Chapter 13

Random Spanning Trees

13.1 Introduction

In this chapter we present one of the most fundamental results in Spectral Graph Theory: the Matrix-Three Theorem. It relates the number of spanning trees of a connected graph to the determinants of principal minors of the Laplacian. We then extend this result to relate the fraction of spanning trees that contain a given edge to the effective resistance of the entire graph between the edge's endpoints.

13.2 Determinants

To begin, we review some facts about determinants of matrices and characteristic polynomials. We first recall the Leibniz formula for the determinant of a square matrix $\boldsymbol{A}$:

$$\det(\boldsymbol{A}) = \sum_{\pi} \left(\mathsf{sgn}(\pi) \prod_{i=1}^{n} \boldsymbol{A}(i, \pi(i)) \right), \tag{13.1}$$

where the sum is over all permutations π of $\{1, \ldots, n\}$.

Also recall that the determinant is multiplicative, so for square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$

$$\det(\boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{A}) \det(\boldsymbol{B}). \tag{13.2}$$

Elementary row operations do not change the determinant. If the columns of $\boldsymbol{A}$ are the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, then for every c

$$\det\left(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\right) = \det\left(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n + c\boldsymbol{a}_1\right).$$

This fact gives us two ways of computing the determinant. The first comes from the fact that we can apply elementary row operations to transform $\boldsymbol{A}$ into an upper triangular matrix, and (13.1) tells us that the determinant of an upper triangular matrix is the product of its diagonal entries.

The second comes from the observation that the determinant is the volume of the parallelepiped with axes $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$: the polytope whose corners are the origin and $\sum_{i \in S} \boldsymbol{a}_i$ for every $S \subseteq \{1, \dots, n\}$. Let

$$\boldsymbol{\Pi}_{\boldsymbol{a}_1}$$

be the symmetric projection orthogonal to $\boldsymbol{a}_1$. As this projection amounts to subtracting off a multiple of $\boldsymbol{a}_1$ and elementary row operations do not change the determinant,

$$\det\left(\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n\right) = \det\left(\boldsymbol{a}_1, \boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_2, \dots, \boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_n\right).$$

The volume of this parallelepiped is $\|\boldsymbol{a}_1\|$ times the volume of the parallelepiped formed by the vectors $\boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_2, \dots, \boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_n$. I would like to write this as a determinant, but must first deal with the fact that these are $n-1$ vectors in an n dimensional space. The way we first learn to handle this is to project them into an $n-1$ dimensional space where we can take the determinant. Instead, we will employ other elementary symmetric functions of the eigenvalues.

13.3 Characteristic Polynomials

Recall that the characteristic polynomial of a matrix $\boldsymbol{A}$ is

$$\det(x\boldsymbol{I} - \boldsymbol{A}).$$

I will write this as

$$\sum_{k=0}^{n} x^{n-k}(-1)^k \sigma_k(\boldsymbol{A}),$$

where $\sigma_k(\boldsymbol{A})$ is the kth elementary symmetric function of the eigenvalues of $\boldsymbol{A}$, counted with algebraic multiplicity:

$$\sigma_k(\boldsymbol{A}) = \sum_{|S|=k} \prod_{i \in S} \lambda_i.$$

Thus, $\sigma_1(\boldsymbol{A})$ is the trace and $\sigma_n(\boldsymbol{A})$ is the determinant. From this formula, we know that these functions are invariant under similarity transformations.

In Exercise 3 from Lecture 2, you were asked to prove that

$$\sigma_k(\boldsymbol{A}) = \sum_{|S|=k} \det(\boldsymbol{A}(S, S)). \tag{13.3}$$

This follows from applying the Leibnitz formula (13.1) to $\det(x\boldsymbol{I} - \boldsymbol{A})$.

If we return to the vectors $\boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_2, \dots, \boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_n$ from the previous section, we see that the volume of their parallelepiped may be written

$$\sigma_{n-1}\left(\boldsymbol{0}_n, \boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_2, \dots, \boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_n\right),$$

as this will be the product of the $n-1$ nonzero eigenvalues of this matrix.

