MIT/LCS/TM-249

THE MARKOV CHAIN TREE THEOREM

F.T. Leighton

R.L. Rivest

November 1983

The Markov Chain Tree Theorem

F. Thomson Leighton

Mathematics Department and
Laboratory for Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Ronald L. Rivest

Laboratory for Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract: Let M be a finite first-order stationary Markov chain. We define an *arborescence* to be a set of edges in the directed graph for M having at most one edge out of every vertex, no cycles, and maximum cardinality. The weight of an arborescence is defined to be the product over each edge in the arborescence of the probability of the transition associated with the edge. We prove that if M starts in state i, its limiting average probability of being in state j is proportional to the sum of the weights of all arborescences having a path from i to j and no edge out of j. We present two proofs. The first is derived from simple graph theoretic identities. The second is derived from the closely-related Matrix Tree Theorem.

Key Words: arborescence, long-run probability, Markov chain, Matrix Tree Theorem

---

This research was supported by the Bantrell Foundation and by NSF grant MCS-8006938.

1. Introduction

Let M denote a finite first-order stationary Markov chain with states $\{1,\ldots,n\}$ and transition probability matrix P such that the ij entry p_{ij} of P is the probability of a transition from state i to state j. Let

$$\overline{P} = \lim_{t\to\infty} \frac{1}{t}(I + P + P^2 + \cdots + P^{t-1})$$

be the *long-run transition matrix* for the Markov chain M. (The limit exists because P is stochastic [5, Theorem 2.1].) The ijth entry of $\overline{P}$, $\overline{p}_{ij}$, is simply the long-run average probability that M will be in state j, given that M started in state i. If M is ergodic, then every row of $\overline{P}$ is the stationary probability vector π for M.

Let $V = \{1,\ldots,n\}$ be the vertices of a directed graph G, with edge set $E = \{(i,j) \mid p_{ij} \neq 0\}$. This is the usual directed graph associated with the Markov chain M. (Note that G may contain self-loops.) We define the *weight* of edge (i,j) to be p_{ij}. An edge set $A \subseteq E$ is an *arborescence* if A contains at most one edge out of every vertex, has no cycles, and has maximum possible cardinality. The *weight* of an arborescence is the product of the weights of the edges it contains. A node which has outdegree zero in A is called a *root* of the arborescence.

Clearly every arborescence contains the same number of edges. In fact, if G contains exactly k minimal closed subsets of vertices, then every arborescence has size $|V| - k$ and contains one root in each minimal closed subset. (A subset of nodes is said to be *closed* if no edges are directed out of the subset.) In particular, if G is strongly connected (i.e. the Markov chain is irreducible), then every arborescence is a set of $|V|-1$ edges that form a directed spanning tree with all edges flowing towards a single vertex (the root of the tree).

Let $\mathcal{A}(V)$ denote the set of arborescences of G, $\mathcal{A}_j(V)$ denote the set of arborescences having j as a root, and $\mathcal{A}_{ij}(V)$ denote the set of arborescences having root j and a directed path from i to j. (In the special case $i = j$, we define $\mathcal{A}_{jj}(V)$ to be $\mathcal{A}_j(V)$.) In addition, let $||\mathcal{A}(V)||$, $||\mathcal{A}_j(V)||$ and $||\mathcal{A}_{ij}(V)||$ denote the sums of the weights of the arborescences in $\mathcal{A}(V)$, $\mathcal{A}_j(V)$ and $\mathcal{A}_{ij}(V)$, respectively.

The standard method of computing $\overline{P}$ is to use linear algebraic techniques. In this paper, we derive a simple combinatorial technique for computing $\overline{P}$. In particular, we prove the following.

Theorem 1 (The Markov Chain Tree Theorem): *Let the stochastic $n \times n$ matrix P define a finite Markov chain M with long-run transition matrix $\overline{P}$. Then*

$$\overline{p}_{ij} = \frac{||\mathcal{A}_{ij}(V)||}{||\mathcal{A}(V)||}.$$

Corollary 1: *If M is ergodic, then*

$$\overline{p}_{ij} = \frac{||\mathcal{A}_j(V)||}{||\mathcal{A}(V)||}.$$

Proof of Corollary: When M is ergodic, the underlying graph is strongly connected and every arborescence with root j is a spanning tree with root j. ∎

Although Corollary 1 is proved in [7], the proof is complicated and the result is not well known. In this paper, we give two proofs of the stronger result. The first proof is derived from some simple graph theoretic identities involving arborescences and paths. The second proof is derived from the closely related Matrix-Tree Theorem. The proofs are presented in Section 3. The graph identities are included in Section 2. We conclude this section with some examples of how the Markov Chain Tree Theorem can be used to calculate long-run transition probabilities.

Example 1. Consider the Markov chain of Figure 1.

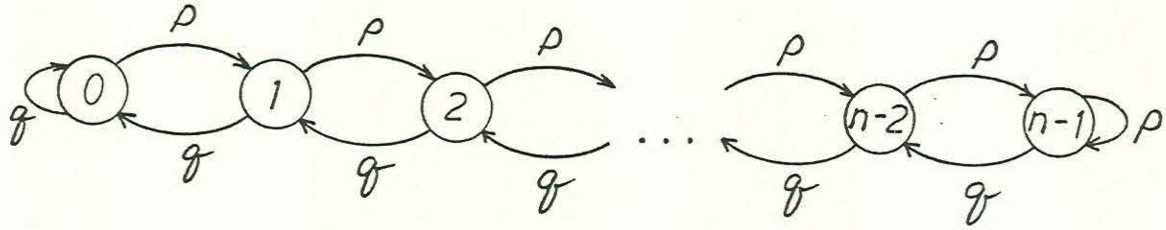

Figure 1: *A strongly connected Markov chain.*

In this strongly connected chain, $q = 1 - p$ and the single tree with root j has j "right-going" arcs of weight p ($0 \to 1 \to \cdots \to j$) and $n-1-j$ "left-going" arcs of weight q ($j \leftarrow j+1 \leftarrow \cdots \leftarrow n-1$). The weight of the spanning tree with root j is thus $p^j q^{n-1-j}$, so

