AN ELEMENTARY PROOF OF THE MARKOV CHAIN TREE THEOREM

ALEX KRUCKMAN, AMY GREENWALD, JOHN WICKS

Contents

1. Introduction

The Markov Chain Tree Theorem is a classical result which expresses the stable distribution of an irreducible Markov matrix in terms of directed spanning trees of its associated graph. In this article, we present what we believe to be an original elementary proof of the theorem (Theorem 5.1). Our proof uses only linear algebra and graph theory, and in particular, it does not rely on probability theory. For this reason, this article could serve as a pedagogical tool or a gentle introduction to the theory of Markov matrices for undergraduate computer science and mathematics students.

A version of our proof of the Markov Chain Tree Theorem appeared in John Wicks' PhD thesis [4]. Other proofs of the Markov Chain Tree Theorem which use probability theory can be found in Broder [2, Theorem 1], or in more general form in Anantharam and Tsoucas [1]. The interested reader can find more information on Markov chains, matrices, and graphs in Kemeny and Snell [3].

In Section 2, we introduce basic facts and terminology that we will need when working with graphs. In Section 3, we define Markov matrices and provide an algebraic formula for the stable distribution of a unichain Markov matrix. In Section 4, we discuss directed trees and prove the existence of directed spanning trees of unichain graphs. In Section 5, we prove the Markov Chain Tree Theorem by rewriting the algebraic formula for the stable distribution provided in Section 3 as a sum of weights of directed spanning trees.

2. Graph Theory Basics

A *finite directed graph* G is a nonempty finite set of vertices, V, together with a set of edges, $E \subseteq V \times V$. We depict a finite directed graph $G = (V, E)$ by drawing a circle to represent each vertex $v \in V$ and an arrow from the vertex u to the vertex v to represent each edge $(u, v) \in E$. We say that (u, v) *starts* at u and *ends* at v, or that (u, v) is *outgoing* from u and *incoming* to v. An edge of the form (v, v) from a vertex v to itself is called a *self-loop*.

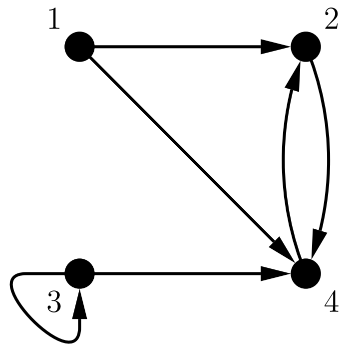


FIGURE 1. A graph

For example, the following figure represents the finite directed graph with vertex set $\{1, 2, 3, 4\}$ and edge set $\{(1, 2), (1, 4), (2, 4), (3, 3), (3, 4), (4, 2)\}$. This graph has a self-loop at vertex 3. From now on, we will refer to finite directed graphs simply as graphs.

To any graph $G = (V, E)$, we may associate a weight function $d : E \to \mathbb{R}$. In this case, we call G a *weighted graph*. We depict a weighted graph by labeling each edge with its weight.

Given a weighted graph G with weight function d, we define the *weight* of G to be the product of the weights of its edges:

$$||G||_d = \prod_{(v_i, v_j) \in E} d(v_i, v_j).$$

A *walk* of length l in a graph G is a sequence of $l + 1$ vertices $(v_0, \ldots, v_l)$ such that for each $1 \leq i \leq l$, there is an edge $(v_{i-1}, v_i) \in E$. The length l refers to the number of edges traversed on the walk. Note that there is always a walk from a vertex to itself, namely the walk of length 0 consisting of that vertex alone. In the example graph of Figure 1, $(1, 4, 2, 4)$ is a walk, because $(1, 4)$, $(4, 2)$, and $(2, 4)$ are all edges in the graph. However, $(1, 3, 4)$ is not a walk, since $(1, 3)$ is not an edge.

We define a binary relation $\sim$ on V, where $u \sim v$ if and only if there is a walk from u to v and a walk from v to u. We would like to show that $\sim$ is an equivalence relation. Let $u, v, w \in V$ be arbitrary vertices. We have $u \sim u$ since there is a walk of length 0 from u to itself, so $\sim$ is reflexive. If $u \sim v$, then $v \sim u$ by definition, so $\sim$ is symmetric. If $u \sim v$ and $v \sim w$, we can concatenate the walks from u to v and from v to w to obtain a walk from u to w. Similarly, we can concatenate the walks from w to v and from v to u to obtain a walk from w to u. Thus $u \sim w$, and $\sim$ is transitive.

We have established that $\sim$ is an equivalence relation. This relation partitions V into equivalence classes, called *strongly connected components* (SCCs). An SCC is called a *closed class* if and only if it has no outgoing edges. Vertices that are do not belong to a closed class are called *transient*. The SCCs of the example graph in Figure 1 are $\{1\}$, $\{3\}$, and $\{2, 4\}$. The class $\{2, 4\}$ is closed, and the vertices 1 and 3 are transient.

The following fundamental lemma shows that every graph contains at least one closed class. A graph is called *unichain* if and only if it contains exactly one closed class.

Lemma 2.1. *Starting from any vertex in a graph G, there exists a walk in G that terminates in a closed class. In particular, every graph contains at least one closed class.*

Proof. Let v be a vertex of G, and let $\mathcal{C}_1$ be its SCC. If $\mathcal{C}_1$ is closed, then we have a walk (of length 0) starting at v and terminating in a closed class, and we are done. Otherwise, $\mathcal{C}_1$ has an outgoing edge to some other SCC $\mathcal{C}_2$, say (u_1, v_2), with $u_1 \in \mathcal{C}_1$ and $v_2 \in \mathcal{C}_2$. Now since v and u_1 are in the same SCC, there is a walk from v to u_1, and continuing along the edge (u_1, v_2), there is a walk from v terminating in $\mathcal{C}_2$.

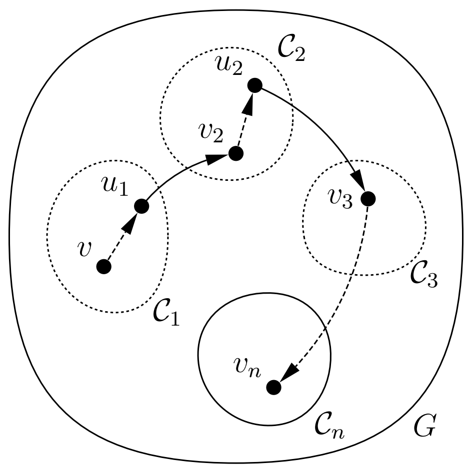


FIGURE 2. Constructing a walk to a closed class

We now repeat the process starting with v_2. If $\mathcal{C}_2$ is closed, we have constructed a walk from v terminating in a closed class. Otherwise, there is a walk starting from v_2 to a vertex v_3 in another SCC, $\mathcal{C}_3$. Concatenating these walks, there is a walk from v terminating in $\mathcal{C}_3$.

