The Matrix Tree Theorem

Rich Schwartz

January 30, 2020

0.1 The Main Result

Let G be a connected graph with vertex set V. Given a function $f : V \to \boldsymbol{R}$ we define the *Laplacian* $\Delta f : V \to \boldsymbol{R}$ by the formula

$$\Delta f(v) = kf(v) - \sum_{i=1}^{k} f(w_i). \tag{1}$$

Here $w_1, ..., w_k$ are the vertex neighbors of v. The Laplacian is a discrete analogue of the similarly named operator from calculus. The Laplacian of a function measures how far a function is from being equal at each vertex to the average of its neighbors.

If we label the vertices of G from 1 to n then we can identify $\boldsymbol{R}^n$ with vector space of all functions $f : V \to \boldsymbol{R}$. We have the *Laplace operator* $\Delta : \boldsymbol{R}^n \to \boldsymbol{R}^n$, which is just the Laplacian defined relative to the standard basis.

The representation of Δ in the standard basis is the matrix M such that all entries are 0 except:

- $M_{ij} = -1$ if and only if an edge of G joins i to j.
- M_{ii} is the degree of vertex i.

The matrix M is symmetric and every row and column sums to 0.

Let M_{11} denote the minor of M obtained by crossing off the first row and column. Here is the famous Matrix-Tree Theorem:

Theorem 0.1 *The number of spanning trees of G is $|\det(M_{11})|$.*

At the end, I'll prove two reformulations of this result.

1. The number of spanning trees of G is $|\det(M_{ij})|$ where M_{ij} is the minor of M obtained by crossing off the ith row and jth column from M. In other words, all these minors have the same determinant up to sign.

2. All the nonzero eigenvalues of M are real, and the number of spanning trees of G is
$$\frac{1}{n}\lambda_2 ... \lambda_n,$$
where $\lambda_2, ..., \lambda_n$ are the nonzero eigenvalues of M.

0.2 The Cauchy-Binet Theorem

We start with a result from linear algebra. Let A be an $n \times N$ matrix and let B be an $N \times n$ matrix. Here $n \geq N$. The matrix AB is an $n \times n$ matrix. Given any subset $S \subset \{1, ..., N\}$ having n-elements, form the two $n \times n$ matrices A_S and B_S, obtained by just using the rows (or columns) indexed by the set S. Define

$$f(A,B) = \det(AB), \qquad g(A,B) = \sum_S \det(A_S)\det(B_S). \tag{2}$$

The sum ranges over all choices of S.

Theorem 0.2 (Cauchy-Binet) $f(A,B) = g(A,B)$.

Proof: Think of A and B each as n-tuples of vectors in $\boldsymbol{R}^N$. We get these vectors by listing out the rows of A and the columns of B. So, we can write

$$f(A,B) = f(A_1, ..., A_n, B_1, ..., B_n), \tag{3}$$

and likewise for g. The values of f and g change in the same way when the following operations are performed:

- Replace A_i by λA_i.
- Replace A_i to $A_i + A_j$.
- Swap A_i and A_j.

The same goes when B is used in place of A. So, we can do elementary row and column operations on A and B the rows of A are $(1,0,...,0)$, $(0,1,0,...,0)$, etc. and similarly for the columns of B. But in this case the result is quite easy. ♠

The proof of the Cauchy-Binet Theorem is done, but here's a bit more discussion. One can think about it in a slightly more geometric way. Let $\mathcal{M}_{n,N}$ denote the set of $n \times N$ matrices. Let (N,n) denote N choose n. There is a nice map from $\mathcal{M}_{n,N}$ into $\boldsymbol{R}^{(N,n)}$ called the *Plucker embedding*. Given the matrix A, you just enumerate the subsets S above, as $S_1, ..., S_{(N,n)}$ and then you define

$$\phi(A) = (\det(A_{S_1}), \det(A_{S_2}), ...)$$

To make the notation nicer, we define $\phi(B) = \phi(B^t)$ when B is an $N \times n$ matrix. The Cauchy-Binet Theorem says that

$$\det(AB) = \phi(A) \cdot \phi(B).$$

In other words, you take the Plucker embedding of the two matrices and then take the dot product of the result, and this computes the determinant of the product.