$$\pi_j = \frac{p^j q^{n-1-j}}{\sum_{i=0}^{n-1} p^i q^{n-1-i}}.$$

Example 2. This is like the last example except that states 0 and $n-1$ are "absorbing barriers". Consider a gambler who begins with i dollars and keeps betting 1 dollar at a time, with probability p of winning two dollars back and $q = 1 - p$ of losing his dollar each time. We assume that the gambler doesn't quit until he is "ruined" (has no money left) or until he holds $n-1$ dollars. Then $\overline{p}_{i0}$ is the probability he will be ruined and $\overline{p}_{in-1}$ is the probability he will go home with $n-1$ dollars.

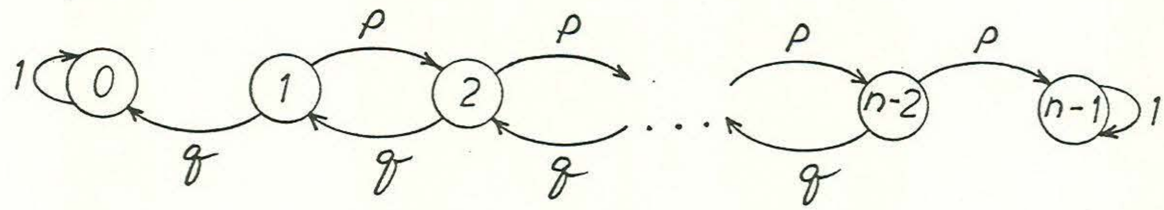

Figure 2: *A Markov chain for the gambler's problem.*

Each arborescence consists of exactly $n-2$ arcs, j of which will be "left-going" with weight q $(0 \leftarrow 1 \leftarrow \cdots \leftarrow j)$ and $n-j-2$ of which will be "right-going" with weight p $(j+1 \rightarrow j+2 \rightarrow \cdots \rightarrow n-1)$. Thus

$$\overline{p}_{i0} = \frac{\sum_{j=i}^{n-2} q^j p^{(n-2)-j}}{\sum_{j=0}^{n-2} q^j p^{(n-2)-j}} \qquad \text{and} \qquad \overline{p}_{in-1} = \frac{\sum_{j=0}^{i-1} q^j p^{(n-2)-j}}{\sum_{j=0}^{n-2} q^j p^{(n-2)-j}}.$$

Example 3. Consider the Markov chain of Figure 3.

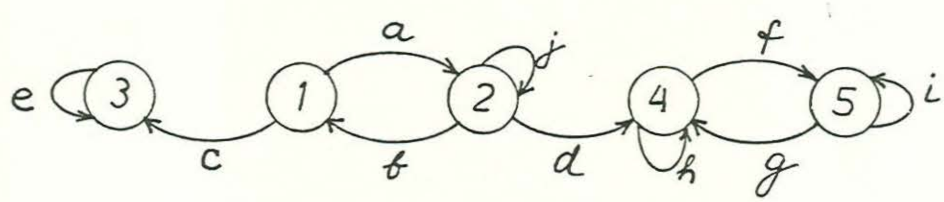

Figure 3: *A Markov chain with six arborescences.*

In this example, there are six arborescences: $adf, adg, bcf, bcg, cdf,$ and cdg. Thus $\|\mathcal{A}(V)\| = (ad+bc+cd)\cdot(f+g)$ and

$$\overline{P} = \frac{1}{\|\mathcal{A}(V)\|} \cdot \begin{pmatrix} 0 & 0 & (bc+cd)(f+g) & adg & adf \\ 0 & 0 & bc(f+g) & (ad+cd)g & (ad+cd)f \\ 0 & 0 & (ad+bc+cd)(f+g) & 0 & 0 \\ 0 & 0 & 0 & (ad+bc+cd)g & (ad+bc+cd)f \\ 0 & 0 & 0 & (ad+bc+cd)g & (ad+bc+cd)f \end{pmatrix}.$$

For example, $\overline{p}_{24} = \frac{(ad+cd)g}{(ad+bc+cd)(f+g)}$ because the only two arborescences which have 4 as a root and a path from 2 to 4 are those with weights adg and cdg.

2. Graph Identities

It is well-known that the states of any Markov chain can be decomposed into a set T of transient states and sets $B_1, B_2, \ldots, B_m$ of minimal closed subsets of states. For example, $T = \{1,2\}$, $B_1 = \{3\}$ and $B_2 = \{4,5\}$ in Figure 3. For any subset of states $W \subseteq V$, define $c(W)$ to be the number of minimal closed subsets of states contained in W. For example, $c(\{1,2,4\}) = 0$ and $c(\{3,4\}) = 1$ in Figure 3. As we remarked in the introduction, every arborescence has $|V| - c(V)$ edges. The following lemma states a simple but important fact about $c(W)$.

