Counting trees with random walks

Giulio Iacobelli\textsuperscript{a,}\textsuperscript{*}, Daniel R. Figueiredo\textsuperscript{b}, Valmir C. Barbosa\textsuperscript{b}

\textsuperscript{a}Instituto de Matemática, Federal University of Rio de Janeiro, Caixa Postal 68530, 21941-909, Rio de Janeiro - RJ, Brazil
\textsuperscript{b}Systems Engineering and Computer Science Program, COPPE, Federal University of Rio de Janeiro, Caixa Postal 68511, 21941-972 Rio de Janeiro - RJ, Brazil

Received 10 June 2017

Abstract

We give a simple proof of Tutte’s matrix-tree theorem, a well-known result providing a closed-form expression for the number of rooted spanning trees in a directed graph. Our proof stems from placing a random walk on a directed graph and then applying the Markov chain tree theorem to count trees. The connection between the two theorems is not new, but it appears that only one direction of the formal equivalence between them is readily available in the literature. The proof we now provide establishes the other direction. More generally, our approach is another example showing that random walks can serve as a powerful glue between graph theory and Markov chain theory, allowing formal statements from one side to be carried over to the other.

© 2018 Elsevier GmbH. All rights reserved.

MSC: primary 05C50; secondary 60G50; 05C20; 05C05 ; 60J10

Keywords: Spanning trees; Random walks; Matrix-tree theorem; Markov chain tree theorem

1. Introduction

The 170-year-old result known as the Kirchhoff matrix-tree theorem [8] is one of the earliest known links between the theory of graphs and the theory of matrices, and as such...
foreshadows by well over a century the modern spectral theory of graphs. Given a finite connected undirected graph, Kirchhoff’s result allows the number of spanning trees in the graph to be given in terms of any of the minors of its Laplacian matrix via a closed-form expression. Over the many decades that followed its publication in 1847, the Kirchhoff theorem was succeeded by two related results.

The first of these came in 1860 (by Borchardt) and in 1889 (by Cayley), and in the end seems to have become better known than its predecessor. It explores the particular case in which the graph is complete, yielding the well-known Cayley’s formula for the graph’s number of spanning trees \((n^{n-2}\) for \(n\) vertices; see [7] for various proofs of this result). The second came considerably later in 1948, after about a century since Kirchhoff’s theorem, and is known as Tutte’s matrix-tree theorem [17]. Tutte’s theorem extends Kirchhoff’s result to directed graphs and states that the number of spanning trees rooted at vertex \(i\) is equal to the \((i, i)\) minor of the graph’s outgoing Laplacian (see [5] and references therein).

In more recent years, a seemingly independent line of development led to what has become known as the Markov chain tree theorem [3,12], a prominent result allowing the stationary distribution of a finite discrete-time Markov chain to be computed in terms of the spanning trees in the graph associated with the chain. Specifically, it states that the stationary probability of state \(i\) is proportional to a certain weight associated with the set of all spanning trees rooted at \(i\) (see also Chapter 9 in [2], Chapter 4 in [14], and Chapter 6 in [6]). This alone should serve to dispel any impression of independence between this theorem and the descent of Kirchhoff’s result, since the explicit reference to the set of all rooted spanning trees in both Tutte’s matrix-tree theorem and the Markov chain tree theorem clearly suggests a possible connection between the two.

In fact, a formal link between the two theorems is explicitly given in [12], where the authors show that a weighted version of Tutte’s matrix-tree theorem implies the Markov chain tree theorem. The converse implication, however, has to the best of our knowledge not been published anywhere, though curiously, the close tie between the two theorems has, at least to a limited extent, been explicitly acknowledged (see, for example, the notes to Chapter 9 in [2], where the authors mention that Tutte’s matrix-tree theorem and the Markov chain tree theorem can be viewed as part of the “same circle of ideas”). In this paper we remedy this situation by providing a proof that the Markov chain tree theorem implies the same weighted version of Tutte’s matrix-tree theorem as above. The approach we follow converts a weighted directed graph into a Markov chain by placing a random walk on it, and then uses Markov chain theory to count trees. Our proof is simple and, together with the proof in [12], establishes a formal equivalence between the two theorems.

