SPANNING TREES

STEVEN P. LALLEY

1. Preliminaries

1.1. **Spanning Trees.** Let $G = (\mathscr{X}, \mathscr{E})$ be a finite, connected graph with no multiple edges or self-loops, and let $N = |\mathscr{X}|$ be the number of vertices. A *spanning tree* is a connected subgraph graph $T = (\mathscr{X}, \mathscr{E}_T)$ with no cycles. Because a spanning tree is connected, any two vertices $x, y \in \mathscr{X}$ are connected by a path in T; and because T has no cycles, this path is unique.

Any spanning tree T can be assigned a *root* vertex ϱ in N different ways; the pair (T, ϱ) is a *rooted spanning tree*. The edges of a rooted spanning tree (T, ϱ) can be *directed* toward the root, so that for every vertex $x \neq \varrho$ there is a unique edge leading *out* of x. (This is the first edge on the unique path from x to ϱ.) Since every edge of T has a direction, it follows that there must be exactly $N-1$ edges in T, one for every vertex other than the root.

1.2. **Simple Random Walk on a Graph.** The *simple random walk* on the graph $G = (\mathscr{X}, \mathscr{E})$ is the reversible Markov chain on the vertex set $\mathscr{X}$ associated with the conductance function $C_e \equiv 1$, that is, the Markov chain with transition probabilities

$$\begin{aligned} p(x,y) &= 1/d(x) \quad \text{if } x, y \text{ are neighbors;} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Here $d(x)$ is the *degree* of x in the graph, that is, the number of neighbors of x (which is also the number of edges incident to x). Becasue the graph is connected, the simple random walk is irreducible. Its unique stationary distribution ν is proportional to the degree function:

$$\nu(x) = d(x)/D \quad \text{where} \quad D = \sum_{y \in \mathscr{X}} d(y).$$

Since simple random walk is an irreducible Markov chain, there is a two-sided version $(X_n)_{n\in\mathbb{Z}}$ such that for each n, the distribution of X_n is the stationary distribution ν. (See HW 8, Problem 4.)

Proposition 1. *The simple random walk is invariant by time-reversal. Explicitly, the joint distribution of the sequence $(X_n)_{n\in\mathbb{Z}}$ is the same as that of $(X_{-n})_{n\in\mathbb{Z}}$.*

Proof. (Sketch) It suffices to show that the cylinder sets

$$\{X_i = x_i \ \forall\, 0 \le i \le n\} \quad \text{and} \quad \{X_{-i} = x_i \ \forall\, 0 \le i \le n\}.$$

This is a routine calculation (exercise) using the fact that the stationary distribution ν is proportional to the degree function $d(x)$. □

2. Aldous-Hoover Theorem

Aldous and Hoover found a simple Markov chain on the space of rooted spanning trees of G whose stationary distribution is the product of the uniform distribution on unrooted trees with the stationary distribution ν on the root. This leads to a relatively simple randomized algorithm for constructing a uniformly distributed spanning tree, as will be explained below. The Markov chain lives on the set of rooted spanning trees (T, x), where T is an unrooted spanning tree and $x \in \mathscr{X}$ is a root vertex; it evolves according to the following rules:

Transition Law: The one-step transition law $(T, x) \mapsto (T', y)$ can occur if and only if

(i) y and x are nearest neighbors, and
(ii) T' is obtained from T by *adding* the (directed) edge from x to y and *deleting* the direced edge *out* of y.

This transition occurs with probability

$$q((T, x), (T', y)) = 1/d(x).$$

Lemma 2. *The transition probability matrix $\mathbb{Q} = (q((T, x), (T', y)))$ is irreducible, and so there is a unique stationary distribution π.*

Proof. (Sketch) It suffices to show that for any two rooted trees (T, x) and (T', y) there is a finite sequence of allowable transitions that take (T, x) to (T', y). This can be done in two stages.

(1) First, take any path in the graph G from x to y and follow the sequence of transitions dictated by this sequence; this moves (T, x) to a rooted tree (T'', y) with root y.
(2) Next, order the "branches" of the target tree T' emanating from the root y; call these $B_1, B_2, \cdots, B_k$. Let γ be the path in G from y back to y that traverses the branches B_i one at a time, by depth-first search. Follow the sequence of transitions dictated by the path γ. This will transform (T'', y) to (T', y).

$\square$

Lemma 2 implies that the Markov chain has a unique stationary distribution π. Consequently (by HW 8, Problem 4) there is a two-sided version $((T_n, X_n))_{n\in\mathbb{Z}}$ such that for any time n the distribution of (T_n, X_n) is π.

Proposition 3. *The two-sided sequence $(X_n)_{n\in\mathbb{Z}}$ is the stationary simple random walk on G.*

Proof. It suffices to show that the sequence $(X_n)_{n\in\mathbb{Z}}$ is a simple random walk, because it will then follow (HW 8 again) that it is stationary. But this is obvious, because by construction, whenever the root X_n is x, it moves to a nearest neighbor y with (conditional) probability $1/(d(x))$. $\square$

Proposition 4. *The tree T_0 is uniquely determined by the past $(X_n)_{n\le 0}$ by the following law: for each vertex $y \neq X_0$, the unique edge of the rooted tree (T_0, X_0) out of y is the edge crossed by the random walk X_n on the step after the* last *visit to y before time* 0.

Proof. Whenever the random walk X_n visits the vertex y, this vertex becomes the root of the tree, and therefore has no edge out. On the next step, the root moves to a nearest neighbor z of y, and the directed edge from y to z is added to the tree. This edge remains until the next visit of the random walk X_n to the vertex y. Thus, between successive visits to y, the last exit from y determines the directed edge out of y in the rooted spanning tree (T_n, X_n). □

Theorem 5. *(Aldous-Hoover) The unique stationary distribution is* $\pi(T,x) = C\,d(x)\prod_y (1/d(y))$. *Consequently, the distribution of the* unrooted *tree* T_0 *is uniform on the space of all spanning trees.*

Proof. It suffices to show that the function $\mu(T,x) = d(x)\prod_y (1/d(y))$ satisfies the defining equations of a stationary distribution, that is,

$$\mu(T,x) = \sum_{(T',z)} \mu(T',z) q((T',z),(T,x)).$$

The rooted trees $(T'z)$ for which the transition probability $q((T',z),(T,x))$ is positive are those for which x is a nearest neighbor of z, and such that T is obtained from T' by adding the edge from z to x and deleting the edge out of x in the rooted tree (T',z). There are precisely $d(x)$ such pairs, one for each nearest neighbor z. Consequently, the equations for stationarity can be rewritten as

$$d(x)\prod_y (1/d(y)) \stackrel{?}{=} \sum_{z\sim x} (d(z)\prod_y (1/d(y)))(1/d(z)).$$

This equation, on second look, is completely obvious. □

Aldous-Hoover Algorithm: To determine the spanning tree T_0, one must, by Proposition 4, follow the path of the random walk X_n backward in time to discover, for each vertex $y \neq X_0$, which edge was crossed upon leaving y for the last time before time 0. Because the simple random walk is time-reversible, it is possible to do this by (i) choosing a root vertex $X_0 = x$ from the stationary distribution ν, and then (ii) running a random walk X_{-n} backward in time until the first n (i.e., last $-n$) when every vertex y has been visited at least once.

3. Wilson's Algorithm

4. The Matrix-Tree Theorem