Lemma 1: *If U and W are disjoint subsets of V and if there are no edges from W to U in E, then $c(U \bigcup W) = c(U) + c(W)$.*

Proof: Every minimal closed subset in U or W is a minimal closed subset in $U \bigcup W$. Thus $c(U \bigcup W) \geq c(U) + c(W)$. If a closed subset of $U \bigcup W$ contains nodes in both U and W, then the portion of the subset in W is also closed (since there are no edges from W to U). Thus the original subset is not minimal, implying that $c(U \bigcup W) \leq c(U) + c(W)$. Thus $c(U \bigcup W) = c(U) + c(W)$, as claimed. ∎

Given any subset of nodes $W \subseteq V$, define an *arborescence from* W to be an acyclic subgraph of $G = (V, E)$ for which the outdegree of nodes in W is at most one and for which the outdegree of nodes in $V - W$ is zero. Let $\mathcal{A}^r(W)$ denote the set of arborescences from W with r edges, $\mathcal{A}^r_j(W)$ denote the set of arborescences from W with root j and r edges, and $\mathcal{A}^r_{ij}(W)$ denote the set of arborescences from W with root j, a path from i to j and r edges. (If $i = j$, then $\mathcal{A}^r_{ij}(W)$ is defined to be $\mathcal{A}^r_j(W)$.) As we are particularly interested in arborescences with $|W| - c(W)$ edges, we use $\mathcal{A}(W)$, $\mathcal{A}_j(W)$ and $\mathcal{A}_{ij}(W)$ to denote the sets $\mathcal{A}^{|W|-c(W)}(W)$, $\mathcal{A}^{|W|-c(W)}_j(W)$ and $\mathcal{A}^{|W|-c(W)}_{ij}(W)$, respectively. For example, $\mathcal{A}_{ij}(W)$ denotes the set of arborescences from W with root j, a path from i to j, and $|W| - c(W)$ edges.

Notice that the definitions for $\mathcal{A}(V)$, $\mathcal{A}_j(V)$ and $\mathcal{A}_{ij}(V)$ provided here are equivalent to those given in the introduction. This is because every *maximum* arborescence has $|V| - c(V)$ edges. Also notice that $\mathcal{A}_j(W)$ and $\mathcal{A}_{ij}(W)$ may be empty for some W. This happens when node j is not contained in a minimal closed subset of W and/or when there is no path from i to j in G. When W is nonempty, $\mathcal{A}(W)$ is nonempty. In general, $\mathcal{A}^r(W)$ will be empty precisely when $r > |W| - c(W)$.

The *weight* of an arborescence from W and the $||\mathcal{A}||$ notation are defined as in the introduction. Using Lemma 1, the following identities are easily established.

Lemma 2: *Let U and W be disjoint subsets of V such that there are no edges from W to U. Also let $i, i' \in U$ and $j, j' \in W$ be arbitrary vertices. Then*

$$\begin{aligned}
||\mathcal{A}(U \bigcup W)|| &= ||\mathcal{A}(U)|| \cdot ||\mathcal{A}(W)|| \\
||\mathcal{A}_i(U \bigcup W)|| &= ||\mathcal{A}_i(U)|| \cdot ||\mathcal{A}(W)|| \\
||\mathcal{A}_j(U \bigcup W)|| &= ||\mathcal{A}(U)|| \cdot ||\mathcal{A}_j(W)|| \\
||\mathcal{A}_{ii'}(U \bigcup W)|| &= ||\mathcal{A}_{ii'}(U)|| \cdot ||\mathcal{A}(W)|| \\
||\mathcal{A}_{jj'}(U \bigcup W)|| &= ||\mathcal{A}(U)|| \cdot ||\mathcal{A}_{jj'}(W)|| \\
||\mathcal{A}_{ij}(U \bigcup W)|| &= \sum^{j' \in W} ||\mathcal{A}_{ij'}(U)|| \cdot ||\mathcal{A}_{j'j}(W)||
\end{aligned}$$

Proof: The union of an arborescence from U with $|U| - c(U)$ edges and an arborescence from W with $|W| - c(W)$ edges is an arborescence from $U \bigcup W$ with $|U| - c(U) + |W| - c(W) = |U \bigcup W| - c(U \bigcup W)$ edges. (No cycles can be formed in the union since there are no edges from W to U.) Conversely, an arborescence from $U \bigcup W$ with $|U \bigcup W| - c(U \bigcup W)$ edges can have at most $|U| - c(U)$ edges from nodes in U and at most $|W| - c(W)$ edges from W. Hence, the arborescence can be uniquely expressed as the union of an arborescence from U with $|U| - c(U)$ edges and an arborescence from W with $|W| - c(W)$ edges. Thus $||\mathcal{A}(U \bigcup W)|| = ||\mathcal{A}(U)|| \cdot ||\mathcal{A}(W)||$. The remaining identities can be similarly proved. ∎

At first glance, it is not at all clear why sums of weighted arborescences should be related to long-run transition probabilities. We will demonstrate this connection by showing that each is related to sums of weighted paths in the chain. For example, let $\mathcal{P}^r_{ij}(W)$ denote the set of paths from i to j through W that have r edges. (A *path from i to j through* W is a path starting at i and ending at j which traverses only nodes in W. Note that i and j need not be in W.) By definition,

$$\overline{p}_{ij} = \lim_{t \to \infty} \frac{1}{t} \sum^{0 \le r < t} ||\mathcal{P}^r_{ij}(V)||$$

where $\|\mathcal{P}^r_{ij}(V)\|$ denotes the sum of the weights of the paths in $\mathcal{P}^r_{ij}(V)$. (The *weight* of a path is simply the product of the weights of its edges.) Thus the long-run probabilities are closely related to the sums of weighted paths.

For the most part, we will be interested in the set of paths from i to j through W which have at least one edge. We denote this set by $\mathcal{P}_{ij}(W)$ and the sum of the weights of these paths by $\|\mathcal{P}_{ij}(W)\|$. If $j \notin W$ and $c(W) = 0$, then $\|\mathcal{P}_{ij}(W)\|$ is the probability that a sequence of transitions will (once having left state i) move to state j upon leaving W. (If $i \notin W$ and the path has just one edge, then the path never enters W.) This value is closely related to the *taboo Greene function* described in [9]. The following lemmas contain identities for $\|\mathcal{P}^r_{ij}(W)\|$ and $\|\mathcal{P}_{ij}(W)\|$ that are useful in establishing the connection between paths and arborescences.

