Undirected Graphs and the Spectral Theorem

Kent Quanrud

October 25, 2020

1 The Laplacian of a graph

In our discussion on electrical networks, we breifly came across the *Laplacian form* of an undirected graph. Let us reintroduce the Laplacian $L : \mathbb{R}^V \to \mathbb{R}^V$ of a weighted undirected graph $G = (V, E)$, this time from a different perspective.

The **Laplacian of an (unweighted) edge** $e = \{u, v\}$ is the rank-1 matrix

$$L_e = (1_u - 1_v) \otimes (1_u - 1_v)$$

where $1_u \in \{0,1\}^V$ denotes the indicator vector[1] for u. Here $a \otimes b$ denotes the outer product of two vectors a, b, defined by $\langle x, (a \otimes b)y \rangle = \langle a, x \rangle \langle b, y \rangle$. Note that the expression for L_e is indifferent to whether we wrote $1_u - 1_v$ or $1_v - 1_u$, as long as it is symmetric. For any input vector $x \in \mathbb{R}^V$, we have

$$\langle x, L_e x \rangle = (x_u - x_v)^2.$$

For an undirected graph $G = (V, E)$ with positive edge weights $w : E \to \mathbb{R}_{>0}$, the **Laplacian of the graph** is the corresponding weighted sum of Laplacians of the edges,

$$L = \sum_e w(e) L_e.$$

Given an input vector $x \in \mathbb{R}^V$, we have

$$\langle x, Lx \rangle = \sum_{e \in E} w(e) \langle x, L_e x \rangle = \sum_{e=(u,v) \in E} w(e)(x_u - x_v)^2.$$

That is, the L induces a simple sum of squared differences on x, based on the edges of the graph.

In fact, it induces a very familiar sum of squares. Recall that the the electrical flow problem is to minimize $\langle f, Rf \rangle$ over $f \in \mathbb{R}^E$ s.t. $Bf = d$. Here $R = \text{diag}(r)$ is the diagonal map of resistances $r \in \mathbb{R}^E_{>0}$. $d \in \mathbb{R}^V$ represents the flow demands and $B : \mathbb{R}^E \to \mathbb{R}^V$ maps flows to the net flow at each vertex. We also saw that, by first-order optimality conditions, the electrical flow is always of the form $f = B^T p$ for a set of vertex potentials p. Then we have

$$\langle f, Rf \rangle = \left\langle B^T p, RB^T p \right\rangle = \sum_{e=(u,v) \in E} r_e (p_u - p_v)^2.$$

That is, we are choosing p as to minimize the Laplacian of the graph with edge weights corresponding to the resistances.

Recall that a linear operator $A : \mathbb{R}^n \to \mathbb{R}^n$ is **symmetric** if $A = A^T$. It is easy to see that the Laplacian L is symmetric: each L_e is symmetric since in general $(a \otimes b)^T = (b \otimes a)$, and L is a positively weighted combination of L_e's. Another salient property of L is that, as a sum of squares,

$$\langle x, Lx \rangle \geq 0 \text{ for all } x \in \mathbb{R}^n.$$

These two properties make L a member of the following very importance class of linear operators.

[1] We are avoiding the conventional notation e_u for the standard basis vectors because e is so frequently used for edges.

Definition 1. *A linear operator $A : \mathbb{R}^n \to \mathbb{R}^n$ is a* **positive semi-definite** *linear operator if*

1. *A is symmetric.*
2. *$\langle x, Ax\rangle \geq 0$ for all $x \in \mathbb{R}^n$.*

A is **(strictly) positive definite** *if in addition to being positive semi-definite,*

3. *A is invertible.*

The Laplacian L is *not* invertible: $L\mathbb{1} = \mathbb{0}$. If G is connected, and we restrict to the $n-1$ space $\mathbb{R}^V/\mathbb{1}$, then L is invertible and (strictly) positive definite (see Exercise **??**).