In this way we begin constructing a sequence of SCCs, $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \ldots$, such that for all $i > 1$, $\mathcal{C}_{i-1} \neq \mathcal{C}_i$, and for all $j \geq i$, there is a walk from $v_i \in \mathcal{C}_i$ terminating in $\mathcal{C}_j$. Since the number of closed classes is finite, we must eventually arrive at either a closed class, in which case we are done, or an SCC which has already been visited. We will show that the latter case is impossible.

Suppose we have the sequence $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n$, where $\mathcal{C}_n = \mathcal{C}_i$ for some $i < n-1$. Then there is a walk from the vertex v_i in $\mathcal{C}_i$ to the vertex v_{i+1} in $\mathcal{C}_{i+1}$. But there is also a walk from v_{i+1} to the vertex v_n in $\mathcal{C}_n$. Since $\mathcal{C}_n = \mathcal{C}_i$, there is a walk from v_n to v_i, and thus there is a walk from v_{i+1} to v_i. So $v_i \sim v_{i+1}$, contradicting the fact that $\mathcal{C}_i \neq \mathcal{C}_{i+1}$.

Thus, we can construct a walk from any vertex v that terminates in a closed class. In particular, this shows that G has at least one closed class. □

3. Markov Matrices

We will work with $n \times n$ square matrices with real-valued entries. For convenience, we will work with a fixed n for the entire article. For such a matrix M, we write $M_{i,j}$ to refer to the element in the i^{th} row and j^{th} column of M.

A matrix M is called *Markov* if and only if all its entries are non-negative and all its columns sum to 1.

Markov matrices are often used to represent discrete random processes, as follows. Consider a system which may at any time be in one of n states, and suppose that at each of a series of discrete time steps, the system transitions randomly to another state. If the probability of transitioning to state j depends only on the current state i, then we can encode these probabilities as a Markov matrix M by setting $M_{j,i}$ to be the probability of transitioning from state i to state j. Since the total probability of transitioning from state i to any other state must be 1, the columns of M sum to 1.

To every Markov matrix M, we associate a weighted graph $G(M) = (V_n, E)$ with n vertices. As a convention, we will take as the vertex set $V_n = \{1, 2, \ldots, n\}$. Then for all $i, j \in V_n$, $(i, j) \in E$ if and only if $M_{j,i} > 0$. We define a weight function $d : E \to \mathbb{R}$ by $d(i, j) = M_{j,i}$.

In the graph $G(M)$ associated to M, each vertex represents a state in the discrete random process, and the weight of an edge (i, j) represents the probability of transitioning from state i to state j.

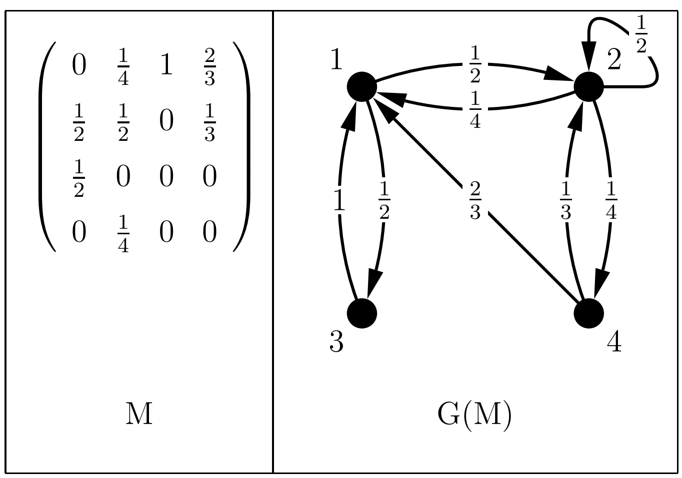


Figure 3. A Markov matrix and its graph

Note that for all i, the weights of the outgoing edges from vertex i correspond to the matrix entries in the i^{th} column. Thus the weight of every edge is positive, and the sum of the weights of the outgoing edges from each vertex is 1.

We call a Markov matrix *unichain*[1] if and only if its corresponding graph is unichain, that is, if it has exactly one closed class. The matrix in Figure 3 is unichain, since its graph has only one SCC, $\{1,2,3,4\}$, which necessarily is a closed class.

A vector in $\mathbb{R}^n$ is called a *distribution* if and only if all its entries are non-negative and its entries sum to 1. A distribution v can be used to represent the probability distribution across states at a given time. That is, the i^{th} entry v_i is the probability that the system is in state i at that time. Multiplying v by M results in the probability distribution across states at the next time step.

We are interested in *stable distributions*, which are fixed by multiplication by M (that is, $Mv = v$). These are eigenvectors with eigenvalue 1. A stable distribution represents a possible limiting behavior of the discrete random process.

Given a Markov matrix, M, the space of eigenvalues with eigenvector 1 is the kernel of $M - I$, since $Mv = v$ if and only if $(M - I)v = Mv - v = 0$. Let $\Lambda = M - I$. This matrix Λ is called the *laplacian* of M. Note that since the columns of M sum to 1, the columns of Λ sum to 0.

In the example of Figure 3, the laplacian of the given Markov matrix M is

$$\begin{pmatrix} -1 & \frac{1}{4} & 1 & \frac{2}{3} \\ \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & -1 & 0 \\ 0 & \frac{1}{4} & 0 & -1 \end{pmatrix}.$$

We claim that $v = (\frac{1}{3}, \frac{2}{5}, \frac{1}{6}, \frac{1}{10})$ is a stable distribution of M. Multiplying, we see that

$$\begin{pmatrix} -1 & \frac{1}{4} & 1 & \frac{2}{3} \\ \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & -1 & 0 \\ 0 & \frac{1}{4} & 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{2}{5} \\ \frac{1}{6} \\ \frac{1}{10} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

so v is an eigenvector of M with eigenvalue 1. Moreover, $\frac{1}{3} + \frac{2}{5} + \frac{1}{6} + \frac{1}{10} = 1$, and all entries are positive, so v is a stable distribution.