Lemma 3: *For all i, j, r and W, $\|\mathcal{P}^r_{ij}(W)\| \leq 1$. If $c(W + \{j\}) = 0$, then there is a constant $\alpha < 1$ such that $\|\mathcal{P}^r_{ij}(W)\| \leq \alpha^{\lfloor r/n \rfloor}$. If $c(W + \{j\}) = 0$ and $j \notin W$, then $\|\mathcal{P}_{jj}(W)\| < 1$.*

Proof: By definition, $\|\mathcal{P}^r_{ij}(W)\|$ is the probability that, starting in state i, r successive transitions will pass through states in W and terminate in state j. This probability is clearly at most one.

If $c(W + \{j\}) = 0$, then $W + \{j\}$ contains no closed subsets. Hence, every node in W is linked by a path (with nonzero weight) of length n to a node outside $W + \{j\}$. Let ϵ be the minimum of the weights of these paths and set $\alpha = 1 - \epsilon$. The probability of staying in $W + \{j\}$ through $s + n$ transitions is thus at most α times the probability of staying in $W + \{j\}$ through s transitions. Hence $\|\mathcal{P}^r_{ij}(W)\|$ is at most $\alpha^{\lfloor r/n \rfloor}$.

If $c(W + \{j\}) = 0$ and $j \notin W$, then there is a path from j to a node not in $W + \{j\}$. If this path has weight ϵ, then $\|\mathcal{P}_{jj}(W)\| \leq 1 - \epsilon$. ∎

Lemma 4: *If $c(W) = 0$, then*

$$\|\mathcal{P}_{ij}(W)\| = \|\mathcal{P}_{ij}(\emptyset)\| + \sum^{k \in W} \frac{\|\mathcal{P}_{ik}(\emptyset)\| \cdot \|\mathcal{P}_{kj}(W - \{k\})\|}{1 - \|\mathcal{P}_{kk}(W - \{k\})\|}.$$

Proof: Excluding the one-edge path from i to j (if it exists), every path from i to j through W can be decomposed into a one-edge path from i to k, some nonnegative number of paths from k to k through $W - \{k\}$, and a path from k to j through $W - \{k\}$. Hence,

$$\|\mathcal{P}_{ij}(W)\| = \|\mathcal{P}_{ij}(\emptyset)\| + \sum^{k \in W} \sum^{t \geq 0} \|\mathcal{P}_{ik}(\emptyset)\| \cdot \|\mathcal{P}_{kk}(W - \{k\})\|^t \cdot \|\mathcal{P}_{kj}(W - \{k\})\|.$$

By Lemma 3, $\|\mathcal{P}_{kk}(W - \{k\})\| < 1$ and thus

$$\sum^{t \geq 0} \|\mathcal{P}_{kk}(W - \{k\})\|^t = \frac{1}{1 - \|\mathcal{P}_{kk}(W - \{k\})\|}.$$

Substitution then gives the desired result. ∎

The connection between paths and arborescences is established in the following theorem.

Theorem 2 (The Path Tree Theorem): *If $c(W+\{i\})=0$, then*

$$\|\mathcal{P}_{ij}(W)\| = \begin{cases} \frac{\|\mathcal{A}_{ij}(W+\{i\}-\{j\})\|}{\|\mathcal{A}(W)\|}, & \text{if } i \neq j; \\ \frac{\|\mathcal{A}(W-\{i\})\|-\|\mathcal{A}(W+\{i\})\|}{\|\mathcal{A}(W)\|}, & \text{if } i = j. \end{cases}$$

Proof: The proof is by induction on $|W|$. If $W=\emptyset$, then $\|\mathcal{P}_{ij}(\emptyset)\| = p_{ij}$. Similarly, $\|\mathcal{A}_{ij}(\{i\})\| = p_{ij}$ for $i \neq j$ and $1-\|\mathcal{A}(\{i\})\| = p_{ii}$. In what follows, we assume that the result is true for sets smaller than W. There are five cases to consider, depending on i and j. Each case starts with the identity from Lemma 4 and proceeds by applying the hypothesis inductively. Indentities like those proved in Lemma 2 are used throughout the proof.

Case 1: $i \neq j$, $i \notin W$, $j \notin W$

$$\begin{aligned}
\|\mathcal{P}_{ij}(W)\| &= \|\mathcal{A}_{ij}(\{i\})\| + \sum^{k\in W} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \frac{\|\mathcal{A}_{kj}(W-\{j\})\|}{\|\mathcal{A}(W-\{k\})\|}}{1-\frac{\|\mathcal{A}(W-\{k\})\|-\|\mathcal{A}(W)\|}{\|\mathcal{A}(W-\{k\})\|}} \\
&= \|\mathcal{A}_{ij}(\{i\})\| + \sum^{k\in W} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{kj}(W-\{j\})\|}{\|\mathcal{A}(W)\|} \\
&= \sum^{k\in W+\{j\}} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{kj}(W-\{j\})\|}{\|\mathcal{A}(W)\|} \\
&= \frac{\|\mathcal{A}_{ij}(W+\{i\}-\{j\})\|}{\|\mathcal{A}(W)\|}.
\end{aligned}$$

Case 2: $i \neq j, i \notin W, j \in W$

$$\begin{aligned}
\|\mathcal{P}_{ij}(W)\| = \|\mathcal{A}_{ij}(\{i\})\| &+ \sum^{k\in W-\{j\}} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{kj}(W-\{j\})\|}{\|\mathcal{A}(W)\|} \\
&+ \frac{\|\mathcal{A}_{ij}(\{i\})\| \cdot (\|\mathcal{A}(W-\{j\})\| - \|\mathcal{A}(W)\|)}{\|\mathcal{A}(W)\|} \\
= \sum^{k\in W} & \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{kj}(W-\{j\})\|}{\|\mathcal{A}(W)\|} \\
= \frac{\|\mathcal{A}_{ij}(W+\{i\}-\{j\})\|}{\|\mathcal{A}(W)\|}&.
\end{aligned}$$