Before proceeding, we note that random walks have been used as a sort of glue to bridge results between graph theory and Markov chain theory. This has been particularly successful in the context of spanning trees (of graphs) and stationary distributions (of Markov chains). This connection has also inspired the development of beautiful algorithms for generating spanning trees at random, such as Aldous and Broder’s algorithm [1,4] as well as Wilson’s [18], both of which leverage random walks. We refer the reader to Section 2 in [18] for a historical account of these algorithms, and to [14], whose Section 4.4 mentions the Markov chain tree theorem as the reason why randomly selecting a state of a Markov chain in accordance with its stationary distribution is closely related to randomly selecting a spanning tree of a graph. Wilson’s result, in particular, is closely
related to the Tutte matrix-tree theorem. In Section 9.7 in [10] (see also [9]), for example, Kirchhoff’s matrix-tree theorem is proved by computing the probability that Wilson’s algorithm outputs a given spanning tree. Further examples can be found in Section 8.2 in [11]. Finally, in [15] the authors derive both Wilson’s result and the Markov chain tree theorem from computing the distribution of loop-erased random walks at selected hitting times.

2. The two theorems

Both Tutte’s matrix-tree theorem and the Markov chain tree theorem make reference to a directed graph, that is, a graph whose edges are ordered pairs of vertices. We assume throughout that such graphs are finite and strongly connected (so a directed path exists leading from any vertex to any other). The out-degree of vertex \( i \) in a directed graph is the number of edges that lead away from \( i \), that is, the number of vertices \( j \) such that \((i, j)\) is an edge of the graph. We denote the out-degree of vertex \( i \) by \( d_i^+ \) and its set of out-neighbors by \( N_i^+ \). In a directed graph \( G \), a spanning tree rooted at vertex \( i \) is a spanning subgraph of \( G \) (i.e., all of \( G \)’s vertices are present) in which \( i \) has no out-neighbor and every other vertex has exactly one out-neighbor. Moreover, the edges of such a subgraph must be such that a directed path exists from any non-root vertex to the root \( i \). Because \( G \) is strongly connected, it is always possible for the tree’s edges to be as required, and of course the desired directed path from any given vertex to \( i \) is unique. We denote the set of all spanning trees rooted at vertex \( i \) in \( G \) by \( T_i \).

Another trait that is common to the two theorems is that they both refer to a weighted version of a directed graph. This weighting comes in two varieties, first by assigning a positive weight to each edge and then by assigning a positive weight to each tree in \( T_i \) for every \( i \). We denote the weight of edge \((i, j)\) by \( a_{i,j} \) and define the weight of a tree \( \tau \in T_i \) to be

\[
w(\tau) \triangleq \prod_{(i, j) \in E(\tau)} a_{i,j},
\]

where \( E(\tau) \) is the set of edges of \( \tau \). Given this weight for a tree in \( T_i \), the weight of set \( T_i \) itself is simply the sum of all its trees’ weights,

\[
w(T_i) \triangleq \sum_{\tau \in T_i} w(\tau).
\]

Stating the Tutte matrix-tree theorem requires the definition of the outgoing Laplacian of a weighted directed graph. The outgoing Laplacian of graph \( G \), denoted by \( L_i^+ \), is the square matrix \( L_i^+ = \Delta^+ - A \), where \( \Delta^+ \) is the diagonal matrix of entry \( \Delta_{ii}^+ = \sum_{j \in N_i^+} a_{i,j} \) for vertex \( i \) and \( A \) is the adjacency matrix of \( G \) (of entry \( a_{i,j} \) if \( j \in N_i^+ \), 0 otherwise). That is,

\[
L_{i,j}^+ = \begin{cases} 
\sum_{k \in N_i^+} a_{i,k} - a_{i,i} & \text{if } j = i, \\
-a_{i,j} & \text{if } j \neq i \text{ and } j \in N_i^+, \\
0 & \text{otherwise}.
\end{cases}
\]
Deleting the $i$th row and column from $L^+$ yields the matrix that we henceforth denote by $L^+_{[i]}$. We are then ready to state the first of the two theorems.\footnote{A more general version of the theorem, relating an arbitrary minor of the outgoing Laplacian to the enumeration of certain types of forests, is given in [5,12].}

**Theorem 1 (Tutte’s Matrix-Tree Theorem).** Let $G$ be a weighted directed graph. Then

$$w(T_i) = \det(L^+_{[i]}),$$

for every vertex $i$.