The following fact is a slight generalization of a well known result, namely that every Markov matrix has a stable vector. For a proof, see for example Wicks [4] Theorem 5.14.

Theorem 3.1. *If M is a Markov matrix whose graph $G(M)$ has k closed classes, then $\dim(\ker(M - I)) = k$. That is, the space of stable vectors of M has dimension k.*

In particular, the space of stable vectors of a unichain Markov matrix M has dimension 1. We will show that there is a simple way of describing the space of stable vectors by way of the adjugate[2] matrix of M.

Given a matrix A, the minor $\mathrm{Mi}^{i,j}(A)$ is the matrix formed from A by deleting row i and column j. We define the cofactor $\mathrm{Co}^{i,j}(A) = (-1)^{i+j}|\mathrm{Mi}^{i,j}(A)|$. For example, if we have

[1]A Markov matrix is called irreducible if its corresponding graph has only one strongly connected component. Irreducibility is a stronger condition, since all irreducible Markov matrices are unichain, and the Markov chain tree theorem is often stated for the irreducible case.

[2]The adjugate matrix is also known as the classical adjoint matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \ \mathrm{Mi}^{1,2}(A) = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}, \ \text{and} \, \mathrm{Co}^{1,2}(A) = (-1)^3 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} = 6.$$

Before continuing, we will pause briefly to remind the reader of a few elementary properties of the determinant, which will be useful later.

Theorem 3.2. *The determinant of a matrix M can be calculated by the following formula. If M is a 1×1 matrix, its determinant is the single entry. Otherwise, for any row i or column j,*

$$|M| = \sum_{k=1}^{n} M_{i,k} Co^{i,k}(A) = \sum_{k=1}^{n} M_{k,j} Co^{k,j}(A).$$

Note that this formula is recursive, since the cofactor is defined in terms of a determinant. If the columns of M are the vectors $v_1, \ldots, v_n$, we will sometimes write $|M|$ as $\det(v_1, \ldots, v_n)$.

The determinant satisfies the following properties:

- *The determinant is multilinear in the rows and columns of M. For any i, $\det(v_1, \ldots, a(v_i + v_i'), \ldots, v_n) = a(\det(v_1, \ldots, v_i, \ldots, v_n) + \det(v_1, \ldots, v_i', \ldots, v_n))$; that is, the determinant is linear in the i^{th} column, and similarly, the determinant is linear in the j^{th} row for any j.*
- *M is invertible if and only if $|M| \neq 0$ if and only if the nullspace of M is not empty.*
- *If any two rows or any two columns of M are the equal, then $|M| = 0$.*
- *The determinant is unaffected by permutations. That is, given some permutation σ of the indices $\{1, \ldots, n\}$, let $\sigma(M)$ be the result of reordering the rows and columns of M so that the elements of the i^{th} row are now in the $\sigma(i)^{th}$ row, and the same holds for columns. Then $|\sigma(M)| = |M|$.*
- *The determinant of an upper-triangular matrix is the product of the diagonal entries. The determinant of a block upper-triangular matrix is the product of the determinants of the diagonal blocks.*

The adjugate matrix $\mathrm{adj}(A)$ is defined by $(\mathrm{adj}(A))_{i,j} = \mathrm{Co}^{j,i}(A)$. In the example above,

$$\mathrm{adj}(M) = \begin{pmatrix} -3 & 6 & -3 \\ 6 & -12 & 6 \\ -3 & 6 & -3 \end{pmatrix}.$$

The entry $(\mathrm{adj}(A))_{2,1}$ is $\mathrm{Co}^{1,2}(A) = 6$, as we computed above.

We recall the following fact from linear algebra.

Lemma 3.3. *For any $n \times n$ matrix, A,*

$$A \, adj(A) = adj(A) \, A = \begin{pmatrix} |A| & 0 & \cdots & 0 \\ 0 & |A| & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & |A| \end{pmatrix}.$$

Proof. We will compute $(A\,\mathrm{adj}(A))_{i,j}$. If $i = j$, we have

$$\begin{aligned}(A\,\mathrm{adj}(A))_{i,j} &= \sum_{k=1}^{n} A_{i,k}(\mathrm{adj}(A))_{k,j} \\ &= \sum_{k=1}^{n} A_{i,k}(\mathrm{Co}^{j,k}(A)) \\ &= \sum_{k=1}^{n} A_{i,k}(\mathrm{Co}^{i,k}(A)) \\ &= |A|,\end{aligned}$$

by the usual formula for the determinant expanded along row i.

On the other hand, if $i \neq j$, let $\overline{A}$ be the matrix obtained by replacing the j^{th} row of A with a copy of the i^{th} row. Since two rows of $\overline{A}$ are equal, $|\overline{A}| = 0$.

Now computing the determinant along row j, $|\overline{A}| = \sum_{k=1}^{n} \overline{A}_{j,k}\mathrm{Co}^{j,k}(\overline{A})$. But $\overline{A}_{j,k} = A_{i,k}$, and since the j^{th} row of $\overline{A}$ is deleted when computing $\mathrm{Co}^{j,k}(\overline{A})$, $\mathrm{Co}^{j,k}(\overline{A}) = \mathrm{Co}^{j,k}(A) = (\mathrm{adj}(A))_{k,j}$. So $0 = |\overline{A}| = \sum_{k=1}^{n} A_{i,k}(\mathrm{adj}(A))_{k,j} = (A\,\mathrm{adj}(A))_{i,j}$.

These computations show that $A\,\mathrm{adj}(A)$ is $|A|$ along the diagonal and 0 elsewhere, as required. The same argument holds for $\mathrm{adj}(A)A$, except that we expand the determinant formula along columns instead of rows. □

Theorem 3.4. *Given a unichain Markov matrix M with laplacian $\Lambda = M - I$, the vector v_M whose entries are the diagonal entries of the adjugate of Λ, $(v_M)_i = (adj(\Lambda))_{i,i}$, is a stable vector of M. That is, $Mv_M = v_M$.*

Proof. Consider $\Lambda\mathrm{adj}(\Lambda)$. By Lemma 3.3, the off-diagonal entries of the product are 0, and the diagonal entries are $|\Lambda|$. But by Theorem 3.1, $\dim(\ker(\Lambda)) = 1$, so $|\Lambda| = 0$. Thus the product is the zero matrix, and every column of $\mathrm{adj}(\Lambda)$ is in $\ker(\Lambda)$.

