *with probability at least* 0.99.

The idea of the analysis is easy: each time we add a vertex, the size of the common neighborhood slashes by $\sim 1/2$ and the edges between the vertices in the common neighborhood are independent of all the "randomness" we have seen so far.

Karp asked if we could improve on this algorithm. No improved algorithm is known so far. There are lower bounds on restricted class of methods that at least justify our failure in finding better algorithms somewhat [2].

[2]

The Planted Clique Model Planted clique model $G(n, 1/2, k)$ is a variant of the above question of Karp. In this model, the input graph is generated by taking a $G \sim G(n, 1/2)$, taking a fixed set of k vertices and adding edges to create a k-clique on it. We choose $k \gg 2\log_2 n$ so the maximum clique in the new graph must be the added or planted clique since there was no clique larger than $2\log_2 n$ in the graph before. The goal of the algorithm is to find the vertices in the planted k-clique.

There is a simple quasipolynomial algorithm for the problem that searches for $3\log_2 n$ size cliques iteratively and observes that each such clique must be a subclique of the planted k-clique.

Exercise 3. *Flesh out the details of this simple method to recover the planted k-clique*

The key algorithmic question is to find the planted k-clique (with high probability over the draw of the graph) in polynomial time.

We will solve a related but easier question in this lecture. Given a graph G that is either generated by taking a sample from $G(n, 1/2)$ ("random") or from $G(n, 1/2, k)$ ("planted"), decide correctly which model was used to generate it.

The following is an easy by key observation that relates the spectral norm of a (variant of) the adjacency matrix of a graph and the size of its largest clique.

Lemma 1. *Let G be a graph with a clique of size ω. Let A be the ± 1-adjacency matrix of G. That is, $A(i, j) = +1$ if $\{i, j\}$ is an edge in G and -1 otherwise. Then,*

$$\|A\|_2 + 1 \geq \omega$$

Proof. Let $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$ be the eigenvalues of A. By the Courant-Fisher theorem, we know that

$$\|A\|_2 \geq \lambda_1(A) = \max_{x \neq 0} \frac{x^\top A x}{\|x\|_2^2}$$

Thus, to prove the lemma, it is enough to find a vector x such that the ratio in the RHS is at least $\omega - 1$. We take the 0-1 indicator x vector of the ω-clique, say S, in G. Note that $\|x\|_2^2 = \omega$.

Further, $x^\top Ax = \sum_{i,j} x_i x_j A(i,j) = \omega(\omega-1)$ since only the $\{i,j\}$ such that $i,j \in S$ contribute a non-zero value and in that case, in fact each term contributes a $+1$ whenever $i \neq j$.

Thus, $\frac{x^\top Ax}{\|x\|_2^2} = \omega(\omega-1)/\omega = \omega - 1$ as desired.

□

We can now describe and analyze a distinguishing algorithm.

1. Construct the ± 1 adjacency matrix of the input graph.
2. Compute $\|A\|_2$.
3. If $\|A\|_2 \leq C\sqrt{n}$, output "random". Otherwise output "planted".

Lemma 2. *There is a constant $C > 0$ such that the above distinguishing algorithm succeeds correctly with high probability whenever $k \geq 2C\sqrt{n}$.*

Proof. The analysis is simple. Observe that A is a random matrix with independent entries up to symmetry when $G \sim G(n, 1/2)$. So by the theorem above (for $p = q = 1/2$), $\|A\|_2 \leq C\sqrt{n}$ for some constant $C > 0$.

On the other hand, by the lemma above, in the case when $k \geq 2C\sqrt{n}$, the spectral norm of $\|A\|_2 \geq 2C\sqrt{n}$. □

Bibliography

[1] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. ITCS 2014

[2] Le Gall, François. Powers of Tensors and Fast Matrix Multiplication. Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation, 296–303. 2014.

[3] Sushant Sachdeva, Nisheeth K. Vishnoi. Faster Algorithms via Approximation Theory. Foundations and Trends in Theoretical Computer Science. 2013. https://theory.epfl.ch/vishnoi/Publications_files/approx-survey.pdf.

[4] Daniel Spielman. Spectral Partitioning in a Stochastic Block Model. Lecture, Yale University. 2015. http://www.cs.yale.edu/homes/spielman/561/lect21-15.pdf.

[5] Chandler Davis, William Morton Kahan. The rotation of eigenvectors by a perturbation. SIAM Journal on Numerical Analysis, 7(1):1–46, 1970.

[6] Van Vu. Spectral norm of random matrices. Combinatorica, 27(6):721–736, 2007.