Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

In this lecture we prove the Burton,Pemantle Theorem [BP93].

3.1 Properties of Matrix Trace

In this part we record the properties of the trace that we talked about in the last lecture. For a matrix $A \in \mathbb{R}^{n \times n}$,

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} A(i,i).$$

It is easy to see that for any pair of matrices $A \in \mathbb{R}^{n \times k}$ and $B \in \mathbb{R}^{k \times n}$,

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA).$$

The matrix dot product is defined analogous to the vector dot product. For matrices $A, B \in \mathbb{R}^{k \times n}$,

$$A \bullet B := \sum_{i=1}^{k} \sum_{j=1}^{n} A(i,j)B(i,j).$$

It is easy to see that

$$A \bullet B = \mathrm{Tr}(AB^{\mathsf{T}}).$$

Lemma 3.1. *For any symmetric matrix A with eigenvalues $\lambda_1, \ldots, \lambda_n$,*

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i.$$

Proof. Let $x_1, \ldots, x_n$ be orthonormal eigenvectors corresponding to $\lambda_1, \ldots, \lambda_n$. Also, let $\mathbf{1}_i$ be the indicator vector of the i-th coordinate. Then, by definition of the trace,

$$
\begin{aligned}
\mathrm{Tr}(A) &= \sum_{i=1}^{n} \mathbf{1}_i A \mathbf{1}_i \\
&= \sum_{i=1}^{n} \mathbf{1}_i \left(\sum_{j=1}^{n} \lambda_j x_j x_j^{\mathsf{T}} \right) \mathbf{1}_i \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \langle x_j, \mathbf{1}_i \rangle^2
\end{aligned}
$$

Now, we swap the sums and we use that each x_j has norm 1, so $\sum_{i=1}^{n} \langle x_j, \mathbf{1}_i \rangle^2 = 1$.

$$
\begin{aligned}
\mathrm{Tr}(A) &= \sum_{j=1}^{n} \lambda_j \sum_{i=1}^{n} \langle x_j, \mathbf{1}_i \rangle^2 \\
&= \sum_{j=1}^{n} \lambda_j.
\end{aligned}
$$

$\square$

3.2   Characteristic Polynomial

For an indeterminant t, the *characteristic polynomial* of A, $\chi[a](t)$ is defined as follows,

$$
\chi[A](t) = \det(tI - A).
$$

Note that the characteristic polynomial is of degree n with respect to t. The roots of this polynomial are the eigenvalues of A. This is because if λ is an eigenvalue of A, then $(\lambda I - A)x = 0$. So, $\lambda I - A$ is singular, and $\det(\lambda I - A) = 0$, i.e., λ is a root of $\det(tI - A)$.

Recall that if A is symmetric then all eigenvalues of A are real. Therefore, if A is symmetric with eigenvalues $\lambda_1, \ldots, \lambda_n$, the characteristic polynomial is a real-rooted polynomial with roots $\lambda_1, \ldots, \lambda_n$ and we can write,

$$
\chi[A](t) = (t - \lambda_1)(t - \lambda_2) \ldots (t - \lambda_n)
$$

Note that by definition the characteristic polynomial is invariant under rotation. That is, it only captures the eigenvalues of A; so any matrix with the same eigenvalues as A have the same characteristic polynomial. Opening up the above identity we can write,

$$
\begin{aligned}
\chi[A](t) &= t^n - (\lambda_1 + \cdots + \lambda_n)t^{n-1} + (\lambda_1\lambda_2 + \lambda_2\lambda_3 + \cdots + \lambda_{n-1}\lambda_n)t^{n-2} + \ldots \\
&= \sum_{k=0}^{n} t^{n-k}(-1)^k \sum_{S \in \binom{[n]}{k}} \prod_{i \in S} \lambda_i.
\end{aligned}
\tag{3.1}
$$