The columns of Λ sum to 0, so the rows of Λ^T sum to 0, and thus

$$\Lambda^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

so letting J be the vector consisting of all 1s, $J \in \ker(\Lambda^T)$. But since Λ is a square matrix, $\dim(\ker(\Lambda^T)) = \dim(\ker(\Lambda)) = 1$, so every vector in $\ker(\Lambda^T)$ is a multiple of J.

Now consider $\mathrm{adj}(\Lambda)\Lambda$. Again, this product is the zero matrix, so $(\mathrm{adj}(\Lambda)\Lambda)^T = \Lambda^T\mathrm{adj}(\Lambda)^T = 0$, and every column of $\mathrm{adj}(\Lambda)^T$ is in $\ker(\Lambda^T)$. Thus every column of $\mathrm{adj}(\Lambda)^T$ is a multiple of J, so each row of $\mathrm{adj}(\Lambda)^T$ contains only a single value.

But this means that all column vectors of $\mathrm{adj}(\Lambda)$ are equal, and they are all equal to the vector v_M given by the diagonal entries. Thus $v_M \in \ker(\Lambda)$, so v_M is a stable vector of M. □

As an example of this theorem, consider the Markov matrix

$$M = \begin{pmatrix} 0 & 0 & \frac{2}{3} \\ 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$$

with

$$\Lambda = \begin{pmatrix} -1 & 0 & \frac{2}{3} \\ 1 & -\frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{2} & -1 \end{pmatrix}, \text{ and } \operatorname{adj}(\Lambda) = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & 1 & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Note that the columns of $\operatorname{adj}(\Lambda)$ are all the same. The vector v_M of diagonal entries is $(\frac{1}{3}, 1, \frac{1}{2})$, and indeed, we have

$$Mv_M = \begin{pmatrix} 0 & 0 & \frac{2}{3} \\ 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ 1 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ 1 \\ \frac{1}{2} \end{pmatrix} = v_M,$$

so v_M is a stable vector of M.

The Markov Chain Tree Theorem will give us an alternate way of calculating a stable vector of M from the graph $G(M)$. We will prove the theorem by showing that the result of this calculation is a multiple of the vector v_M of diagonal entries of the laplacian.

4. Directed Trees

A graph $G = (V, E)$ is called a *directed tree* rooted at $v \in V$ if and only if G contains a unique walk from each vertex in V to v. The next theorem is a useful alternate characterization of directed trees.

Theorem 4.1. *A graph $G = (V, E)$ is a directed tree rooted at $v \in V$ if and only if*

1. *v has no outgoing edges, while every $u \in V \setminus \{v\}$ has exactly one outgoing edge, and*
2. *G does not contain any cycles.*

Proof. Let G be a directed tree rooted at v. Suppose G contains a cycle which starts and ends at some vertex $u \in V$. There exists a walk from u to v, but we can construct a distinct walk from u to v by following the cycle from u back to itself, then taking the original walk from u to v. This contradicts the uniqueness of the walk, so G contains no cycles.

If v had an outgoing edge (v, u), then the walk consisting of this edge followed by the unique walk from u to v would constitute a cycle starting and ending at v. But we have already established that G contains no cycles, so v has no outgoing edges.

Let u be a vertex other than v. There is a walk from u to v, so u has at least one outgoing edge. Suppose u had two outgoing edges, (u, w_1) and (u, w_2). Concatenating these edges to the walks from w_1 to v and from w_2 to v respectively, we could construct two distinct walks from u to v. Thus, u has exactly one outgoing edge.

Conversely, let G be a graph containing no cycles, in which one vertex, v, has no outgoing edges, while every $u \in V \setminus \{v\}$ has exactly one outgoing edge. The SCCs of G each contain exactly one vertex, since if there were a pair of vertices $u \sim w$ in an SCC, there would a cycle containing them both made up of the walk from u to w concatenated to the walk from w to u.

Since every vertex but v possesses an outgoing edge, G is unichain with unique closed class, $\{v\}$. By Lemma 2.1, for any vertex $u \in V \setminus \{v\}$, G contains a walk from u to v. If there were more than one such walk, the first vertex at which the walks diverged would have two outgoing edges, contradicting our assumption. Thus, the walk is unique, and G is a directed tree rooted at v. □

The two properties in the statement of Theorem 4.1 give rise two two sets of graphs which will be of interest to us. Fix $n > 0$, the number of vertices, and consider the vertex set $V_n = \{1, 2, \ldots, n\}$.

We will denote by $\mathcal{D}_i$ the set of graphs with vertex set V_n which have property 1 for the vertex v_i. That is, v_i has no outgoing edges, while every other vertex v_j has exactly one outgoing edge.

We will denote by $\mathcal{T}_i$ the subset of $\mathcal{D}_i$ consisting of graphs which also have property 2: they contain no cycles. Theorem 4.1 tells us that $\mathcal{T}_i$ is exactly the set of directed trees with vertex set V_n rooted at v_i.

For now we will work in $\mathcal{D}_i$ for greater generality. Let $D = (V_n, E) \in \mathcal{D}_i$. Each vertex other than v_i has exactly one outgoing edge, so we can define a function which describes E completely. Let $\text{map}(D) : V_n \setminus \{i\} \to V_n$ be the function $\text{map}(D)(j) = k$ if $(j, k) \in E$.

Similarly, to any $D = (V_n, E) \in \mathcal{D}_i$, we can associate an $n \times n$ matrix $\text{mat}(D)$, defined by $\text{mat}(D)_{k,j} = 1$ if $(j, k) \in E$ and $\text{mat}(D)_{j,k} = 0$ otherwise. Note that for $j \neq i$, if $\text{map}(D)(j) = k$, then the j^{th} column of $\text{mat}(D)$ is the k^{th} standard basis vector e_k. The i^{th} column is the zero vector. Thus if $\sigma = \text{map}(D)$, then $\text{mat}(D) = (e_{\sigma(1)}, \ldots, 0, \ldots, e_{\sigma(n)})$.