Case 3: $i \neq j, i \in W$

$$\begin{aligned}
\|\mathcal{P}_{ij}(W)\| &= \sum^{k\in W+\{j\}-\{i\}} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{kj}(W-\{j\})\|}{\|\mathcal{A}(W)\|} + \frac{(1-\|\mathcal{A}(\{i\})\|) \cdot \|\mathcal{A}_{ij}(W+\{i\}-\{j\})\|}{\|\mathcal{A}(W)\|} \\
&= \frac{\|\mathcal{A}_{ij}(W+\{i\}-\{j\})\|}{\|\mathcal{A}(W)\|} + \sum^{k\in W+\{j\}-\{i\}} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{kj}(W-\{j\})\|}{\|\mathcal{A}(W)\|} \\
&\qquad - \frac{\|\mathcal{A}(\{i\})\| \cdot \|\mathcal{A}_{ij}(W-\{j\})\|}{\|\mathcal{A}(W)\|} \\
&= \frac{\|\mathcal{A}_{ij}(W+\{i\}-\{j\})\|}{\|\mathcal{A}(W)\|}.
\end{aligned}$$

Case 4: $i = j, i \notin W$

$$
\begin{aligned}
\|\mathcal{P}_{ii}(W)\| &= 1 - \|\mathcal{A}(\{i\})\| + \sum^{k \in W} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{ki}(W - \{i\})\|}{\|\mathcal{A}(W)\|} \\
&= \frac{\|\mathcal{A}(W - \{i\})\|}{\|\mathcal{A}(W)\|} - \left(\frac{\|\mathcal{A}(\{i\})\| \cdot \|\mathcal{A}(W)\|}{\|\mathcal{A}(W)\|} - \sum^{k \in W} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{ki}(W)\|}{\|\mathcal{A}(W)\|} \right) \\
&= \frac{\|\mathcal{A}(W - \{i\})\| - \|\mathcal{A}(W + \{i\})\|}{\|\mathcal{A}(W)\|}.
\end{aligned}
$$

Case 5: $i = j, i \in W$

$$
\begin{aligned}
\|\mathcal{P}_{ii}(W)\| = 1 - \|\mathcal{A}(\{i\})\| &+ \sum^{k \in W - \{i\}} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{ki}(W - \{i\})\|}{\|\mathcal{A}(W)\|} \\
&\qquad + \frac{(1 - \|\mathcal{A}(\{i\})\|) \cdot (\|\mathcal{A}(W - \{i\})\| - \|\mathcal{A}(W)\|)}{\|\mathcal{A}(W)\|} \\
&= \frac{(1 - \|\mathcal{A}(\{i\})\|) \cdot \|\mathcal{A}(W - \{i\})\|}{\|\mathcal{A}(W)\|} + \sum^{k \in W - \{i\}} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{ki}(W - \{i\})\|}{\|\mathcal{A}(W)\|} \\
&= \frac{\|\mathcal{A}(W - \{i\})\|}{\|\mathcal{A}(W)\|} - \left(\frac{\|\mathcal{A}(\{i\})\| \cdot \|\mathcal{A}(W - \{i\})\|}{\|\mathcal{A}(W)\|} - \sum^{k \in W - \{i\}} \frac{\|\mathcal{A}_{ik}(\{i\})\| \cdot \|\mathcal{A}_{ki}(W - \{i\})\|}{\|\mathcal{A}(W)\|} \right) \\
&= \frac{\|\mathcal{A}(W - \{i\})\| - \|\mathcal{A}(W + \{i\})\|}{\|\mathcal{A}(W)\|}.
\end{aligned}
$$

This completes the induction. ∎

Corollary 2: *If B is a minimal closed set and $i, j \in B$, then*

$$\|\mathcal{P}_{ij}(B - \{i\})\| = \frac{\|\mathcal{A}_{ij}(B - \{j\})\|}{\|\mathcal{A}(B - \{i\})\|}.$$

Proof: For $i \neq j$, apply case 2 of Theorem 2. When $i = j$, apply case 4. (Note that when $i = j$, $\|\mathcal{P}_{ii}(B - \{i\})\| = 1$.) ∎

3. Proofs

In what follows, we present two proofs of the Markov Chain Tree Theorem. The first proof is derived from the graph identities described in Section 2. The second proof is derived from the Matrix Tree Theorem.

3a. Graph Theoretic Proof

As we remarked in Section 2,

$$\overline{p}_{ij} = \lim_{t\to\infty} \frac{1}{t} \sum^{0\le r<t} \|\mathcal{P}^r_{ij}(V)\|.$$

In what follows, we will prove the Markov Chain Tree Theorem by showing that

$$\lim_{t\to\infty} \frac{1}{t} \sum^{0\le r<t} \|\mathcal{P}^r_{ij}(V)\| = \frac{\|\mathcal{A}_{ij}(V)\|}{\|\mathcal{A}(V)\|}.$$

If there is no path from i to j in M, then both sides of the equation are zero. Otherwise we must consider three cases.

Case 1: $i, j \in T$, the set of transient states.

Here $\|\mathcal{A}_{ij}(V)\| = 0$ since every arborescence with j as a root has less than $|V| - c(V)$ edges. On the other hand, $\|\mathcal{P}^r_{ij}(V)\| = \|\mathcal{P}^r_{ij}(T)\|$ which is at most $\alpha^{\lfloor r/n \rfloor}$ for some $\alpha < 1$ by Lemma 3. Thus

$$\begin{aligned}\lim_{t\to\infty} \sum^{0\le r<t} \|\mathcal{P}^r_{ij}(V)\| &\le \lim_{t\to\infty} \frac{1}{t} \sum^{0\le r<t} \alpha^{\lfloor r/n \rfloor} \\ &\le \lim_{t\to\infty} \frac{1}{t}\left(\frac{n}{1-\alpha}\right) \\ &= 0.\end{aligned}$$

Case 2: $i, j \in B$, a minimal closed block.