2 The Spectral Theorem for Symmetric Maps

We note about that the Laplacian $L : \mathbb{R}^V \to \mathbb{R}^V$ of an undirected graph is a symmetric and positive semi-definite map. We have seen before in our discussion of random walks that the eigenvectors can be very insightful in understanding the behavior of a linear map. We will see here that the eigenvectors of symmetric linear maps are particularly well behaved.

Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. Recall that a vector $x \in \mathbb{R}^n$ is an **eigenvector** of A with **eigenvalue** $\lambda \in \mathbb{C}$ of $Ax = \lambda x$. Recall the following facts proven in our earlier discussions, which apply generally to all linear maps.

Fact 1. *Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. Then A has an eigenvalue $\lambda \in \mathbb{C}$ and eigenvector $x \in \mathbb{C}^n$.*

If A is symmetric, then we can strengthen this fact to assert a real-valued eigenvalue and eigenvector with real-valued coordinates.

Lemma 2. *Let $L : X \to X$ be a symmetric linear map in a vector space X over $\mathbb{R}$. Let x maximize $\langle x, Lx\rangle$ subject to $\|x\| = 1$. Then $Lx = \lambda x$ for $\lambda = \langle x, Lx\rangle$.*

Proof. We claim that for any $u \in X$ with $\|u\| = 1$ and $\langle u, x\rangle = 0$, $\langle u, Lx\rangle = 0$. If Lx is orthogonal to u for every u orthogonal to x, then we must have $Lx \in \operatorname{span}(x)$; i.e., $Lx = \lambda x$ for some $\lambda \in \mathbb{R}$. Upon inspection, $\lambda = \lambda\langle x, x\rangle = \langle x, Lx\rangle$, as claimed.

Let $u \in X$ with $\|u\| = 1$ and $\langle u, x\rangle = 0$. Define

$$f(\epsilon) = \left\langle \frac{x + \epsilon u}{\sqrt{1+\epsilon^2}}, L\left(\frac{x+\epsilon u}{\sqrt{1+\epsilon^2}}\right)\right\rangle = \frac{\langle x + \epsilon u, L(x + \epsilon u)\rangle}{1+\epsilon^2}.$$

$f(\epsilon)$ can be interpreted as perturbing x slightly in the direction of u and renormalizing, and then computing the inner product over L. Note that $\|x + \epsilon u\|^2 = \|x\|^2 + \epsilon^2\|u\|^2 = 1 + \epsilon^2$, so $\frac{x+\epsilon u}{\sqrt{1+\epsilon u}}$ is indeed a normal vector that competes with x in maximizing $\langle x, Lx\rangle$. In particular, by choice of x, $f(\epsilon)$ is maximized at $f(0) = \langle x, Lx\rangle$. Optimality at 0 implies that $f'(0) = 0$. Expanding out $f'(0)$, we find that $\langle u, Lx\rangle = 0$, as desired. (See Exercise **??**). ■

Remark 3. An alternative proof starts from the fact there exists a complex eigenvalue and eigenvector, and goes on to show that this eigenvalue must be real-valued and that there is a corresponding eigenvector with real-valued coordinates.

The above lemma implies that beyond simply having an eigenvalue, a symmetric map has a real-valued eigenvalue without extending the field to include complex values. This simple fact leads to the following sweeping theorem about symmetric matrices.

Theorem 4. *Let X be an n-dimensional vector space over $\mathbb{R}$. Let $A : X \to X$ be a symmetric linear map. Then there exists an orthonormal basis $u_1, \ldots, u_n$ of X and n scalar values $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that*

$$A = \lambda_1(u_1 \otimes u_1) + \lambda_2(u_2 \otimes u_2) + \cdots + \lambda_n(u_n \otimes u_n).$$

Proof. If $n = 0$, then the claim is tautological, as X is the trivial vector space $\{\mathbb{0}\}$ and A can be expressed as an empty sum. Suppose $n \geq 1$. By Lemma 2, A has a real-valued eigenvalue λ with a corresponding eigenvector $u \in X$. By scaling u, we may assume $\|u\| = 1$. Consider the map $B = A - \lambda(u \otimes u)$. B is also symmetric, and maps the space $\text{span}(x) = \{\alpha x : \alpha \in \mathbb{R}\}$ to $\mathbb{0}$. Let $Y = \{y \in X : \langle x, y \rangle = 0\}$ be the subspace of X orthogonal to x. We have $\dim(Y) = n - 1$.