Suppose $d : E \to \mathbb{R}$ is a weight function on the edge set of a graph $D \in \mathcal{D}_i$. Since there is a unique edge out of j for $j \neq i$, letting $\sigma = \text{map}(D)$, we can express the weight of D in terms of σ in the following way:

$$||D||_d = \prod_{(j,k) \in E} d(j, k) = \prod_{j \neq i} d(j, \sigma(j)).$$

If M is an $n \times n$ matrix, we define a weight function d_M on D by letting $d_M(j, k) = M_{k,j}$. We have

$$||D||_{d_M} = \prod_{(j,k) \in E} M_{k,j} = \prod_{j \neq i} M_{\sigma(j),j}.$$

Lemma 4.2. *Let M be a Markov matrix, and let $D \in \mathcal{D}_i$. Then $||D||_{d_M} \neq 0$ if and only if D is a subgraph of $G(M)$.*

Proof. By the definition of the Markov graph $G(M)$, an edge (j, k) is in $G(M)$ if and only if the corresponding matrix entry $M_{k,j}$ is nonzero. Thus, if an edge (j, k) in the graph D is also an edge in the graph $G(M)$ associated to M, then $d_M(j, k) = M_{k,j} > 0$. Otherwise, $d_M(j, k) = M_{k,j} = 0$. Hence the product $||D||_{d_M}$ is nonzero if and only if every edge in E is also an edge in $G(M)$. □

Now we turn out attention to directed trees. Given a graph $G = (V, E)$, a *directed spanning tree* (DST) of G is a directed tree $T = (V_T, E_T)$ such that $V_T = V$ and $E_T \subseteq E$. That is, T is a subgraph which spans all vertices of G.

We know from Lemma 2.1 that every graph contains at least one closed class. Our goal is to show that if a graph contains exactly one closed class, that is, if it is unichain, then it contains DSTs rooted at each of the vertices in that class.

Theorem 4.3. *A graph G contains a DST rooted at a vertex v if and only if G is unichain and v is in its closed class*

Proof. Let $G = (V, E)$ be a graph which contains some DST rooted at $v \in V$. Then for every vertex $u \in V$, there is a walk from u to v in G. By Lemma 2.1, G contains at least one closed class $\mathcal{C}$. We must have $v \in \mathcal{C}$, for if not, there would be no walk from $w \in \mathcal{C}$ to v since $\mathcal{C}$ has no outgoing edges. Suppose G contains another closed class, $\mathcal{C}'$. By the same argument, $v \in \mathcal{C}'$, so $\mathcal{C} = \mathcal{C}'$. Thus G is unichain.

Conversely, suppose $G = (V, E)$ is unichain with closed class $\mathcal{C}$, and $v \in \mathcal{C}$. Then by Lemma 2.1, for any vertex $u \in V$, there is a walk from u which terminates at some vertex $w \in \mathcal{C}$. Since $v \in \mathcal{C}$, there is a walk from w to v, so, concatenating these, there is a walk from u to v. Thus we can define a function $l_v : V \to \mathbb{N}$ such that $l(u)$ is the minimum length over all walks from u to v.

We will construct a DST, T, of G by selecting one outgoing edge for each vertex $u \in V \setminus \{v\}$, then demonstrating that T contains no cycles. Consider $\{(u, w_1), \dots, (u, w_d)\}$, the set of edges in E outgoing from u. From this set, there must be at least one (u, w_i) such that $l_v(w_i) = l_v(u) - 1$. In particular, the first edge along any minimum length walk from u to v will have this property. Select this edge, and let E_T be the set of the edges selected in this way for each vertex $u \neq v$. Let $T = (V, E_T)$.

Now the value of l_v decreases by one along each edge in E_T. Suppose that there is a cycle of length $m \geq 1$ in T, $(u_0, u_1, \dots, u_m)$ with $u_m = u_0$. Since l_v decreases by one along each edge of the cycle, we have $l_v(u_i) = l_v(u_0) - i$. But then $l_v(u_0) = l_v(u_m) = l_v(u_0) - m$, a contradiction. Thus T is a graph containing no cycles, in which one vertex, v, has no outgoing edges, while every $u \in V \setminus \{v\}$ has exactly one outgoing edge, so by Theorem 4.1, T is a DST of G. ◻

5. A Proof of the Markov Chain Tree Theorem

For a unichain Markov matrix M, the Markov Chain Tree Theorem is concerned with the sum of the weights of all DSTs of $G(M)$ rooted at a vertex i. We will define a vector w_M whose i^{th} component is this quantity. That is, $(w_M)_i = \sum_T ||T||_{d_M}$ where the sum is taken over all $T \in \mathcal{T}_i$ which are DSTs of $G(M)$.

Theorem 5.1 (Markov Chain Tree Theorem)**.** *If M is a unichain Markov matrix, with w_M defined as above, then w_M is a stable vector of M, and there exists a normalizing factor $c \in \mathbb{R}$ such that $c(w_M)$ is the unique stable distribution of M.*

Our plan is to relate the vector w_M to the vector v_M, which we defined in Section 3 to be the diagonal entries of $\operatorname{adj}(\Lambda)$. Since we proved in Theorem 3.4 that v_M is a stable vector of M, the Markov Chain Tree Theorem will follow.

We will need a simple lemma describing the diagonal entries of the adjugate matrix. For any $n \times n$ matrix A, we will denote by $R_i(A)$ the matrix formed by replacing the i^{th} column of A by the i^{th} standard basis vector, e_i.

Lemma 5.2. *Given an $n \times n$ matrix A, for all i, $(adj(A))_{i,i} = |R_i(A)|$.*

Proof. Computing the determinant along the i^{th} column, we have

$$\begin{aligned} |R_i(A)| &= \sum_{k=1}^{n} (R_i(A))_{k,i} \mathrm{Co}^{k,i}(R_i(A)) \\ &= \sum_{k=1}^{n} (e_i)_k \mathrm{Co}^{k,i}(R_i(A)). \end{aligned}$$

Since e_i is 1 in the i^{th} component and 0 elsewhere, all terms in the sum drop out except $k = i$. Then $|R_i(A)| = \mathrm{Co}^{i,i}(R_i(A))$. Now $\mathrm{Co}^{i,i}(R_i(A)) = \mathrm{Co}^{i,i}(A)$, since we remove the i^{th} column when calculating the cofactor. Hence $|R_i(A)| = \mathrm{Co}^{i,i}(A) = (\mathrm{adj}(A))_{i,i}$. □

The first lemma shows that v_M can be computed in terms of weights of graphs. The proof of the lemma is followed by an example, which the reader may find enlightening.

Lemma 5.3. *For M a unichain Markov matrix with laplacian $\Lambda = M - I$, if v_M is the vector consisting of the diagonal entries of $adj(\Lambda)$, then*

$$(v_M)_i = \sum_{D \in \mathcal{D}_i} ||D||_{d_M} |R_i(mat(D) - I))|$$

Proof. By definition, $(v_M)_i = (\mathrm{adj}(\Lambda))_{i,i} = |R_i(\Lambda)|$ by Lemma 5.2.