For a set $S \subseteq [n]$, we write A_S to denote the sub-matrix of A indexed by the elements of S. A_S is also known as the principal sub-matrix of A. We use $\det_k(A)$ to denote the sum of all principal minors of A of size k, i.e.,

$$
\det_k(A) = \sum_{S \in \binom{[n]}{k}} \det(A_S).
$$

It is easy to see that the coefficient of t^{n-k} in the characteristic polynomial is $(-1)^k \det_k(A)$. Therefore, we can write

$$
\chi[A](t) = \sum_{k=0^n} t^{n-k}(-1)^k \det_k(A).
\tag{3.2}
$$

Therefore, by (3.1), for any $k \geq 0$, we can write

$$
\det_k(A) = \sum_{S \in \binom{[n]}{k}} \prod_{i \in S} \lambda_i.
$$

In particular, for $k = 1$ we get that

$$\text{Tr}(A) = \det_1(A) = \sum_{i=1}^{n} \lambda_i.$$

This gives another proof of Lemma 3.1.

The following elementary statement follows from the definition of the characteristic polynomial and the Cauchy-Binet formula.

Theorem 3.2 ([MSS13])**.** *For rank 1 symmetric matrices* $W_1, \ldots, W_m \in \mathbb{R}^{n \times n}$, *and scalars* $z_1, \ldots, z_m$,

$$\det \left(tI + \sum_{i=1}^{m} z_i W_i \right) = \sum_{k=0}^{n} t^{n-k} \sum_{S \in \binom{[m]}{k}} z^S \det_k \left(\sum_{i \in S} W_i \right).$$

We leave the proof of this as an exercise.

3.3   Proof of the Burton, Pemantle Theorem

Recall that for any edge e we defined

$$y_e = L_G^{\dagger/2} b_e,$$

if G is weighted we defined

$$y_e = \sqrt{w(e)} L_G^{\dagger/2} b_e.$$

In this section we prove Theorem 3.3.

Theorem 3.3 (Burton and Pemantle [BP93])**.** *Given any (weighted) graph* G, *let* $Y \in \mathbb{R}^{E \times E}$ *where for each pair of edges* $e, f \in F$,

$$Y(e, f) = \langle y_e, y_f \rangle.$$

Let μ *be the weighted uniform distribution of spanning trees,* $\mu(T) = \prod_{e \in T} w(e)$. *Then, for any set of edges* $F \subseteq E$,

$$\mathbb{P}_{T \sim \mu} [F \subseteq T] = \det(Y_F).$$

Proof of Theorem 3.3. We prove this by induction on the size of F. Say the claim is true for any set of size less than $|F|$. By the inclusion-exclusion principle we can write,

$$
\begin{aligned}
\mathbb{P}[F \cap T \neq \emptyset] &= \sum_{k=1}^{|F|} (-1)^{k-1} \sum_{S \in \binom{F}{k}} \mathbb{P}[S \subseteq T] \\
&= \sum_{k=1}^{|F|-1} (-1)^{k-1} \sum_{S \in \binom{F}{k}} \det(Y_S) + (-1)^{|F|-1} \mathbb{P}[F \subseteq T] \quad\quad (3.3)
\end{aligned}
$$

By Cauchy-Binet, for any set $S \subseteq F$,

$$\det(Y_S) = \det_{|S|} \left(\sum_{e \in S} y_e y_e^{\mathsf{T}} \right).$$

So,

$$
\begin{aligned}
\mathbb{P}\left[F \cap T \neq \emptyset\right] &= \sum_{k=1}^{|F|-1} (-1)^{k-1} \sum_{S \in \binom{F}{k}} \det_k \left(\sum_{e \in S} y_e y_e^{\mathsf{T}}\right) + (-1)^{|F|-1}\mathbb{P}\left[F \subseteq T\right] \\
&= \sum_{k=1}^{|F|-1} (-1)^{k-1} \det_k \left(\sum_{e \in F} y_e y_e^{\mathsf{T}}\right) + (-1)^{|F|-1}\mathbb{P}\left[F \subseteq T\right]. \quad (3.4)
\end{aligned}
$$