We claim that B maps Y into Y. Indeed, for any $y \in Y$, we have

$$\langle x, By \rangle = \langle Bx, y \rangle = \langle \mathbb{0}, y \rangle = 0,$$

so $By \in Y$.

Thus B restricts to a linear and symmetric operator on Y. By induction on n, there is an orthonormal basis $u_1, \ldots, u_{n-1}$ of Y and scalar values $\lambda_1, \ldots, \lambda_{n-1} \in \mathbb{R}$ such that

$$B = \lambda_1(u_1 \otimes u_1) + \cdots + \lambda_{n-1}(u_{n-1} \otimes u_{n-1}).$$

Let $\lambda_n = \lambda$ and $u_n = u$. Observe that $u_1, \ldots, u_n$ is an orthonormal basis of X. We have

$$\lambda_1(u_1 \otimes u_1) + \cdots + \lambda_n(u_n \otimes u_n) = B + \lambda_n(u_n \otimes u_n) = A,$$

as desired. ■

Theorem 4 makes the structure of any symmetric map $A : \mathbb{R}^n \to \mathbb{R}^n$ extremely transparent. By Theorem 4, let $u_1, \ldots, u_n \in \mathbb{R}^n$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ be such that

$$A = \lambda_1(u_1 \otimes u_1) + \cdots + \lambda_n(u_n \otimes u_n).$$

It will be convenient to assume that that λ_i's are in nonincreasing order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

For any input vector $x \in \mathbb{R}^n$, we can write x uniquely in the form

$$x = \alpha_1 u_1 + \cdots + \alpha_n u_n;$$

namely, with $\alpha_i = \langle x, u_i \rangle$. Then we have

$$Ax = \lambda_1(u_1 \otimes u_1)x + \cdots + \lambda_n(u_n \otimes u_n)x = \lambda_1 \alpha u_1 + \cdots + \lambda_n \alpha_n u_n.$$

That is, in the basis $\{u_1, \ldots, u_n\}$, A simply rescales the ith coordinate by a factor of λ_i. That is to say:

Every symmetric matrix is a diagonal matrix up to a rotation (i.e., change in basis).

We can see from the construction in the proof that the λ_i's are the eigenvalues of A and the u_i's are eigenvectors. But this fact is even more obvious in hindsight given the representation

For each i, we have

$$Au_i = \lambda_i(u_i \otimes u_i)u_i = \lambda_i u_i,$$

by orthonormality of the u_i's.

Theorem 5. *Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a symmetric linear operator. Let $\lambda_1, \ldots, \lambda_n$ be the n eigenvalues of A (with multiplicity) in decreasing order. Then*

$$\lambda_k = \min_{S : \dim(S) = k-1} \max_{x \in X/S} \frac{\langle x, Lx \rangle}{\langle x, x \rangle}.$$

3 The Laplacian and cuts

Let $G = (V, E)$ be an undirected graph with m edges and n vertices, and positive edge weights $w : E \to \mathbb{R}_{>0}$. Let $L : \mathbb{R}^V \to \mathbb{R}^V$ be the Laplacian of G. A helpful way to interpret to Laplacian is to see that it encodes all the cuts of G. Let $S \subset V$ be a set of vertices, and let $\mathbb{1}_S \in \{0,1\}^V$ be the indicator vector for S. Then we can extract the weight of the cut $\partial(S)$ from the Laplacian as

$$\langle \mathbb{1}_S, L\mathbb{1}_S \rangle = \sum_{e \in \partial(S)} w(e).$$

Recall that the *(uniform) sparsest cut* of an G is the set S minimizing the ratio

$$\frac{\sum_{e \in \partial(S)} w(e)}{|S|(n - |S|)}.$$

The minimum such quantity over a graph G is called the *sparsity* of G and denoted

$$\Psi(G) = \min_{S \subset V} \frac{\sum_{e \in \partial(S)} w(e)}{|S|(n - |S|)}.$$

