COMS W6998-5: Algorithms through Geometric Lens (Fall'17) October 18, 2017

Lecture 13 – Graph Coloring (end), Spectral Graph Theory I

Instructors: *Alex Andoni, Ilya Razenshteyn* Scribes: *Geelon So*

In today's lecture, we finish up our discussion on graph coloring from last time. Specifically, we left Lemma 7 in the previous lecture notes unproven. Then, we'll shift gears to begin talking about spectral graph theory.

1 Graph Coloring

To quickly recap the previous lectures, we've seen that if a graph G with n vertices is 3-colorable, then there must be a collection of vectors $x_1, \ldots, x_n \in \mathbb{R}^n$ such that

1. $||x_i|| = 1$, and
2. for every edge (i, j) in G, $\langle x_i, x_j \rangle = -\frac{1}{2}$.

The intuition is made clearer from the following diagram, where condition (1) restricts us to the unit sphere, and condition (2) says that any edge must be 120° apart, corresponding to the 3 colors. This embedding of the graph into $\mathbb{R}^n$ places adjacent vertices relatively far apart from each other.

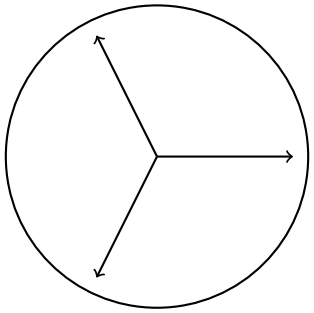

Figure 1: An embedding of the vertices of G into $\mathbb{R}^n$, where adjacent vertices are placed 120° away from each other.

We saw that using semidefinite programming (SDP), we can find an independent set I in polynomial time (then we color the independent set with a single color, and iterate on the rest of the graph). In the analysis from last class, we left unproved the following claim.

Claim 1. *We can find an independent set I such that $|I| \geq \Omega\left(\frac{n}{\Delta^{1/3} \ln n}\right)$, where Δ is the max degree of any vertex of G.*

The way we'll prove this is by cutting our collection of $\{x_1, \ldots, x_n\}$ by a random hyperplane of $\mathbb{R}^n$, and taking all the vertices on one side of the hyperplane. The hope is that since adjacent vertices are far apart, we won't capture too many edges. However, a cut through the origin may not be good enough.

For example, in Figure 2 we can see that any cut through the origin will contain an edge with probability 1/2 (it just depends on which side of the hyperplane we take).

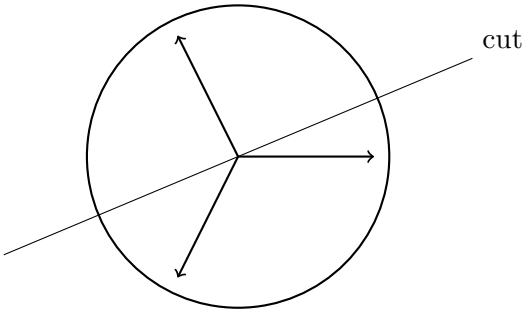

Figure 2: A random hyperplane through the origin easily fails to obtain an independent set.

So, instead of taking cuts through the origin, let's take hyperplanes lifted some distance off the origin:

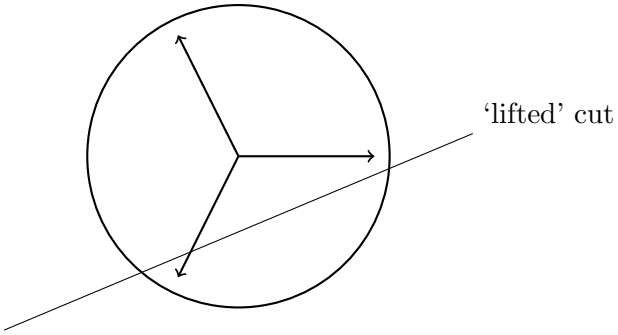

Figure 3: A random hyperplane some distance away from the origin does a better job at cutting the graph.

In this manner, once we've found a subset of $\{x_1, \ldots, x_n\}$ containing relatively few edges, call it V, then we can pick out an independent set $I \subset V$. Our goal will be to determine how far we should 'lift' these hyperplanes so that we cut out many edges, but not so much that we also cut out too many vertices of the whole graph.

Formally, we represent a random hyperplane by a random vector $g \in \mathbb{R}^n$, sampled according to an n-dimensional Gaussian distribution: $g = (g_1, \ldots, g_n) \sim \mathcal{N}(0,1)^n$. The set of points V that is captured by the lifted hyperplane is defined to be

$$V := \{i \text{ a vertex of } G : \langle x_i, g \rangle \geq t\},$$

where t is a distance parameter we hope to determine. Once we obtain V, we can find an independent set $I \subseteq V$. That is,

$$I := \{i \in V : i \text{ has no neighbors in } V\}.$$

Let's proceed with the analysis.