We first show that $\overline{p}_{ij} = \|\mathcal{P}_{ij}(B - \{i\})\| \cdot \overline{p}_{ii}$. By Corollary 2, $\|\mathcal{P}_{ii}(B - \{i\})\| = 1$ and thus we need only consider the case when $i \ne j$. By definition,

$$\begin{aligned}\|\mathcal{P}_{ij}(B-\{i\})\| \cdot \overline{p}_{ii} - \overline{p}_{ij} &= \lim_{t\to\infty} \frac{1}{t}\left(\sum^{0\le r<t} \|\mathcal{P}^r_{ii}(B)\| \cdot \sum^{s\ge 1} \|\mathcal{P}^s_{ij}(B-\{i\})\| - \sum^{0\le q<t} \|\mathcal{P}^q_{ij}(B)\| \right) \\ &= \lim_{t\to\infty} \frac{1}{t}\left(\sum^{0\le r<t} \|\mathcal{P}^r_{ii}(B)\| \cdot \sum^{s\ge t-r} \|\mathcal{P}^s_{ij}(B-\{i\})\| \right) \\ &\le \lim_{t\to\infty} \frac{1}{t}\left(\sum^{0\le r<t} \sum^{s\ge t-r} \alpha^{\lfloor s/n \rfloor} \right) \quad \text{for some } \alpha < 1 \text{ by Lemma 3} \\ &\le \lim_{t\to\infty} \frac{1}{t}\left(\sum^{0\le r<t} \frac{n\alpha^{\lfloor \frac{t-r}{n} \rfloor}}{1-\alpha} \right) \\ &\le \lim_{t\to\infty} \frac{1}{t}\left(\frac{n^2}{(1-\alpha)^2} \right) \\ &= 0, \text{ as claimed.}\end{aligned}$$

By Corollary 2, this means that

$$\overline{p}_{ij} = \frac{\|\mathcal{A}_{ij}(B-\{j\})\|}{\|\mathcal{A}(B-\{i\})\|} \cdot \overline{p}_{ii}.$$

Summing over j, we find that

$$1 = \sum^{j\in B} \frac{\|\mathcal{A}_{ij}(B-\{i\})\|}{\|\mathcal{A}(B-\{i\})\|} \cdot \overline{p}_{ii}$$
$$= \frac{\|\mathcal{A}(B)\|}{\|\mathcal{A}(B-\{i\})\|} \cdot \overline{p}_{ii}$$

and thus that

$$\overline{p}_{ii} = \frac{\|\mathcal{A}(B-\{i\})\|}{\|\mathcal{A}(B)\|}.$$

Substituting, this gives

$$\overline{p}_{ij} = \frac{\|\mathcal{A}_{ij}(B-\{j\})\|}{\|\mathcal{A}(B)\|}$$
$$= \frac{\|\mathcal{A}_{ij}(B)\|}{\|\mathcal{A}(B)\|}$$
$$= \frac{\|\mathcal{A}_{ij}(V)\|}{\|\mathcal{A}(V)\|}.$$

Case 3: $i \in T, j \in B$.

We first show that

$$\overline{p}_{ij} = \sum^{k\in B} \|\mathcal{P}_{ik}(T)\| \cdot \overline{p}_{kj}.$$

By definition,

$$\sum^{k\in B} \|\mathcal{P}_{ik}(T)\| \cdot \overline{p}_{kj} - \overline{p}_{ij} = \lim_{t\to\infty} \frac{1}{t}\left(\sum^{k\in B} \sum^{s\geq 1} \|\mathcal{P}^s_{ik}(T)\| \cdot \sum^{0\leq r<t} \|\mathcal{P}^r_{kj}(B)\| - \sum^{0\leq q<t} \|\mathcal{P}^q_{ij}(T \bigcup B)\|\right)$$
$$= \lim_{t\to\infty} \frac{1}{t}\left(\sum^{k\in B} \sum^{0\leq r<t} \|\mathcal{P}^r_{kj}(B)\| \cdot \sum^{s\geq t-r} \|\mathcal{P}^s_{ik}(T)\|\right)$$
$$\leq \lim_{t\to\infty} \frac{1}{t}\left(\sum^{k\in B} \sum^{0\leq r<t} \sum^{s\geq t-r} \alpha^{\lfloor s/n \rfloor}\right) \quad \text{for some } \alpha < 1$$
$$\leq \lim_{t\to\infty} \frac{1}{t}\left(\frac{n^3}{(1-\alpha)^2}\right)$$
$$= 0, \quad \text{as claimed.}$$

By Theorem 2 and the analysis in case 2, the preceding means that

$$\overline{p}_{ij} = \sum^{k\in B} \frac{\|\mathcal{A}_{ik}(T)\|}{\|\mathcal{A}(T)\|} \cdot \frac{\|\mathcal{A}_{kj}(B)\|}{\|\mathcal{A}(B)\|}$$
$$= \frac{\|\mathcal{A}_{ij}(T \bigcup B)\|}{\|\mathcal{A}(T \bigcup B)\|}$$
$$= \frac{\|\mathcal{A}_{ij}(V)\|}{\|\mathcal{A}(V)\|}.$$

This completes the proof of Theorem 1. ∎

3b. Proof Using the Matrix Tree Theorem

Let X be an arbitrary real-valued $n \times n$ matrix. We let $C_k(X)$ denote the $n \times n$ matrix obtained from X by replacing its kth column by a length n vector of ones. We let $D_{ij}(X)$ denote the $(n-1) \times (n-1)$ matrix obtained from X by deleting its ith row and jth column. If A and B are sets we also let $D_{AB}(X)$ denote the matrix obtained from X by deleting all rows in A and all columns in B. The following lemma contains some simple identities for the determinants of these matrices. (The *determinant* of a matrix X is denoted by $|X|$.)