Let us now relate cuts to some of the eigenvectors of L. Observe that because L is positive semi-definite, all its eigenvalues are nonnegative. Moreover, we know that $\mathbb{1}$ is an eigenvector with eigenvalue of 0 – this gives us our smallest eigenvector. The eigenvector corresponding to the second smallest eigenvalue, denoted λ_{-2}, is given by

$$\lambda_{-2} = \min_{x : \langle \mathbb{1}, x \rangle = 0} \frac{\langle x, Lx \rangle}{\langle x, x \rangle}.$$

Now, consider any cut $\mathbb{1}_S$, and let x be the orthogonal projection from $\mathbb{1}$; namely,

$$x = \mathbb{1}_S - \alpha \mathbb{1} \text{ where } \alpha = \langle \mathbb{1}_S, \mathbb{1} \rangle / \langle \mathbb{1}, \mathbb{1} \rangle = |S|/n$$

Observe that

$$\langle x, x \rangle = \langle x, \mathbb{1}_S \rangle = (1 - \alpha)|S| = (n - |S|)|S|/n.$$

Thus

$$\lambda_{-2} \leq n \frac{\sum_{e \in \partial(S)} w(e)}{|S|(n - |S|)}.$$

Taking the minimum over all sets S, we obtain the following.

Theorem 6. *Let $G = (V, E)$ be an undirected graph with m edges and n vertices, and positive edge weights $w : E \to \mathbb{R}_{>0}$. Let $L : \mathbb{R}^V \to \mathbb{R}^V$ be the Laplacian of G. Let λ_{-2} be the second smallest eigenvalue of L and let $\Psi(G)$ be the sparsity of G. Then*

$$\lambda_{-2} \leq n\Psi(G).$$

We will come back to this discussion later.

4 Random walks in undirected graphs

Let $G = (V, E)$ be an undirected graph with m edges and n vertices, and positive edge weights $w : E \to \mathbb{R}_{>0}$. Let us assume that G is connected. Let $R : \mathbb{R}^V \to \mathbb{R}^V$ be the random walk map of G. Recall that $R = AD^{-1}$, where $A : \mathbb{R}^V \to \mathbb{R}^V$ is the weighted adjacency map and $D = \text{diag}(A\mathbb{1})$ is the diagonal map of weighted vertex degrees. Recall that R was the beneficiary of the *Perron-Frobenius theorem*, which for random walks gave us a lot of information about the eigenvalues and eigenvectors of R. In particular, all of the eigenvalues of R lie in the range $[-1, 1]$, and it has eigenvalue 1 with multiplicity 1. There is a strictly positive eigenvector for eigenvalue value that defines a unique **stationary distribution**. Before, we proved the existence of a unique stationary distribution for strongly connected *directed* random walk. For undirected random walk, the stationary distribution is very straightforward.

Theorem 7. *Let $G = (V, E)$ be an undirected graph with m edges and n vertices, and positive edge weights $w : E \to \mathbb{R}_{>0}$. Let $R : \mathbb{R}^V \to \mathbb{R}^V$ be the random walk map of G. Then R has stationary distribution proportional to the weighted degrees of its vertices.*

Proof. We have

$$R(A\mathbb{1}) = A(\operatorname{diag}(A\mathbb{1}))^{-1} A\mathbb{1} = A\mathbb{1}.$$

■

While R is not a symmetric map[2], we can extend the spectral theorem for symmetric maps to R by way of *similarity*.

Definition 8. *Two linear maps $A, B : X \to X$ are* **similar** *if $A = C^{-1}BC$ for an invertible map $C : X \to X$.*

Lemma 9. *Let A and B be similar. Then their kernels are isomorphic. In particular, if $A = C^{-1}BC$, then C restricts is an isomorphism between $\ker(A)$ and $\ker(B)$.*

Lemma 10. *Let $A, B : X \to X$ be two linear maps. If A and B are similar, then A and B have the same eigenvalues with the same multiplicities. If $A = C^{-1}BC$, then C maps eigenvectors of A to eigenvectors of B with the same eigenvalues.*