Step 1: Let's bound the expected size of V, which is

$$\mathbb{E}[|V|] = \sum_{i=1}^{n} \Pr\left[\langle x_i, g\rangle \geq t\right] = n \Pr\left[\langle x, g\rangle \geq t\right],$$

where x is any one of the x_i's. And since $\mathcal{N}(0,1)^n$ is spherically symmetric, without loss of generality, we may assume $x = (1, 0, \ldots, 0)$, in which case,

$$\mathbb{E}[|V|] = n \Pr[g_1 \geq t] = n \cdot \Theta\left(\frac{1}{t} \cdot e^{-t^2/2}\right), \tag{1}$$

where the latter bound applies when $t \geq 1$.

Step 2: Now, let's find the expected size of an independent set in V:

$$\mathbb{E}[|I|] = \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\{i \in I\}}\right] = \sum_{i=1}^{n} \Pr[i \in I],$$

where the probability that i is contained in the selected independent set I is given by

$$\Pr[i \in I] = \Pr[i \in V] \cdot \underbrace{\Pr\left[\text{no neighbors of } i \text{ are contained in } V \;\middle|\; i \in V\right]}_{s}.$$

If we call the latter term s, then it follows that s must be parametrized by our choice of t. Suppose that we can bound $s \geq \frac{1}{2}$. Then, $\mathbb{E}[|I|]$ becomes

$$\mathbb{E}[|I|] \geq \sum_{i=1}^{n} \Pr[i \in V] \cdot \frac{1}{2} = \frac{1}{2}\mathbb{E}[|V|] = \Theta\left(n \cdot \frac{1}{t} \cdot e^{-t^2/2}\right). \tag{2}$$

This is the term we're hoping to provide a lower bound on. So now, we should bound s from below.

Step 3: The probability that no neighbors of i are contained in V is the complement of the probability at least one neighbor of i is in V. We can apply union bound here, obtaining

$$\begin{aligned} s &= 1 - \Pr\left[\exists \text{ some neighbor of } i \in V \;\middle|\; i \in V\right] \\ &\geq 1 - \Delta \Pr\left[j \in V \;\middle|\; i \in V\right], \end{aligned}$$

where j is an arbitrary neighbor of i, and Δ bounds the number of neighbors vertex i may have. Rewriting the latter term in terms of x_i, x_j, and g, we get

$$\Pr\left[j \in V \;\middle|\; i \in V\right] = \Pr\left[\langle x_j, g\rangle \geq t \;\middle|\; \langle x_i, g\rangle \geq t\right].$$

Now, let's compute this last term. Because g is spherically symmetric, we may assume that x_i and x_j are placed as in Figure 4. Then the above probability becomes

$$\Pr\left[j \in V \;\middle|\; i \in V\right] = \Pr\left[-\frac{1}{2}g_1 + \frac{\sqrt{3}}{2}g_2 \geq t \;\middle|\; g_1 \geq t\right] \leq \Pr\left[-\frac{1}{2}t + \frac{\sqrt{3}}{2}g_2 \geq t\right] = \Pr\left[g_2 \geq \sqrt{3}t\right].$$

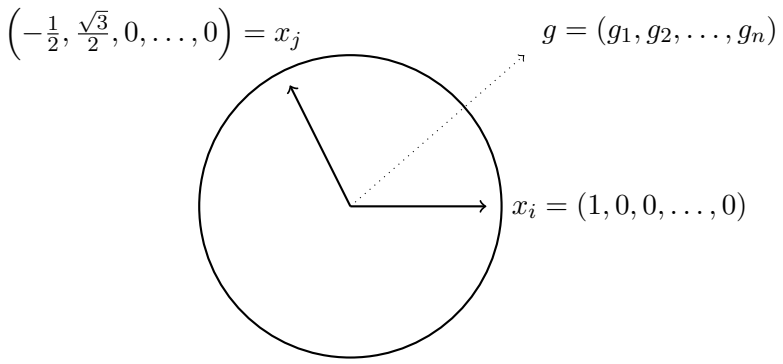

Figure 4: Spherical symmetry allows us to assume that x_i, x_j, and g are in this configuration.