Consider the j^{th} column of Λ, $\lambda_j = (\Lambda_{1,j}, \ldots, \Lambda_{n,j})$. Since Λ is a laplacian, each column sums to 0. So we can write $\Lambda_{j,j} = -\sum_{k \neq j} \Lambda_{k,j}$. Let $\overline{e}_{k,l} = e_k - e_l$. If $k \neq l$, this is the vector which is 1 in its k^{th} coordinate, -1 in its l^{th} coordinate, and 0 elsewhere, and $\overline{e}_{k,k}$ is the zero vector. Then we can write $\lambda_j = (\Lambda_{1,j}, \ldots, -\sum_{k \neq j} \Lambda_{k,j}, \ldots, \Lambda_{n,j}) = \sum_{k \neq j} \Lambda_{k,j} \overline{e}_{k,j} = \sum_{k=1}^{n} M_{k,j} \overline{e}_{k,j}$. We can change the entries to entries of M because M and Λ agree off of the diagonal, and every entry on the diagonal, $M_{j,j}$, is multiplied by the zero vector $\overline{e}_{j,j}$.

We will use this expression for the columns of Λ to compute the determinant of $R_i(\Lambda)$. By the multilinearity of the determinant,

$$\begin{aligned} |R_i(\Lambda)| &= \det(\lambda_1, \ldots, e_i, \ldots, \lambda_n) \\ &= \det\left(\sum_{k_1} M_{k_1,1} \overline{e}_{k_1,1}, \ldots, e_i, \ldots, \sum_{k_n} M_{n,k_n} \overline{e}_{k_n,n}\right) \\ &= \sum_{k_1} \cdots \sum_{k_n} \det(M_{k_1,1} \overline{e}_{k_1,1}, \ldots, e_i, \ldots, M_{k_n,n} \overline{e}_{k_n,n}) \\ &= \sum_{k_1} \cdots \sum_{k_n} \prod_{j \neq i} M_{k_i,i} \det(\overline{e}_{k_1,1}, \ldots, e_i, \ldots, \overline{e}_{k_n,n}) \\ &= \sum_{\sigma} \prod_{j \neq i} M_{\sigma(j),j} \det(\overline{e}_{\sigma(1),1}, \ldots, e_i, \ldots, \overline{e}_{\sigma(n),n}), \end{aligned}$$

where each σ represents a choice of values for each of the k_j, $\sigma(j) = k_j$. Since there is no k_i (the i^{th} column is just e_i and is not expressed as a sum), σ is a function which assigns to each $j \neq i$ a value $1 \leq \sigma(j) \leq n$.

The key observation is that these σ are in bijection with the graphs $D \in \mathcal{D}_i$. That is, each $\sigma = \text{map}(D)$ for some $D \in \mathcal{D}_i$. The property of assigning a value to each $j \neq i$ is equivalent to having an edge out of the vertex j for all $j \neq i$.

Thus we can reinterpret the sum over σ as a sum over $D \in \mathcal{D}_i$. Given $\sigma = \text{map}(D)$, the product $\prod_{j \neq i} M_{\sigma(j),j}$ is exactly our expression for $||D||_{d_M}$. Consider $\text{mat}(D) = (e_{\sigma(1)}, \ldots, 0, \ldots, e_{\sigma(n)})$, where the 0 is the i^{th} column. Then $\text{mat}(D) - I = (e_{\sigma(1)} - e_1, \ldots, -e_i, \ldots, e_{\sigma(n)} - e_n)$. Finally, $R_i(\text{mat}(D) - I) = (\overline{e}_{\sigma(1),1}, \ldots, e_i, \ldots, \overline{e}_{\sigma(n),1})$. But this is exactly the matrix whose determinant we take in the sum.

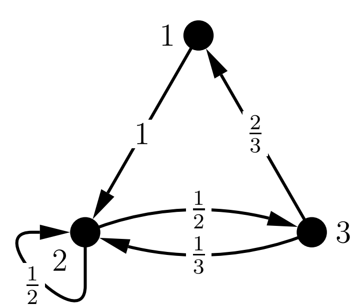
So $|R_i(\Lambda)| = \sum_{D \in \mathcal{D}_i} ||D||_{d_M} |R_i(\text{mat}(D) - I)|$, as was to be shown. $\square$

$$\begin{pmatrix} 0 & 0 & \frac{2}{3} \\ 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$$

M

$$\begin{pmatrix} -1 & 0 & \frac{2}{3} \\ 1 & -\frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{2} & -1 \end{pmatrix}$$

Λ



$G(M)$

FIGURE 4. A running example

We will now pause to give an example, using the Markov matrix M in Figure 4. In Section 3, we calculated the stable distribution $v_M = (\frac{1}{3}, 1, \frac{1}{2})$ for M by taking the diagonal entries of the adjugate matrix $\text{adj}(\Lambda)$.

We will follow the proof of Lemma 5.3 to write the first entry of v_M in terms of the weights of graphs in $\mathcal{D}_1$.

$$
\begin{aligned}
(v_M)_1 &= Co^{1,1}(\Lambda) \\
&= |R_i(\Lambda)| \\
&= \begin{vmatrix} 1 & 0 & \frac{2}{3} \\ 0 & -\frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{2} & -1 \end{vmatrix} \\
&= \det\left(\left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 0 \\ 0 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 0 \\ 0 \\ 0 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 0 \\ -\frac{1}{2} \\ \frac{1}{2} \end{smallmatrix}\right), \left(\begin{smallmatrix} \frac{2}{3} \\ 0 \\ -\frac{2}{3} \end{smallmatrix}\right) + \left(\begin{smallmatrix} 0 \\ \frac{1}{3} \\ -\frac{1}{3} \end{smallmatrix}\right) + \left(\begin{smallmatrix} 0 \\ 0 \\ 0 \end{smallmatrix}\right) \right) \\
&= \det\left(\left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right), 0\left(\begin{smallmatrix} 1 \\ -1 \\ 0 \end{smallmatrix}\right) + 0\left(\begin{smallmatrix} 0 \\ 0 \\ 0 \end{smallmatrix}\right) + \frac{1}{2}\left(\begin{smallmatrix} 0 \\ -1 \\ 1 \end{smallmatrix}\right), \frac{2}{3}\left(\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right) + \frac{1}{3}\left(\begin{smallmatrix} 0 \\ 1 \\ -1 \end{smallmatrix}\right) + 0\left(\begin{smallmatrix} 0 \\ 0 \\ 0 \end{smallmatrix}\right) \right) \\
&= 0 \cdot \frac{2}{3} \begin{vmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix} + 0 \cdot \frac{1}{3} \begin{vmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{vmatrix} + 0 \cdot 0 \begin{vmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{vmatrix} \\
&\quad + 0 \cdot \frac{2}{3} \begin{vmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{vmatrix} + 0 \cdot \frac{1}{3} \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{vmatrix} + 0 \cdot 0 \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} \\
&\quad + \frac{1}{2} \cdot \frac{2}{3} \begin{vmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 1 & -1 \end{vmatrix} + \frac{1}{2} \cdot \frac{1}{3} \begin{vmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{vmatrix} + \frac{1}{2} \cdot 0 \begin{vmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{vmatrix}
\end{aligned}
$$