Unit weights throughout imply $\sum_{j \in N^+_i} a_{i,j} = d^+_i$. Moreover, by Theorem 1 they allow the spanning trees rooted at vertex $i$ to be counted via $|T_i| = \det(L^+_{[i]}).

In order to state the Markov chain tree theorem, we must first describe how a weighted directed graph $G$ can be obtained from a finite irreducible Markov chain $M$, say of state space $\Omega$ and transition-probability matrix $P = (p(i,j))_{i,j \in \Omega}$. This is achieved by letting the vertex set of $G$ be $\Omega$ and the set $N^+_i$ of out-neighbors of $i$ be $\{ j \in \Omega \mid p(i,j) > 0 \}$. The graph $G$ thus obtained is clearly strongly connected, by virtue of the irreducibility of $M$. We complete the description of $G$ by letting $a_{i,j} = p(i,j)$ for every vertex $i$ and every $j \in N^+_i$. The Markov chain tree theorem, given next, allows the stationary distribution of the Markov chain to be expressed in terms of the weights of the sets of rooted trees, one set per vertex, associated with $G$.\footnote{A more general version of the theorem, dispensing with the need for irreducibility, is given in [3,12].}

This distribution is henceforth denoted by the row vector $\pi = (\pi(i))_{i \in \Omega}$.

**Theorem 2 (Markov Chain Tree Theorem).** Let $G$ be the weighted directed graph associated with the irreducible Markov chain $M$. Then

$$\pi(i) = \frac{w(T_i)}{\sum_{j \in \Omega} w(T_j)},$$

for every $i \in \Omega$.

One curious special case of Theorem 2 is that in which the Markov chain $M$ corresponds to a symmetric random walk on $G$. In this case, we have $p(i,j) = 1/d^+_i$ for every vertex $i$ and every $j \in N^+_i$, so $w(\tau) = (\prod_{k \neq i} d^+_k)^{-1}$ for every $\tau \in T_i$. It then follows from Eq. (1) that $w(T_i) = (\prod_{k \neq i} d^+_k)^{-1} |T_i|$, and by Theorem 2 we have

$$\pi(i) = \frac{(\prod_{k \neq i} d^+_k)^{-1} |T_i|}{\sum_{j \in \Omega} (\prod_{k \neq j} d^+_k)^{-1} |T_j|} = \frac{d^+_i |T_i|}{\sum_{j \in \Omega} d^+_j |T_j|},$$

for every $i \in \Omega$. This expression, referred to in [13] as “folklore”, gives rise to a known combinatorial identity relating the number of spanning trees rooted at vertex $i$ to the numbers of spanning trees rooted at the in-neighbors of $i$ (vertices $j$ such that $i \in N^+_j$).

To see this, simply consider that the vector $\pi$ whose entries are given by Eq. (2) must satisfy the balance equation $\pi = \pi P$ for $M$, or $\sum_j \pi(j) p(j,i) = \pi(i)$ for each $i$. Using Equation (2) in the latter expression, we easily recover the said identity, $\sum_{j|i \in N^+_j} |T_j| = d^+_i |T_i|$ (see, e.g., Proposition 5 in [16]).
3. Tree counts from random walks

Given the weighted directed graph $G$ of the previous section and a positive $w_{i,j}$ for each edge $(i, j)$ in $G$, let $\delta_i^+ \triangleq \sum_{k \in \mathbb{N}^+} w_{i,k}$. We provide a proof of Theorem 1 (for weights $a_{i,j} = w_{i,j}$) from the validity of Theorem 2 (for weights $a_{i,j} = w_{i,j} / \delta_i^+$). We do so by considering the Markov chain $M$ with transition probabilities $p(i, j) = w_{i,j} / \delta_i^+$, which represents the random walk on $G$ whose probabilities for moving out of vertex $i$ are proportional to the $w_{i,j}$s. Clearly, setting $w_{i,j} = 1$ for every edge $(i, j)$ leads to the symmetric random walk.