Applying the same tail probability bound for Gaussians, we obtain

$$\Pr\left[j \in V \;\middle|\; i \in V\right] \leq \Theta\left(\frac{1}{\sqrt{3}t} \cdot e^{-3t^2/2}\right) \leq \Theta\left(e^{-3t^2/2}\right).$$

So, finally, we obtain

$$s \geq 1 - \Delta\Theta\left(e^{-3t^2/2}\right). \tag{3}$$

Step 4. At this point, if we want to bound $s \geq \frac{1}{2}$, this means that we require

$$\frac{1}{2} \leq \Delta\Theta\left(e^{-3t^2/2}\right) = \Delta\Theta\left(e^{-t^2/2}\right)^3,$$

which implies we want to choose t so that

$$\Theta\left(\frac{1}{\Delta^{1/3}}\right) \leq e^{-t^2/2}. \tag{4}$$

In particular, we obtain

$$\sqrt{\frac{2}{3}\ln\Theta(\Delta)} \leq t.$$

All that's left to do is to plug Equation 4 back into Equation 2 to get a bound $\mathbb{E}[|I|]$. We obtain

$$\mathbb{E}[|I|] \geq \Theta\left(n \cdot \frac{1}{t} \cdot e^{-t^2/2}\right) \geq \Omega\left(\frac{n}{\Delta^{1/3}\sqrt{\ln n}}\right),$$

since we may bound Δ with n. This completes the proof of Claim 1. □

2 Spectral Graph Theory

So far, the relationship between graph theory and linear algebra has been relatively *ad hoc*—we've found it useful to 'embed' a specific graph problem into the language of linear algebra. And from there, we can borrow existing machinery to find a solution.

But now, we'll start to view graphs more as linear-algebraic objects in their own right. For example, every undirected graph naturally gives rise to an *adjacency matrix*, fully encapsulating everything about that graph. Unsurprisingly, there's a lot we can learn about a graph under this framework.

2.1 Motivation: Diffusion

Consider a graph with n vertices. We'll imagine that each vertex represents a different state of an n-state system. At each time step, we may move from the ith state to the jth state only if i and j are connected by an edge. Furthermore, suppose that we'll move to any adjacent vertex with equal probability.

After moving for some time, we can now ask where we're likely to end up. That is, how has our *probability mass* diffused throughout the n states? Equivalently, what is the probability distribution of a random walker's position on a graph at time t? (We've formulated this question probabilistically, but we can ask the analogous question for the diffusion of physical mass).

To answer this, let's define an initial distribution at time $t = 0$: for each vertex j, assign a probability mass of $x_j(0)$, where we naturally require

$$\sum_{j=1}^{n} x_j(0) = 1.$$

Define $x_j(t)$ to be the probability mass of vertex j at time t; $x(t)$ is a probability vector of length n.

Claim 2. *Let A_{ij} be an indicator for when j is adjacent to i. Let d_j be the degree of vertex j. Then,*

$$x_i(t+1) = \sum_{j=1}^{n} A_{ij} \frac{1}{d_j} x_j(t).$$

The claim above is easily reasoned out. The mass on vertex i must all come from vertices adjacent to it. Each of these neighboring vertices j has mass $x_j(t)$, but since vertex j itself has d_j neighbors, only $\frac{1}{d_j}$ of the mass goes to vertex i.

While the formalism is perhaps a bit clunky, it does take on the form of matrix multiplication. In particular, we can make notation more compact by defining the adjacency and degree matrices:

Definition 3. *Let G be an undirected graph with n vertices. The* adjacency matrix *of G is the $n \times n$ symmetric matrix A where*

$$A_{ij} = \begin{cases} 1 & (i,j) \text{ is an edge in } G \\ 0 & \text{otherwise.} \end{cases}$$

Definition 4. *Let G be as above. The* degree matrix *of G is the $n \times n$ diagonal matrix D where*

$$D_i = d_i,$$

and d_i is the degree of the ith vertex.

It follows from Claim 2 that

$$x(t+1) = AD^{-1}x(t),$$

and so, by induction, we obtain $x(t)$ by applying the linear map AD^{-1} to the initial distribution t times:

$$x(t) = \left[AD^{-1}\right]^t x(0).$$

Finally, we can also try to find the *stationary distribution* x_*, which is naturally a distribution where $x_* = AD^{-1}x_*$. Let's rewrite D^{-1} as $D^{-\frac{1}{2}}D^{-\frac{1}{2}}$, so we obtain

$$x_* = AD^{-\frac{1}{2}}D^{-\frac{1}{2}}x_*.$$

We can now let $y_* = D^{-\frac{1}{2}}x_*$, so that $y_* = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}y_*$, where $M = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is a symmetric matrix.