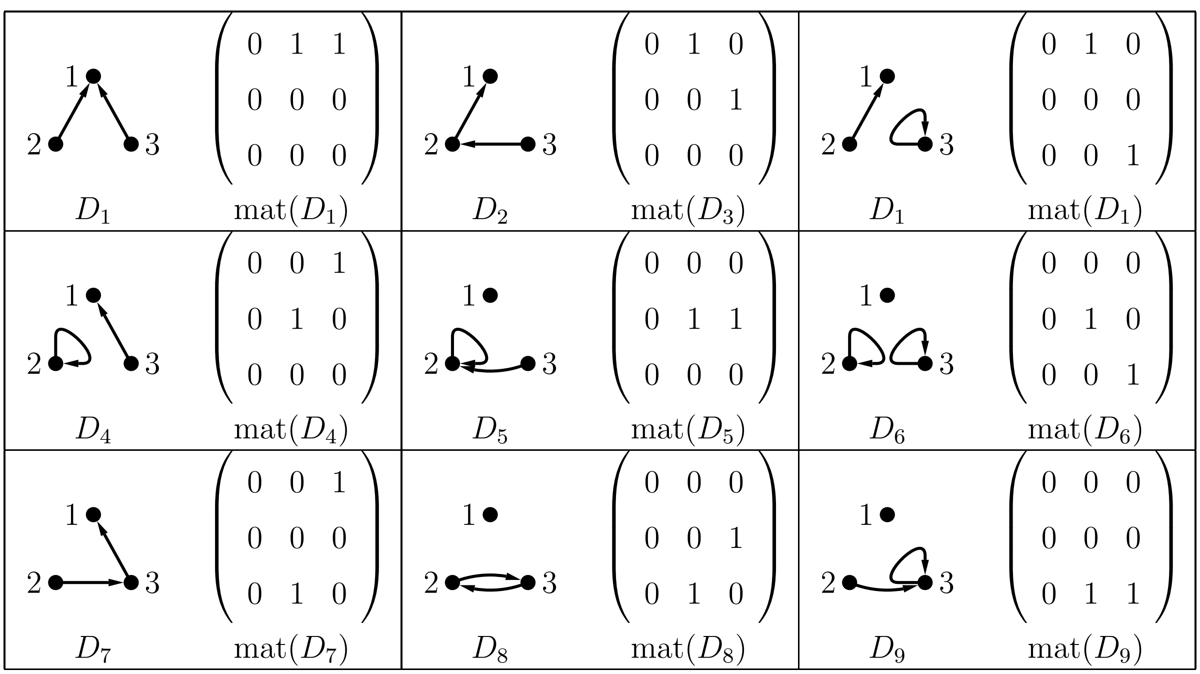


FIGURE 5. $\mathcal{D}_1$ on three vertices

There are nine graphs in $\mathcal{D}_1$ on three vertices (see Figure 5). Taking any of the matrices associated to these graphs, if we subtract the identity and replace the first column with the first standard basis vector, we get a matrix which appears in the expression for $(v_M)_1$ obtained above.

For example,

$$|R_1(\text{mat}(D_1) - I)| = \left| R_1 \begin{pmatrix} -1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \right| = \begin{vmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix},$$

which is the first determinant appearing in the expression. The coefficient of this determinant is the product of the weights of the edges $(2, 1)$ and $(3, 1)$ in D_1 given by $M_{1,2}$ and $M_{1,3}$, $0 \cdot \frac{2}{3}$.

Examining the matrices in Figure 5 and the corresponding determinants in the expression for $(v_M)_1$, we can check that for those j such that D_j is a directed tree, $|R_1(\text{mat}(D_j)-I)| = 1$, and for those j such that D_j contains a cycle, $|R_1(\text{mat}(D_j)-I)| = 0$. The directed trees are D_1, D_2, and D_7. The others all contain cycles.

Thus the the terms in the expression for $(v_M)_i$ corresponding to graphs which are not directed trees drop out. Of the remaining terms, each has a coefficient which is $||D_j||_{d_M}$, the weight of D_j in M. But by Lemma 4.2, $||D_j||_{d_M}$ is nonzero if and only if D_j is a subgraph of $G(M)$. So the terms which do not correspond to DSTs of $G(M)$ also drop out. In our example, D_1 and D_2 contain the edge $(2, 1)$ which is not an edge in $G(M)$, so they drop out.

We are left with a sum over the weights of the DSTs of $G(M)$, which is exactly the expression for $(w_M)_i$. In our example, we are left with the weight of D_7, which is $\frac{1}{3}$.

Our second lemma is the key step that the the determinant of $R_1(\text{mat}(D_j) - I)$ is zero when D_j is not a directed tree. It gives a simple expression for the term $|R_i(\text{mat}(D)-I)|$ in the formula for $(v_M)_i$ given in Lemma 5.3.

Lemma 5.4. *For $D \in \mathcal{D}_i$,*

$$|R_i(mat(D) - I)| = \begin{cases} (-1)^{n-1} & \text{if } D \in \mathcal{T}_i, \text{ i.e. } D \text{ contains no cycles} \\ 0 & \text{otherwise, i.e. } D \text{ contains a cycle} \end{cases}$$

Proof. Suppose $D \in \mathcal{T}_i$. Then D has no cycles, so in particular D has no self-loops, and each diagonal entry (except the i^{th}) of $R_i(\text{mat}(D) - I)$ is -1. The i^{th} diagonal entry is a 1, since the i^{th} column is the standard basis vector e_i.