Lemma 5: Let X be an $n \times n$ stochastic matrix. Then

$$
\begin{aligned}
|C_i(X)| &= |C_j(X)| \text{ for } 1 \leq i,j \leq n \\
|D_{ij}(X)| &= (-1)^{i+j}|D_{ii}(X)| \text{ for } 1 \leq i,j \leq n \\
|C_k(X)| &= \sum_{i=1}^{n} |D_{ii}(X)| \text{ for } 1 \leq k \leq n.
\end{aligned}
$$

Proof: Straightforward. ∎

A general version of the Matrix Tree Theorem [1] can be stated as follows.

Theorem 3 (Matrix Tree Theorem): *Let the $n \times n$ matrix X have entries x_{ij} where*

$$
\begin{aligned}
x_{ij} &= -y_{ij} \textit{ for } i \neq j, \textit{ and} \\
x_{ii} &= -y_{ii} + \sum_{k=1}^{n} y_{ik}.
\end{aligned}
$$

Define an associated graph G with $V = \{1, \ldots, n\}$ and $E = \{(i,j) \mid y_{ij} \neq 0\}$ having weight y_{ij} on edge (i,j). Let $B \subseteq V$, $i,j \in V - B$ and $r = n - |B|$. Then

$$
\begin{aligned}
|D_{B,B}(X)| &= ||\mathcal{A}^r(V-B)||, \quad \textit{and} \\
(-1)^{i+j}|D_{B+j,B+i}(X)| &= ||\mathcal{A}_{ij}^{r-1}(V-B-\{j\})||.
\end{aligned}
$$

Proof: See [1]. ∎

We now proceed with the second proof of the Markov Chain Tree Theorem, starting first with the case that the Markov chain M is irreducible. In this case each row of $\overline{P}$ is equal to the vector π which is defined as the unique solution to:

$$\pi P = \pi, \text{ and } \sum_{k=1}^{n} \pi_k = 1.$$

The vector π is the stationary probability vector for M if M is aperiodic.

Since P is stochastic, the above defining conditions on π can be combined to read:

$$\pi C_k(I-P) = \epsilon_k$$

where I denotes the identity matrix and ϵ_k denotes the vector having a one in column k and zeros elsewhere. This equation uniquely defines π, for any $k, 1 \leq k \leq n$.

We now use Cramer's Rule to solve for π:

$$\pi_k = \frac{|D_{kk}(I-P)|}{|C_k(I-P)|}.$$

Note that Lemma 5 implies that $|C_k(I-P)| = |C_l(I-P)|$ even if $k \neq l$, so the denominators of the equations for the π_k are all the same.

A simple application of the Matrix Tree Theorem to the evaluation of $|D_{kk}(I-P)|$ then completes the proof for irreducible Markov chains.

We now generalize our result to include all Markov chains. As before, partition the states of M into a set T transient states, and sets $B_1, \ldots, B_m$ of minimal closed subsets of states.

We let P_k denote the $|B_k| \times |B_k|$ submatrix of P giving the transition probabilities within B_k, Q denote the $|T| \times |T|$ matrix of transition probabilities within T, and R_k denote the $|B_k| \times |T|$ matrix of transition probabilities from B_k to T.

By appropriate reordering the rows and columns of P we have

$$P = \begin{pmatrix} Q & R_1 & R_2 & \ldots & R_m \\ 0 & P_1 & 0 & \ldots & 0 \\ 0 & 0 & P_2 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & P_m \end{pmatrix}.$$

It is well-known that $\overline{P}$ then has the following form:

$$\overline{P} = \begin{pmatrix} 0 & U_1 & U_2 & \ldots & U_m \\ 0 & \overline{P}_1 & 0 & \ldots & 0 \\ 0 & 0 & \overline{P}_2 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & \overline{P}_m \end{pmatrix},$$

where $\overline{P}_k$ is the long-run transition matrix for P_k,

$$U_k = NR_k\overline{P}_k, \text{ and}$$

$$N = (I + Q + Q^2 + \cdots) = (I-Q)^{-1}.$$

Here n_{ij} is the average number of times M will visit state j, when M starts in state i. The matrix N always exists [6, Lemma III.4.1]. In fact, we will show in what follows that

$$n_{ij} = \frac{||\mathcal{A}_{ij}(T - \{j\})||}{||\mathcal{A}(T)||}.$$

By definition,

$$\begin{aligned} n_{ij} &= ((I-Q)^{-1})_{ij} \\ &= \frac{(-1)^{i+j}|D_{ji}(I-Q)|}{|I-Q|} \\ &= \frac{(-1)^{i+j}|D_{V-T+\{j\},V-T+\{i\}}(I-P)|}{|D_{V-T,V-T}(I-P)|} \\ &= \frac{||\mathcal{A}_{ij}(T-\{j\})||}{||\mathcal{A}(T)||} \quad \text{by the Matrix Tree Theorem.} \end{aligned}$$

It is clear that both $\overline{p}_{ij}$ and $||\mathcal{A}_{ij}(V)||$ are zero unless $i, j \in B_k$ (one of the closed subsets), or $i \in T$ (the set of transient states) and $j \in B_k$. In the former case, $\overline{p}_{ij} = (\overline{P}_k)_{ij}$. From the analysis of irreducible chains, this means that $\overline{p}_{ij} = \frac{||\mathcal{A}_{ij}(B_k)||}{||\mathcal{A}(B_k)||}$ and thus that $\overline{p}_{ij} = \frac{||\mathcal{A}_{ij}(V)||}{||\mathcal{A}(V)||}$.