We can compute the iterated applications of a matrix M and/or find its stationary distribution by calculating its eigenvalues/eigenvectors. This is just one of the motivations for us to study spectral graph theory (the *spectrum* is the set of eigenvalues of a matrix).

2.2 The Spectral Theorem

Let's review some linear algebra.

Definition 5. *Let $M \in \mathbb{R}^{n\times n}$ be a real $n \times n$ matrix. A nonzero vector $v \in \mathbb{R}^n$ is an* eigenvector *of M if there exists some $\lambda \in \mathbb{R}$ such that $Mv = \lambda v$. Such a λ is called an* eigenvalue *of v.*

If we view M as a linear transformation, then an eigenvector is a vector that is merely scaled by M. This property of eigenvectors makes it very easy for us to calculate an iterated application of M, recalling to one of our questions at the end of the previous section. Here, applying M to an eigenvector t times is just multiplying by a scalar t times, $M^tv = \lambda^t v$.

Notice that if v is an eigenvector of M, then rv is another eigenvector of M for any $r \in \mathbb{R}$, simply because $M(rv) = \lambda(rv)$. While there may be uncountably many eigenvectors, the number of eigenvalues is much smaller.

Claim 6. *Let M as above. There are at most n eigenvalues of M.*

To see this, note that the condition $Mv = \lambda v$ is equivalent to $(M - \lambda I)v = 0$. Since v is nonzero, this implies $\det(M - \lambda I) = 0$. Finally, $\det(M - \lambda I)$ is a degree n polynomial in λ, so it has at most n roots including multiplicity. While in general, these roots might be complex, under certain conditions, we have:

Theorem 7 (Spectral Theorem)**.** *If $M \in \mathbb{R}^{n\times n}$ is a symmetric $n \times n$ matrix, then there exists an* orthonormal eigenbasis *of M. That is, there is a set of n eigenvectors of M, say $v_1, \ldots, v_n$, that are pairwise orthogonal and normalized to unit length. Furthermore, the corresponding set of eigenvalues $\lambda_1, \ldots, \lambda_n$ are real. Without loss of generality, we may assume*

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

Claim 8. *Let M, v_i's, and λ_i's as above. We may rewrite M in its orthonormal eigenbasis as*

$$M = \sum_{i=1}^{n} \lambda_i v_i \cdot v_i^T.$$

(Note that $v \cdot v_i^T$ is the outer product, and not the inner product). In particular, M represented in its eigenbasis is a diagonal matrix where the ith term is λ_i.

Proof as exercise.

While the spectral theorem asserts the existence of certain eigenvectors, its proof (which we defer to a linear algebra textbook) gives a way of construction. This involves defining the Rayleigh quotient of a matrix M and vector x.

Definition 9. *Let $M \in \mathbb{R}^{n \times n}$ be a symmetric $n \times n$ matrix, and $x \in \mathbb{R}^n$ a vector. The* Rayleigh quotient *$R(M, x)$ is*

$$R(M, x) := \frac{\langle Mx, x \rangle}{||x||_2^2} = \frac{x^T M x}{||x||_2^2}.$$

If the context of M is clear, we'll suppress its notation and write $R(x)$ instead.

One property of the Rayleigh quotient is that it maps eigenvectors to its corresponding eigenvalues. That is, if v is an eigenvector of M, with eigenvalue λ, then

$$R(v) = \frac{\langle Mv, v \rangle}{||v||^2} = \frac{\langle \lambda v, v \rangle}{||v||^2} = \lambda \frac{||v||^2}{||v||^2} = \lambda.$$

Another important property is that it is scale invariant, in the sense that $R(x) = R(rx)$ for all nonzero $r \in \mathbb{R} - \{0\}$. This follows from

$$R(rx) = \frac{\langle Mrx, rx \rangle}{||rx||^2} = \frac{r^2 \langle Mx, x \rangle}{r^2 ||x||^2} = \frac{\langle Mx, x \rangle}{||x||^2} = R(x).$$

From these two properties and the spectral theorem, we can show:

Proposition 10. *Let M, v_i, and λ_i be as in Theorem 7. Denote the maximal eigenvalue λ_n by $\lambda_{\max}$. Let*

$$v_* = \arg\max_{v \in \mathbb{R}^n} R(v).$$

Then, v_* is an eigenvector of M with eigenvalue $\lambda_{\max}$.

Of course, a proof of the spectral theorem generally uses Proposition 10 as a lemma. But for now, let's not quibble about circularity.