Note that the coefficients in the RHS are just the coefficients of the characteristic polynomial of the matrix $\sum_{e \in F} y_e y_e^{\mathsf{T}}$. By (3.2),

$$
\begin{aligned}
\det_{|F|}\left(I - \sum_{e \in F} y_e y_e^{\mathsf{T}}\right) &= 1 - \sum_{k=1}^{|F|}(-1)^{k-1}\det_k\left(\sum_{e \in F} y_e y_e^{\mathsf{T}}\right) \\
&= 1 - \mathbb{P}\left[F \cap T \neq \emptyset\right] + (-1)^{|F|+1}\mathbb{P}\left[F \subseteq T\right] + (-1)^{|F|}\det_{|F|}\left(\sum_{e \in F} y_e y_e^{\mathsf{T}}\right).
\end{aligned}
$$

On the other hand, we can write

$$
\begin{aligned}
1 - \mathbb{P}\left[F \cap T \neq \emptyset\right] &= \mathbb{P}\left[F \cap T = \emptyset\right] \\
&= \frac{\det(\tilde{L}_G - \sum_{e \in F} b_e b_e^{\mathsf{T}})}{\det(\tilde{L}_G)} \\
&= \det\left(I - \sum_{e \in F} y_e y_e^{\mathsf{T}}\right).
\end{aligned}
$$

Therefore,

$$
\mathbb{P}\left[F \subseteq T\right] = \det_{|F|}\left(\sum_{e \in F} y_e y_e^{\mathsf{T}}\right) = \det(Y_F).
$$

$\square$

In words, the Burton, Pemantle theorem implies that the probability that a set $S \subseteq E$ is in a random spanning tree is equal to the square of the volume of the $|S|$-dimensional parallelepiped defined on the set of points $\{y_e\}_{e \in S}$. Having this characterization characterization we can write a geometric variant of the algorithm for sampling random spanning trees that we presented in lecture 2 (see Algorithm 1).

Note that Algorithm 1 is purely algebraic and it has nothing to do with the graph G. Indeed the same algorithm can be applied to sample a basis from a weighted uniform distribution of bases of a given set of isotropic vectors in $\mathbb{R}^d$ where the probability of each basis is proportional to the square of the volume of the d-dimensional parallelepiped defined on the vectors of that basis.

3.4 Negative Correlation & Concentration

Recall that in the last lecture we used the Burton, Pemantle theorem to show that edges are negatively correlated in a random spanning tree distribution. That is, for any set of edges F,

$$
\mathbb{P}_{T \sim \mu}\left[F \subseteq T\right] \leq \prod_{e \in F} \mathbb{P}\left[e \in T\right].
$$

Algorithm 1 An equivalent geometric translation of Random Spanning Tree algorithm in Lecture 2.

Let $y_e = L_G^{\dagger/2} b_e$ for each $e \in E$. and $T = \{\}$.

Choose an arbitrary ordering of the edges, $e_1, \ldots, e_m$.

for $i = 1 \to m$ **do**

With probability $\|y_{e_i}\|^2$ add e_i to T, and for every $j > i$ let

$$y_{e_j} = y_{e_j} - \frac{\langle y_{e_j}, y_{e_i} \rangle y_{e_i}}{\|y_{e_i}\|^2}.$$

With the remaining probability do not add e_i and for every $j > i$ let

$$y_{e_j} = \frac{y_{e_j}}{\sqrt{1 - \|y_{e_i}\|^2}}.$$

end for

It is also not hard to see that

$$\mathbb{P}_{T \sim \mu}[|F \cap T| = 0] \geq \prod_{e \in F}(1 - \mathbb{P}[e \in F]).$$

Negative correlation is generally a nontrivial property of probability distributions. It is a widely open problem to find families of matroids that satisfy negative correlation property. For example, the following conjecture is proposed independent by Pemantle [Pem00], Grimmett and Winkler [GW04] on the negative correlation property on uniform forests of a graph.

Conjecture 3.1. *Let μ be a uniform distribution on all spanning forests of G, i.e., all forests that include all vertices of G. Then, for any two edges e, f in G,*

$$\mathbb{P}_{F \sim \mu}[e, f \in F] \leq \mathbb{P}[e \in F] \cdot \mathbb{P}[f \in F].$$

See [Goe+15] for an application of this conjecture in credit networks.