If $i \in T$ and $j \in B_k$, then

$$
\begin{aligned}
\overline{p}_{ij} &= (NR_k\overline{P}_k)_{ij} \\
&= \sum^{l \in B_k} \sum^{l' \in T} \frac{||\mathcal{A}_{il'}(T - \{l'\})||}{||\mathcal{A}(T)||} \cdot ||\mathcal{A}_{l'l}(\{l'\})|| \cdot \frac{||\mathcal{A}_{lj}(B_k)||}{||\mathcal{A}(B_k)||} \\
&= \sum^{l \in B_k} \frac{||\mathcal{A}_{il}(T)||}{||\mathcal{A}(T)||} \cdot \frac{||\mathcal{A}_{lj}(B_k)||}{||\mathcal{A}(B_k)||} \\
&= \frac{||\mathcal{A}_{ij}(T \bigcup B_k)||}{||\mathcal{A}(T \bigcup B_k)||} \\
&= \frac{||\mathcal{A}_{ij}(V)||}{||\mathcal{A}(V)||}. \quad \blacksquare
\end{aligned}
$$

4. Remarks

Throughout the paper, arborescences and paths are defined for graphs with *nonzero*-weight edges. This restriction complicates the proofs somewhat but is necessary to insure their correctness. For example, a "simpler method" might involve substitution of ϵ-weight edges for zero-weight edges in the Markov chain. Then all of the arborescences would be isomorphic (except for weights). By letting $\epsilon \to 0$ and taking the limit in Theorem 1, one might hope to derive a simpler proof. This is not possible, however, as is evidenced by the 2-state example in Figure 4. When ϵ-weight edges are substituted for the zero-weight edges, $\overline{p}_{12} = 1/2$ no matter what the value of ϵ is. Hence $\lim_{\epsilon \to 0} \overline{p}_{12}(\epsilon) = 1/2$. As can be easily seen, however, $\overline{p}_{12}(0) = 0$.

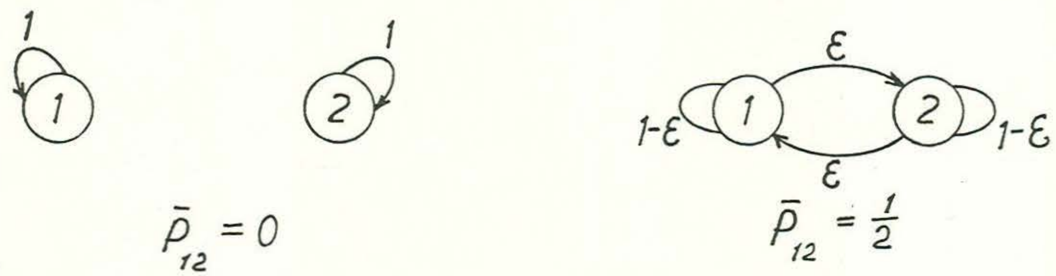

Figure 4: *The difference between 0-weight edges and ϵ-weight edges.*

It is worth noting that $||\mathcal{A}_{ij}(V)||$ and $||\mathcal{A}(V)||$ are *nonnegative* polynomials of degree $|V| - c(V)$ in the variables $\{p_{ij} \mid i \neq j\}$. (By *nonnegative*, we mean that *every coefficient* in the polynomials is nonnegative.) Thus the long-run transition probabilities can be expressed as the ratio of nonnegative polynomials with degree $n - 1$. This result is surprisingly powerful. For example, we use this fact in [8] to show that no n-state automata can estimate a probability with mean-square-error less than $\Omega(1/n)$.

It is also worth noting that the graph theoretic proof of the Markov Chain Tree Theorem can be simply modified to give a natural proof of the Matrix Tree Theorem. Although the proof gives some insight as to why arborescences are important, it does not illuminate their meaning. In fact, it would be very nice to have an interpretation for the weight of an arborescence in the Markov Chain Tree Theorem.

5. Acknowledgements

We first would like to thank Daniel Kleitman for suggesting that the Matrix Tree Theorem might be applicable when computing stationary probabilities of Markov chains. We would also like to thank Seth Chaiken, Herman Chernoff, Richard Dudley, Samuel Karlin, Gary Miller, Larry Shepp and Lorie Snell for helpful remarks and references.

6. References

[1] S. Chaiken, "A Combinatorial Proof of the All Minors Matrix Tree Theorem," *SIAM Journal of Algebraic and Discrete Methods*, Vol. 13, No. 3, September 1982, pp. 319-329.

[2] S. Chaiken and D. Kleitman, "Matrix Tree Theorems," *Journal of Combinatorial Theory*, Series A, Vol. 24, May 1978, pp. 377 - 381.

[3] C. Coates, "Flow-Graph Solutions of Linear Algebraic Equations," *IRE Transactions on Circuit Theory*, Vol. CT-6, 1959, pp. 170-187.

[4] D. Cvetković, M. Doob and H. Sachs, *Spectra of Graphs, Theory and Applications*, Academic Press, New York, 1979.

[5] J. Doob, *Stochastic Processes*, Wiley, New York, 1953.

[6] D. Isaacson and R. Madsen, *Markov Chains - Theory and Applications*, Wiley, New York, 1976.

[7] H. Kohler and E. Vollmerhaus, "The Frequency of Cyclic Processes in Biological Multistate Systems," *Journal of Math Biology,* No. 9, 1980, pp. 275-290.

[8] F. Leighton and R. Rivest, "Estimating a Probability Using Finite Memory," *Proceedings of the 1983 International Conference on Foundations of Computation Theory*, Lecture Notes in Computer Science 158, Springer-Verlag, 1983, pp. 255-269.

[9] Q. Minping and Q. Min, "Circulation for Recurrent Markov Chains," *Zeitshrift Varshinenaka*, Vol. 59, No. 2, 1982, pp. 203-210.