Proof. For all λ, $A - \lambda I$ and $B - \lambda I$ are also similar. ■

We introduce the **normalized walk matrix** as the map $Q : \mathbb{R}^V \to \mathbb{R}^V$ defined by

$$Q = D^{-1/2} R D^{1/2} = D^{-1/2} A D^{-1/2}.$$

By the first equality above, Q is similar to R, and thus has all its eigenvalues in the range $[-1, 1]$ and eigenvalue 1 with multiplicity 1. On the other hand, by the second equality, Q *is symmetric*. As such, it has an orthonormal basis of eigenvectors. Let $1 = \lambda_1, \ldots, \lambda_n \geq -1$ list the eigenvalues of Q in decreasing order. Let $u_1, \ldots, u_n$ be an orthonormal basis of $\mathbb{R}^V$ such that

$$Q = u_1 \otimes u_1 + \lambda_2(u_2 \otimes u_2) + \cdots + \lambda_n(u_n \otimes u_n).$$

(Here we substituted $\lambda_1 = 1$). We also know, from Theorem **??**, that $D^{1/2}u_1$ must correspond to the uniform distribution, which is proportional to d. Thus

$$u_1 = \frac{D^{-1/2}(d)}{\left\|D^{-1/2}(d)\right\|} = \frac{1}{\sqrt{m}}\sqrt{d},$$

where $\sqrt{d}$ represents the entrywise square root of d.

$$Q = \frac{1}{m}\left(\sqrt{d} \otimes \sqrt{d}\right) + \lambda_2(u_2 \otimes u_2) + \cdots + \lambda_n(u_n \otimes u_n).$$

Let us now consider the *convergence rate* of a random walk. Let $x \in \Delta^V$ be any initial probability distribution over Q. We want to understand the distribution $R^k x$ obtained after k steps of the random walk. Observe first that

$$\begin{aligned}
R^k x &\overset{\text{(a)}}{=} D^{1/2} Q^k D^{-1/2} x \\
&= D^{1/2}\left(\frac{1}{m}\left(\sqrt{d} \otimes \sqrt{d}\right) + \lambda_2(u_2 \otimes u_2) + \cdots + \lambda_n(u_n \otimes u_n)\right)^k D^{-1/2} x \\
&\overset{\text{(b)}}{=} D^{1/2}\left(\frac{1}{m}\left(\sqrt{d} \otimes \sqrt{d}\right) + \lambda_2^k(u_2 \otimes u_2) + \cdots + \lambda_n^k(u_n \otimes u_n)\right) D^{-1/2} x \\
&= \frac{1}{m} d + \left(\lambda_2^k(u_2 \otimes u_2) + \cdots + \lambda_n^k(u_n \otimes u_n)\right) D^{-1/2} x
\end{aligned}$$

[2] unless G is regular; see Exercise **??**

(a) substitutes in $R = D^{1/2}QD^{-1/2}$, where the $D^{-1/2}$ and $D^{1/2}$ terms between Q's cancel out. (b) is because the u_i's are othonormal[3] (!) – here we see some of the power of the spectral theorem.

Consider the RHS of the last equation above. We see the stationary distribution, $(1/m)d$, followed by a messy term induced by the non-dominant eigenvalues and eigenvectors. That is, the difference between $R^k x$ and the stationary distribution is precisely

$$D^{1/2}\big(\lambda_2^k(u_2 \otimes u_2) + \cdots + \lambda_n^k(u_n \otimes u_n)\big)D^{-1/2}x.$$

Let $S = \lambda_2^k(u_2 \otimes u_2) + \cdots + \lambda_n^k(u_n \otimes u_n)$; S is symmetric, with eigenvalues $0, \lambda_2^k, \ldots, \lambda_n^k$. Let $\Delta_{\max}$ be the maximum degree in G and let $\Delta_{\min}$.