Recall that the matrices $\boldsymbol{B}\boldsymbol{B}^T$ and $\boldsymbol{B}^T\boldsymbol{B}$ have the same eigenvalues, up to some zero eigenvalues if they are rectangular. So,

$$\sigma_k(\boldsymbol{B}\boldsymbol{B}^T) = \sigma_k(\boldsymbol{B}^T\boldsymbol{B}).$$

This gives us one other way of computing the absolute value of the product of the nonzero eigenvalues of the matrix

$$\left(\boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_2, \dots, \boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_n\right).$$

We can instead compute their square by computing the determinant of the square matrix

$$\begin{pmatrix} \boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_n \end{pmatrix} \left(\boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_2, \dots, \boldsymbol{\Pi}_{\boldsymbol{a}_1}\boldsymbol{a}_n\right).$$

When $\boldsymbol{B}$ is a singular matrix of rank k, $\sigma_k(\boldsymbol{B})$ acts as the determinant of $\boldsymbol{B}$ restricted to its span. Thus, there are situations in which σ_k is multiplicative. For example, if $\boldsymbol{A}$ and $\boldsymbol{B}$ both have rank k and the range of $\boldsymbol{A}$ is orthogonal to the nullspace of $\boldsymbol{B}$, then

$$\sigma_k(\boldsymbol{B}\boldsymbol{A}) = \sigma_k(\boldsymbol{B})\sigma_k(\boldsymbol{A}). \tag{13.4}$$

We will use this identity in the case that $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric and have the same nullspace.

13.4 The Matrix Tree Theorem

We will state a slight variant of the standard Matrix-Tree Theorem. Recall that a spanning tree of a graph is a subgraph that is a tree.

Theorem 13.4.1. *Let $G = (V, E, w)$ be a connected, weighted graph. Then*

$$\sigma_{n-1}(\boldsymbol{L}_G) = n \sum_{\text{spanning trees } T} \prod_{e \in T} w_e.$$

Thus, the eigenvalues allow us to count the sum over spanning trees of the product of the weights of edges in those trees. When all the edge weights are 1, we just count the number of spanning trees in G.

We first prove this in the case that G is just a tree.

Lemma 13.4.2. *Let $G = (V, E, w)$ be a weighted tree. Then,*

$$\sigma_{n-1}(\boldsymbol{L}_G) = n \prod_{e \in E} w_e.$$

Proof. For $a \in V$, let $S_a = V - \{a\}$. We know from (13.3)

$$\sigma_{n-1}(\boldsymbol{L}_G) = \sum_{a \in V} \det(\boldsymbol{L}_G(S_a, S_a)).$$

We will prove that for every $a \in V$,

$$\det(\boldsymbol{L}_G(S_a, S_a)) = \prod_{e \in E} w_e.$$

Write $\boldsymbol{L}_G = \boldsymbol{U}^T \boldsymbol{W} \boldsymbol{U}$, where $\boldsymbol{U}$ is the signed edge-vertex adjacency matrix and $\boldsymbol{W}$ is the diagonal matrix of edge weights. Write $\boldsymbol{B} = \boldsymbol{W}^{1/2} \boldsymbol{U}$, so

$$\boldsymbol{L}_G(S_a, S_a) = \boldsymbol{B}(:, S_a)^T \boldsymbol{B}(:, S_a),$$

and

$$\det(\boldsymbol{L}_G(S_a, S_a)) = \det(\boldsymbol{B}(:, S_a))^2,$$

where we note that $\boldsymbol{B}(:, S_a)$ is square because a tree has $n-1$ edges and so $\boldsymbol{B}$ has $n-1$ rows.

To see what is going on, first consider the case in which G is a weighted path and a is the first vertex. Then,

$$\boldsymbol{U} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{B}(:, S_1) = \begin{pmatrix} -\sqrt{w_1} & 0 & \cdots & 0 \\ \sqrt{w_2} & -\sqrt{w_2} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & -\sqrt{w_{n-1}} \end{pmatrix}.$$

We see that $\boldsymbol{B}(:, S_1)$ is a lower-triangular matrix, and thus its determinant is the product of its diagonal entries, $-\sqrt{w_i}$.