0.3 Proof of the Matrix Tree Theorem

We already have said that G has n vertices. Suppose that G has N edges. Let us call the edges $e_1, ..., e_N$. We know that $N \geq n-1$. Let A be the $n \times N$ matrix (n rows and N columns) with the following description: Suppose v_i and v_j are the two vertices incident to e_k and $i < j$. Then $A_{ik} = 1$ and $A_{jk} = -1$. All other entries are 0. Figure 1 shows an example graph.

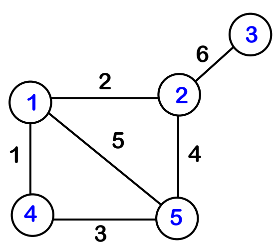

Figure 1: An example graph.

Here is the matrix:

$$\begin{array}{cc} & \textbf{edges:} \\ \textbf{vertices:} & \begin{bmatrix} +1 & +1 & 0 & 0 & +1 & 0 \\ 0 & -1 & 0 & +1 & 0 & +1 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ -1 & 0 & +1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 & 0 \end{bmatrix} \end{array}.$$

Lemma 0.3 *$M = AA^t$, Where A^t is the transpose of A.*

Proof: To see this note that:

$$(AA^t)_{ij} = A_{i1}A_{j1} + A_{i2}A_{j2} + ... + A_{iN}A_{jN}.$$

When $i = j$, this is exactly the number of edges incident to v_i. When $i \neq j$ this expression is only nonzero if there is some k such that $a_{ik} = \pm 1$ and $a_{jk} = \mp 1$. In this case, the expression is -1. Since M has the same description, we see that $M = AA^t$. ♠

Remark: The calculus analogue of this statement is that $\Delta f = \mathrm{Div}(\nabla f)$, the divergence of the gradient of f. A fancier way to write this is that $\Delta = *d*d$ where d is exterior differentiation and $*$ is the Hodge star operator.

Letting A_1 be the matrix obtained from A by deleting the first row (corresponding to vertex 1) we get

$$M_{11} = A_1 A_1^t. \tag{4}$$

Given an $n-1$ element subset $S \subset \{1, ..., N\}$, consider the square matrix $A_{1,S}$ consisting of the columns of A_1 corresponding to S. At the same time, let G_S denote the subgraph of G whose edges are in S. Taking $A = A_1$ and $B = A_1^t$ in the Cauchy-Binet Theorem, we have

$$|\det(M_{11})| = \sum_S \left(\det(A_{1,S})\right)^2 \tag{5}$$

The following two facts finish the proof:

1. If G_S is not a spanning tree for G then $\det(A_{1,S}) = 0$.

2. If G_S is a spanning tree for G then then $\det(A_{1,S}) = \pm 1$.

Proof of Fact 1: Suppose that G_S is not a spanning tree. Since G_S has $n-1$ edges, we see that G_S must contain a cycle. Orient the edges in the cycle so that they point from the smaller vertex to the larger vertex. Assign -1 to the edge if it is going clockwise around the cycle and $+1$ if it is going counterclockwise. Then the sum of the columns of $A_{1,S}$, with these signs, is 0. Since some column-sum is 0, the matrix $A_{1,S}$ has determinant 0. Figure 2 shows an example of how this works for the set $S = \{1, 2, 4, 5\}$. The cycle in this case involves edges $2, 4, 5$ and the identity is $\mathrm{Col2} + \mathrm{Col4} - \mathrm{Col5} = 0$.