$$\begin{aligned}
\left\|R^k x - \frac{1}{m}d\right\|^2 &= \left\|D^{1/2}SD^{-1/2}x\right\|^2 \\
&= \left\langle SD^{-1/2}x, D\left(SD^{-1/2}x\right)\right\rangle \\
&\leq \Delta\left\|SD^{-1/2}x\right\|^2 \\
&= \Delta\left\langle D^{-1/2}x, S^2\left(D^{-1/2}x\right)\right\rangle \\
&\leq \max\left\{\lambda_2^{2k}, \lambda_n^{2k}\right\}\Delta\left\langle x, D^{-1}x\right\rangle \\
&\leq \max\left\{\lambda_2^{2k}, \lambda_n^{2k}\right\}\frac{\Delta_{\max}}{\Delta_{\min}}.
\end{aligned}$$

Recall that $\lambda_2, \lambda_n \in [-1, 1)$. If λ_2 and λ_n are both bounded away from both 1 and -1, then $\max\left\{\lambda_2^{2k}, \lambda_n^{2k}\right\} = \max\{\lambda_2, |\lambda_n|\}^{2k} \to 0$ as $k \to \infty$. To this end, the **spectral gap** of a random walk R is defined as the difference

$$\gamma = 1 - \max\{\lambda_2, |\lambda_n|\},$$

where λ_2 is the second largest eigenvalue.

Theorem 11. *Let $G = (V, E)$ be an undirected graph with m edges and n vertices, and positive edge weights $w : E \to \mathbb{R}_{>0}$. Let G be connected. Let $d \in \mathbb{R}^V_{>0}$ be the weighted degrees of the vertices. Let $\Delta_{\max} = \max_v d(v)$ be the maximum weighted degree and let $\Delta_{\min} = \min_v d(v)$ be the minimum weighted degree. Let $R : \mathbb{R}^V \to \mathbb{R}^V$ be the random walk map of G and let γ be the spectral gap of R.*

For any initial distribution $x \in \Delta^V$, x converges to the stationary distribution d/m at a rate of

$$\left\|R^k x - d/m\right\| \leq (1-\gamma)^k\sqrt{\Delta_{\max}/\Delta_{\min}}.$$

5 Exercises

Exercise 1. Let $G = (V, E)$ be an undirected graph with m edges and n vertices, and positive edge weights $w : E \to \mathbb{R}_{>0}$. Let $L : \mathbb{R}^V \to \mathbb{R}^V$ be the Laplacian of G. Prove that G is connected iff for any $x \notin \text{span}(\mathbb{1})$, we have $\langle x, Lx\rangle > 0$.

Exercise 2. Finish the proof of Lemma 2, by deriving the derivative $f'(\epsilon)$ and showing that $'f(0) = 0$ implies that $\langle x, Lu\rangle = 0$. Where do we use the assumption that L is symmetric?

Exercise 3. Let $G = (V, E)$ be an undirected graph with m edges and n vertices, and positive edge weights $w : E \to \mathbb{R}_{>0}$. Let $R : \mathbb{R}^V \to \mathbb{R}^V$ be the random walk map. Prove that R is symmetric iff G is regular[4].

Exercise 4. Suppose your goal was to converge to the uniform distribution via a random walk for a single vertex as fast as possible. Show that one can choose a vertex v such that, starting from an initial distribution of $x = 1_v$, the ℓ_2-distance from the stationary distribution after k steps is at most $(1-\gamma)^k$.

[3] We should point out that $(a \otimes b)(c \otimes d) = \langle b, c\rangle(a \otimes d)$

[4] A graph is regular if every vertex has the same weighted degree

Exercise 5. Let $G = (V, E)$ be an undirected graph with m edges and n vertices, and positive edge weights $w : E \to \mathbb{R}_{>0}$. Let $L : \mathbb{R}^V \to \mathbb{R}^V$ be the Laplacian of G. Suppose the spectral gap γ is at least some constant, say $\gamma = 1/2$. (Such a graph is called an **expander**).

1. Show that the diameter of G is at most $O(\log n)$.
2. Recall the (s, t)-connectivity problem for which we showed that a random walk gives a $O(\log n)$. Suppose also that G is has constant maximum degree (say, maximum degree 42). Give a deterministic, polynomial time, $O(\log n)$-space algorithm for (s, t)-connectivity on G.