Now D is a directed tree, so there is a length function $l_D : V_n \to \mathbb{N}$ on the vertices, where $l_D(j)$ is the length of the unique walk from j to i. Sort the vertices according to l_D and call the resulting permutation σ. If $l_D(j) < l_D(k)$, j comes before k in the sorting, and $\sigma(j) < \sigma(k)$. Then permute the vertices of $R_i(\text{mat}(D) - I)$ according to σ. Let N be the resulting matrix. Since determinant is preserved under permutation, $|N| = |R_i(\text{mat}(D) - I)|$.

Now all the nonzero off-diagonal entries of $R_i(\text{mat}(D) - I)$ represent edges in D, and for an edge (j, k) in D, we must have $l_D(k) < l_D(j)$. Thus $\sigma(k) < \sigma(j)$. Now the 1, which was in the k^{th} row and j^{th} column, is in the $\sigma(k)^{th}$ row and $\sigma(j)^{th}$ column in N, and thus is above the diagonal.

Since diagonal entries of a matrix remain on the diagonal after permutation, and all off-diagonal entries of $R_i(\text{mat}(D) - I)$ are placed above the diagonal, N is upper triangular with -1 in all diagonal positions but one. The determinant of an upper triangular matrix is the product of its diagonal entries, so $|R_i(\text{mat}(D) - I)| = |N| = (-1)^{n-1}$.

Now suppose $D \notin \mathcal{T}_i$. Then D contains a cycle. The vertex i is a closed class of D, since it has no outgoing edges. But also the cycle is a closed class, since each vertex in the cycle has exactly one outgoing edge, which goes to the next vertex in the cycle.

Let c be the number of vertices in the cycle. Permute $\text{mat}(D)$ by a permutation σ such that i is sent to 1, the vertices in the cycle are sent to $2, \ldots, c+1$, and the rest of the vertices are sent to $c+2, \ldots, n$. The result is:

$$\sigma(\text{mat}(D)) = \begin{pmatrix} 0 & 0 & * \\ 0 & C & * \\ 0 & 0 & D \end{pmatrix},$$

where C is the square submatrix consisting of rows and columns 2 through $c+1$ and D is the square submatrix consisting of rows and columns $c+2$ through n. The first column is all 0s because i has no outgoing edges, and the entries above and below C are 0s because the cycle has no edges outgoing from the cycle. The contents of the submatrices labeled * do not concern us.

Now $|R_i(\text{mat}(D) - I)| = |\sigma(R_i(\text{mat}(D) - I))| = |R_1(\sigma(\text{mat}(D)) - I)|$, because determinant is preserved under permutation, and under the permutation σ, the standard basis vector e_i in the i^{th} column becomes the standard basis vector e_1 in the 1^{st} column. Now,

$$R_1(\sigma(\text{mat}(D)) - I) = \begin{pmatrix} 1 & 0 & * \\ 0 & C - I & * \\ 0 & 0 & D - I \end{pmatrix},$$

the determinant of which is $|C - I||D - I|$, since the determinant of a block diagonal matrix is the product of the determinants of the diagonal blocks.

Notice that every vertex in the cycle has one outgoing edge to another vertex in the cycle, and thus each column of C has exactly one nonzero entry, which is 1. Thus the columns of C all sum to 1, and C is a Markov matrix. So $C - I$ is its laplacian, and we have seen (Theorem 3.1) that the laplacian of a Markov matrix always has determinant 0. Thus $|R_i(\text{mat}(D) - I)| = |C - I||D - I| = 0$. □

We are now prepared to prove the Markov Chain Tree Theorem.

Proof of Theorem 5.1. Putting Lemma 5.3 and Lemma 5.4 together, we have

$$\begin{aligned} (v_M)_i &= \sum_{D \in \mathcal{D}_i} ||D||_{d_M} |R_i(\text{mat}(D) - I)| \\ &= \sum_{D \in \mathcal{T}_i} ||D||_{d_M} (-1)^{n-1} + \sum_{D \in \mathcal{D}_i \setminus \mathcal{T}_i} ||D||_{d_M} \cdot 0 \\ &= (-1)^{n-1} \sum_{D \in \mathcal{T}_i} ||D||_{d_M}. \end{aligned}$$

Now by Lemma 4.2, $||D||_{d_M}$ is nonzero if and only if D is a DST of $G(M)$. Thus the terms corresponding to trees which are not DSTs drop out, and we are left with $(-1)^{n-1} \sum_T ||T||_{d_M}$, where the sum is taken over all $T \in \mathcal{T}_i$ which are DSTs of $G(M)$.

This is $(-1)^{n-1}(w_M)_i$ by definition, so $w_M = (-1)^{n-1}v_M$. Since w_M is a scalar multiple of v_M, and v_M is a stable vector of M by Theorem 3.4, w_M is a stable vector of M.

Now we will show that we can normalize w_M to obtain a stable distribution of M. By the definition of w_M, the entries of w_M are sums of weights of DSTs, which are products of positive edge weights, so the each entry is positive unless the corresponding sum is empty. Since M is unichain, it has a DST rooted at some vertex i in the closed class by Theorem 4.3. Then the sum $(w_M)_i$ is nonempty, so $(w_M)_i > 0$, and $w_M \neq 0$. Hence w_M is a nonzero vector made up of non-negative entries.

Let $c = \sum_{i=1}^{n}(w_M)_i$, and let $\overline{w_M} = \frac{w_M}{c}$. Since $c > 0$, $(\overline{w_M})_i = \frac{(w_M)_i}{c} \geq 0$, and $\sum_{i=1}^{n}(\overline{w_M})_i = \sum_{i=1}^{n}\frac{(w_M)_i}{c} = \frac{\sum_{i=1}^{n}(w_M)_i}{c} = 1$. So $(\overline{w_M})_i$ is a distribution. Furthermore, $\overline{w_M}$ is a scalar multiple of the stable vector w_M, so it is a stable distribution. □

References

[1] V. Anantharam and P. Tsoucas. A proof of the Markov chain tree theorem. *Statist. Probab. Lett.*, 8(2):189–192, 1989.

[2] Andrei Broder. Abstract. Generating random spanning trees. `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.94.9095`.

[3] John G. Kemeny and J. Laurie Snell. *Finite Markov chains*. Springer-Verlag, New York, 1976. Reprinting of the 1960 original, Undergraduate Texts in Mathematics.

[4] John Wicks. *An Algorithm to Compute the Stochastically Stable Distribution of a Perturbed Markov Matrix*. PhD thesis, Brown University, 2008. `http://www.cs.brown.edu/~jwicks/diss.pdf`.