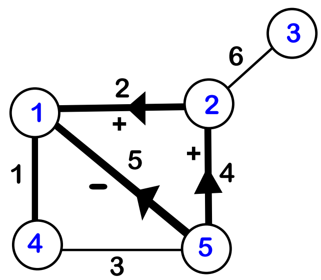

Figure 2: A cycle

We have highlighted the corresponding columns of $A_{1,S}$.

$$A_1 = \begin{bmatrix} 0 & \mathbf{-1} & 0 & \mathbf{+1} & \mathbf{0} & +1 \\ 0 & \mathbf{0} & 0 & \mathbf{0} & \mathbf{0} & -1 \\ -1 & \mathbf{0} & +1 & \mathbf{0} & \mathbf{0} & 0 \\ 0 & \mathbf{0} & -1 & \mathbf{-1} & \mathbf{-1} & 0 \end{bmatrix}.$$

Proof of Fact 2: Before we start, we note that Row $j-1$ of A_1 is Row j of A. That is, Row $j-1$ of A_1 corresponds to v_j.

Suppose that G_S is a spanning tree. We can consider Column k of $A_{1,S}$ which corresponds to a pair (v_j, e_k) where e_k is an edge of G_S and v_j is a leaf of G_S incident to e_k. We can choose $j > 1$ because every tree has at least 2 leaves. In Row $j-1$ of $A_{1,S}$ there is one nonzero entry corresponding to (v_j, e_k) and all other entries are 0 because v_j is incident to no other edges indexed by elements of S. But then

$$\det(A_{1,S}) = \pm \det(A'_{1,S}),$$

where $A'_{1,S}$ is the matrix obtained by deleting the Row $j-1$ and Column k from $A_{1,S}$. But $A'_{1,S}$ corresponds to a spanning tree for $G - v_j$ using edges of $G - e_k$. So, by induction, $\det(A'_{1,S}) = \pm 1$. Hence $\det(A_{1,S}) = \pm 1$.

Figure 3 shows an example involving $S = \{1, 2, 5, 6\}$. Here we take (edge) e_1 and (vertex) v_4. So, $j - 1 = 3$ and $k = 1$.

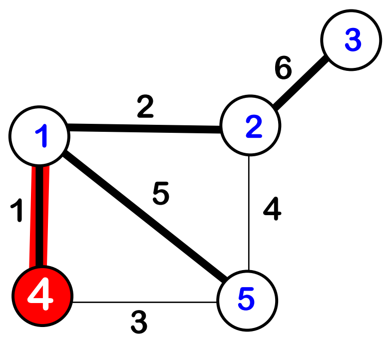

Figure 3: Deleting a leaf

We have highlighted Row 3 of $A_{1,S}$ and you can see how there is just the one nonzero entry.

$$A_{1,S} = \begin{bmatrix} 0 & -1 & * & * & 0 & +1 \\ 0 & 0 & * & * & 0 & -1 \\ \mathbf{-1} & \mathbf{0} & * & * & \mathbf{0} & \mathbf{0} \\ 0 & 0 & * & * & -1 & 0 \end{bmatrix}, \quad A'_{1,S} = \begin{bmatrix} * & -1 & * & * & 0 & +1 \\ * & 0 & * & * & 0 & -1 \\ * & * & * & * & * & * \\ * & 0 & * & * & -1 & 0 \end{bmatrix},$$

0.4 The Reformulations

The next lemma implies the first reformulation:

Lemma 0.4 $|\det(M_{ij})| = |\det(M_{11})|$ *for all* i, j.

Proof: I'll prove this for M_{11} and M_{12}. The general case is similar. Let M_1 denote the matrix obtained by crossing off the first row of M. Then

$$\det M_{1j} = \delta(V_j, V_3, ..., V_n),$$

where δ is the determinant function and V_j is the jth column of M. The function δ is linear in each position and anti-symmetric. Since $V_1 = -(V_2 + ...V_n)$, we have

$$\det M_{11} = -\delta(V_2, V_3, ..., V_m) - \delta(V_3, V_3, ..., V_n) - \delta(V_4, V_3, V_4, ..., V_n) \cdots$$

All the terms vanish except the first one, which is $-\det M_{12}$. ♠

The following equation implies the second reformulation.

$$\det(M_{11}) = \lambda_2 ... \lambda_n / n. \tag{6}$$

The proof of this equation follows the great argument given in

http://maths.qmul.ac.uk/∼pjc/odds/zero.pdf

but I'll also fill in some elementary details to make it more self-contained.