It was shown by Panconesi and Srinivasan [PS97] that a direct implication of negative correlation is the concentration of linear functions of the edges.

Theorem 3.4 (Panconesi and Srinivasan [PS97])**.** *For graph G a (weighted) uniform spanning tree distribution μ, and any set $F \subseteq E$,*

$$\mathbb{P}_{T \sim \mu}[|F \cap T| \geq (1 + \delta)\mathbb{E}[|F \cap T|]] \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}}\right)^{\mathbb{E}[|F \cap T|]}$$

Proof. The basic idea of the proof is simple. Let $X_e := \mathbb{I}[e \in T]$ be the indicator random variable of e being in the tree and let $\tilde{X}_e$ be a set of independent Bernoulli random variables one for each $e \in F$ each with success probability $\mathbb{E}[\tilde{X}_e] = \mathbb{E}[X_e]$. Also, let $X := \sum_{e \in F} X_e$ and $\tilde{X} := \sum_{e \in F} \tilde{X}_e$. It follows by the negative correlation that for any $k \geq 0$,

$$\mathbb{E}[X^k] \leq \mathbb{E}[\tilde{X}^k]. \tag{3.5}$$

This is because we can write

$$
\begin{aligned}
\mathbb{E}\left[X^k\right] &= \sum_{S\in\binom{F}{k}}\mathbb{E}\left[\prod_{e\in S}X_e\right] \\
&\le \sum_{S\in\binom{F}{k}}\prod_{e\in S}\mathbb{E}\left[X_e\right] \\
&= \sum_{S\in\binom{F}{k}}\mathbb{E}\left[\prod_{e\in S}\tilde{X}_e\right] = \mathbb{E}\left[\tilde{X}^k\right].
\end{aligned}
$$

The inequality follows by the negative correlation property.

It follows from (3.5) that for any $t \ge 0$,

$$
\mathbb{E}\left[e^{tX}\right] \le \mathbb{E}\left[e^{t\tilde{X}}\right].
$$

Now, we can write,

$$
\begin{aligned}
\mathbb{P}_{T\sim\mu}\left[|F\cap T|\ge(1+\delta)\mathbb{E}\left[|F\cap T|\right]\right] &\le \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mathbb{E}[X]}} \\
&\le \frac{\mathbb{E}\left[e^{t\tilde{X}}\right]}{e^{t(1+\delta)\mathbb{E}[\tilde{X}]}}
\end{aligned}
$$

The rest of the proof is similar to the usual Chernoff bound arguments. □

Note that the above theorem is fairly general and it holds for a large families of distributions. It turns out that random spanning tree distributions satisfy many virtues of negative dependence so we can prove much stronger concentration inequality. For example, we can show that the number of vertices of degree 2 in a random spanning tree is highly concentrated around its expected value.

References

[BP93]   R. Burton and R. Pemantle. "Local Characteristics, Entropy and Limit Theorems for Spanning Trees and Domino Tilings via Transfer-Impedances". In: *Ann. Probab.* 21.math.PR/0404048 (1993), pp. 1329–1371 (cit. on pp. 3-1, 3-3).

[Goe+15] A. Goel, S. Khanna, S. Raghvendra, and H. Zhang. "Connectivity in Random Forests and Credit Networks". In: *SODA*. 2015, pp. 2037–2048 (cit. on p. 3-5).

[GW04]   G. R. Grimmett and S. N. Winkler. "Negative association in uniform forests and connected graphs". In: *Random Structures and Algorithms*. 2004, pp. 444–460 (cit. on p. 3-5).

[MSS13]   A. Marcus, D. A. Spielman, and N. Srivastava. "Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem". 2013. URL: http://arxiv.org/abs/1306.3969 (cit. on p. 3-3).

[Pem00]   R. Pemantle. "Towards a theory of negative independence". In: *Journal of Mathematical Physics* 41 (2000), pp. 1371–1390 (cit. on p. 3-5).

[PS97]   A. Panconesi and A. Srinivasan. "Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds". In: *SIAM J. Comput.* 26 (1997), pp. 350–368 (cit. on p. 3-5).