Proof of Proposition 10. Since the Rayleigh quotient is scale-invariant, we can restrict our attention to the unit sphere S^{n-1} instead of all of $\mathbb{R}^n$. So, we may assume that

$$v_* = \arg\max_{v \in S^{n-1}} R(v).$$

(As an aside, this implies the existence of v_*, since S^{n-1} is compact and $R(v)$ is continuous). The spectral theorem gives us an orthonormal eigenbasis, $v_1, \ldots, v_n \in \mathbb{R}^n$, so we can represent v_* as $(\alpha_1, \ldots, \alpha_n)_{\text{eig}}$ in the eigenbasis.[1] In other words,

$$v_* = \sum_{i=1}^{n} \alpha_i v_i.$$

[1]The notation $(\alpha_1, \ldots, \alpha_n)_{\text{eig}}$ is nonstandard—we just use it here to specify when a vector is expanded in the eigenbasis.

From Claim 8, M is diagonal in its eigenbasis, so we readily see that $Mv_* = (\lambda_1\alpha_1, \ldots, \lambda_n\alpha_n)_{\text{eig}}$. Computing the Rayleigh quotient is now easy. Since $v_* \in S^{n-1}$ is unit length, it follows that

$$R(v_*) = \langle Mv_*, v_* \rangle = \sum_{i=1}^{n} \lambda_i \alpha_i^2.$$

And because v_* is unit, $\sum_{i=1}^{n} \alpha_i^2 = 1$. Therefore, $R(v_*)$ is actually a convex combination of the λ_i's, so it is bounded above by $\lambda_{\max}$.

By assumption, v_* achieves this maximum. Suppose the maximal eigenvalue is unique (so that $\lambda_i < \lambda_n$ for all $i < n$). Then the α_i's for $i < n$ must be zero, implying that $v_* = v_n$, up to signs (in other words, v_* is a linear combination over the unique eigenvector with eigenvalue $\lambda_{\max}$). In this case, we've shown that v_* is actually the eigenvector v_n with eigenvalue $\lambda_{\max}$.

But in general, more than one v_i may have eigenvalue $\lambda_{\max}$. In this case, v_* turns out to be a linear combination over all these v_i's. It is straightforward to compute $Mv_* = \lambda_{\max} v_*$, verifying that v_* is an eigenvector with eigenvalue $\lambda_{\max}$. □

In Proposition 10, we produced the eigenvector/value pair $(v_*, \lambda_{\max})$ by finding a vector that maximized $R(v)$ over $v \in S^{n-1}$. We can induct on this procedure (which is actually how the spectral theorem is usually proved[2]).

Let $W = \text{span}(v_*)$. Now, we restrict $\mathbb{R}^n$ to the orthogonal complement of W, denoted by

$$W^\perp := \{v \in \mathbb{R}^n : \langle v, w \rangle = 0 \text{ for all } w \in W\}.$$

For example, if $v_* = v_n$, then the orthogonal complement would correspond to all vectors that have a zero entry in the nth coordinate of the eigenbasis. Explicitly,

$$W^\perp = \{(\alpha_1, \ldots, \alpha_{n-1}, 0)_{\text{eig}} \in \mathbb{R}^n\}.$$

Crucially, the orthogonal complement of W is an invariant subspace under action by M. Put more simply, for any $v \in W^\perp$, the image Mv is contained in $W^\perp$. This is especially clear with the example where $v_* = v_n$. In this case, M just scales the ith coordinate by λ_i, so Mv is indeed in $W^\perp$. Summarized neatly,

Claim 11. *Let $w_1, \ldots w_k \in \mathbb{R}^n$ be a collection of eigenvectors of M. Let $W = \text{span}(w_1, \ldots, w_k)$ be the subspace generated over these w_i's. Then $W^\perp$ is invariant under action by M:*

$$MW^\perp \subseteq W^\perp.$$

As a result, once we've found the first v_*, call it w_n, we can restrict our analysis to the orthogonal complement of $\text{span}(w_n)$. In particular, this implies that we can compute the pair (w_{n-1}, λ_{n-1}) by

$$w_{n-1} = \underset{w \in \text{span}(w_n)^\perp}{\arg\max} R(w),$$

[2]Again, let's not worry about circularity. But as an aside, it is interesting that here, even if we treat the spectral theorem as a black box, the knowledge it provides us—that an orthonormal eigenbasis exists—will actually let us produce an orthonormal eigenbasis.

and $R(w_{n-1}) = \lambda_{n-1}$. In general,

$$w_i = \underset{w \in \text{span}(w_n, \ldots, w_{i+1})^\perp}{\arg\max} R(w),$$

and $R(w_i) = \lambda_i$. By the iterative application of Proposition 10, we've constructed an orthonormal eigenbasis of M, given by $\{w_1, \ldots, w_n\}$.