Lemma 0.5 *M has all real eigenvalues.*

Proof: Let $\overline{v}$ denote the coordinatewide complex conjuatage of v. Since M is symmetric, we have $Mv \cdot w = v \cdot Mw$ for all vectors $v, w \in \boldsymbol{C}^n$. Suppose (λ, v) is an eigenvalue-eigenvector pair for M. That is, $Mv = \lambda v$. All we know in advance is that $\lambda \in \boldsymbol{C}$ and $v \in \boldsymbol{C}^n$. Note that $M\overline{v} = \overline{Mv} = \overline{\lambda}\overline{v}$. But then we have

$$\lambda v \cdot \overline{v} = Mv \cdot \overline{v} = v \cdot M\overline{v} = v \cdot \overline{\lambda}\overline{v} = \overline{\lambda} v \cdot \overline{v}.$$

But then $\lambda = \overline{\lambda}$. Hence $\lambda \in \boldsymbol{R}$. ♠

Lemma 0.6 *Real eigenvectors of M – i.e., those which are vectors in $\boldsymbol{R}^n$ – corresponding to distinct eigenvalues are orthogonal.*

Proof: Suppose that (λ, v) and (μ, w) are two eigenvalue-eigenvector pairs and $\lambda \neq \mu$. Then $\lambda(v \cdot w) = Mv \cdot w = v \cdot Mw = \mu(v \cdot w)$. Hence $v \cdot w = 0$. ♠

Lemma 0.7 *$\boldsymbol{R}^n$ has an orthogonal basis of eigenvectors of M.*

Proof: Since M has all real eigenvalues, $\boldsymbol{R}^n$ has a basis of eigenvectors of M. Within each eigenspace we can choose this basis to be orthogonal (by Graham-Schmidt). The previous result shows that the different eigenspaces are orthogonal. Hence $\boldsymbol{R}^n$ has an orthogonal basis of eigenvectors for M. ♠

Since the rows and columns of M sum to 0, the vector $(1, ..., 1)$ is an eigenvector for M and 0 is the corresponding eigenvector. The remaining eigenvectors lie in the subspace of $\boldsymbol{R}_0^n$ consisting of vectors whose coordinates sum to 0. Why? Because $\boldsymbol{R}_0^n$ is the orthogonal complement to $(1, ..., 1)$.

Now for the magic trick. Consider the matrix J whose every entry is 1. This time $(1, ..., 1)$ is an eigenvector of J corresponding to the eigenvalue n. Every vector in $\boldsymbol{R}_0^n$ is an eigenvector of J corresponding to eigenvalue 0. So, the same orthogonal basis for M works equally for J. But this means that the eigenvalues of $M + J$ are $n, \lambda_2, ..., \lambda_n$. Hence

$$\det(M + J) = n\lambda_2 ... \lambda_n.$$

On the other hand, consider the following column/row operations:

1. Add Rows 2,...,n to Row 1. This makes Row 1 equal to $(n, ..., n)$.
2. Add Cols 2,...,n to Col 1. This make Col 1 equal to $(n^2, n..., n)^t$.
3. Divide Col 1 by n. This makes Col 1 equal to $(n, 1, ..., 1)^t$.
4. Subtract Col 1 from Col2,...,Coln. This makes Row 1 equal to $(n, 0, ..., 0)$ and the $(1, 1)$ minor equal to M_{11}.

Operation 3 divides the determinant by n and the remaining operations do not change the determinant. Hence

$$\det(M + J) = n^2 \times \det M_{11}.$$

Combining the two equations gives Equation 6.