To see that the same happens for every tree, renumber the vertices (permute the columns) so that a comes first, and that the other vertices are ordered by increasing distance from 1, breaking ties arbitrarily. This permutations can change the sign of the determinant, but we do not care because we are going to square it. For every vertex $c \neq 1$, the tree now has exactly one edge (b, c) with $b < c$. Put such an edge in position $c-1$ in the ordering, and let w_c indicate its weight. Now, when we remove the first column to form $\boldsymbol{B}(:, S_1)$, we produce a lower triangular matrix with the entry $-\sqrt{w_c}$ on the cth diagonal. So, its determinant is the product of these terms and

$$\det(\boldsymbol{B}(:, S_a))^2 = \prod_{c=2}^{n} w_c.$$

□

Proof of Theorem 13.4.1 . As in the previous lemma, let $\boldsymbol{L}_G = \boldsymbol{U}^T \boldsymbol{W} \boldsymbol{U}$ and $\boldsymbol{B} = \boldsymbol{W}^{1/2} \boldsymbol{U}$. So,

$$\begin{aligned} \sigma_{n-1}(\boldsymbol{L}_G) &= \sigma_{n-1}(\boldsymbol{B}^T \boldsymbol{B}) \\ &= \sigma_{n-1}(\boldsymbol{B} \boldsymbol{B}^T) \\ &= \sum_{|S|=n-1, S \subseteq E} \sigma_{n-1}(\boldsymbol{B}(S,:) \boldsymbol{B}(S,:)^T) \quad \text{(by (13.3))} \\ &= \sum_{|S|=n-1, S \subseteq E} \sigma_{n-1}(\boldsymbol{B}(S,:)^T \boldsymbol{B}(S,:)) \\ &= \sum_{|S|=n-1, S \subseteq E} \sigma_{n-1}(\boldsymbol{L}_{G_S}), \end{aligned}$$

where by G_S we mean the graph containing just the edges in S. As S contains $n-1$ edges, this graph is either disconnected or a tree. If it is disconnected, then its Laplacian has at least two zero eigenvalues and $\sigma_{n-1}(\boldsymbol{L}_{G_S}) = 0$. If it is a tree, we apply the previous lemma. Thus, the sum equals

$$\sum_{\text{spanning trees } T \subseteq E} \sigma_{n-1}(\boldsymbol{L}_{G_T}) = n \sum_{\text{spanning trees } T} \prod_{e \in T} w_e.$$

□

13.5 Leverage Scores and Marginal Probabilities

The *leverage score* of an edge, written ℓ_e is defined to be $w_e \mathrm{R_{eff}}(e)$. That is, the weight of the edge times the effective resistance between its endpoints. The leverage score serves as a measure of how important the edge is. For example, if removing an edge disconnects the graph, then $\mathrm{R_{eff}}(e) = 1/w_e$, as all current flowing between its endpoints must use the edge itself, and $\ell_e = 1$.

Consider sampling a random spanning tree with probability proportional to the product of the weights of its edges. We will now show that the probability that edge e appears in the tree is exactly its leverage score.

Theorem 13.5.1. *If we choose a spanning tree T with probability proportional to the product of its edge weights, then for every edge e*

$$Pr\left[e \in T\right] = \ell_e.$$

For simplicity, you might want to begin by thinking about the case where all edges have weight 1.

Recall that the effective resistance of edge $e = (a, b)$ is

$$(\delta_a - \delta_b)^T \boldsymbol{L}_G^+ (\delta_a - \delta_b),$$

and so

$$\ell_{a,b} = w_{a,b} (\delta_a - \delta_b)^T \boldsymbol{L}_G^+ (\delta_a - \delta_b).$$

We can write a matrix $\boldsymbol{\Gamma}$ that has all these terms on its diagonal by letting $\boldsymbol{U}$ be the edge-vertex adjacency matrix, $\boldsymbol{W}$ be the diagonal edge weight matrix, $\boldsymbol{B} = \boldsymbol{W}^{1/2} \boldsymbol{U}$, and setting

$$\boldsymbol{\Gamma} = \boldsymbol{B} \boldsymbol{L}_G^+ \boldsymbol{B}^T.$$

The rows and columns of $\boldsymbol{\Gamma}$ are indexed by edges, and for each edge e,

$$\boldsymbol{\Gamma}(e, e) = \ell_e.$$

For off-diagonal entries corresponding to edges (a, b) and (c, d), we have

$$\boldsymbol{\Gamma}((a, b), (c, d)) = \sqrt{w_{a,b}} \sqrt{w_{c,d}} (\boldsymbol{\delta}_a - \boldsymbol{\delta}_b)^T \boldsymbol{L}_G^+ (\boldsymbol{\delta}_c - \boldsymbol{\delta}_d).$$

Claim 13.5.2. *The matrix $\boldsymbol{\Gamma}$ is a symmetric projection matrix and has trace $n-1$.*