In general, these two weighting schemes for the edges of graph $G$ lead to distinct values for the weight of each $T$, the set of all trees rooted at vertex $i$. We denote the two possible values by $w_1(T_i)$ and $w_2(T_i)$, depending respectively on whether the weighting scheme of Theorem 1 or that of Theorem 2 is used. Clearly, it holds that

$$w_2(T_i) = \left( \prod_{k \neq i} \delta_k^+ \right)^{-1} w_1(T_i),$$

and by Theorem 2 we have

$$\pi(i) = \frac{\left( \prod_{k \neq i} \delta_k^+ \right)^{-1} w_1(T_i)}{\sum_{j \in \Omega} \left( \prod_{k \neq j} \delta_k^+ \right)^{-1} w_1(T_j)}.$$  \hspace{1cm} (3)

**Proof of Theorem 1.** As the stationary distribution of Markov chain $M$, the $\pi$ given in Eq. (3) must satisfy both the balance equation $\pi = \pi P$ and the normalizing condition $\sum_{i \in \Omega} \pi(i) = 1$. These two can be combined together into a more compact expression, viz.,

$$\pi C_k(I - P) = e_k,$$

where $I$ is the identity matrix and, for arbitrarily chosen $k$, $C_k(I - P)$ is the matrix obtained from $I - P$ by substituting a vector of ones for its $k$th column and $e_k$ is the row vector having 1 in column $k$ and 0 elsewhere. This expression uniquely defines $\pi$ and yields each $\pi(i)$ from Cramer’s rule and well-known properties of determinants [12]:

$$\pi(i) = \frac{\det((I - P)_{\{i\}})}{\det(C_k(I - P))} = \frac{\det((I - P)_{\{i\}})}{\sum_{j \in \Omega} \det((I - P)_{\{j\}})},$$  \hspace{1cm} (4)

where $(I - P)_{\{i\}}$ denotes the matrix obtained from $I - P$ by deleting the $i$th row and column.

Now, recalling that matrix $\Delta^+$ is diagonal with $i$th entry $\delta_i^+$ and that the incidence matrix $A$ has general entry $w_{i,j}$ yields, for the outgoing Laplacian $L^+$ of $G$,

$$L^+ = \Delta^+ - A = \Delta^+(I - (\Delta^+)^{-1}A) = \Delta^+(I - P),$$

and thus,

$$L_{\{i\}}^+ = (\Delta^+(I - P))_{\{i\}} = \Delta_{\{i\}}^+(I - P)_{\{i\}},$$

where the last equality holds since $\Delta^+$ is diagonal. From this we obtain

$$\det(L_{\{i\}}^+) = \det(\Delta_{\{i\}}^+) \det((I - P)_{\{i\}}) = \left( \prod_{k \neq i} \delta_k^+ \right) \det((I - P)_{\{i\}}),$$
so the \((i, i)\) minor of \(\mathbb{I} - P\) can be expressed in terms of the \((i, i)\) minor of \(L^+\):

\[
\det((\mathbb{I} - P)_{(i,i)}) = \left(\prod_{k \neq i} \delta_k^+\right)^{-1} \det(L^+_{(i,i)}),
\]

which by Eq. (4) leads to

\[
\pi(i) = \frac{\left(\prod_{k \neq i} \delta_k^+\right)^{-1} \det(L^+_{(i,i)})}{\sum_{j \in \Omega} \left(\prod_{k \neq j} \delta_k^+\right)^{-1} \det(L^+_{(j,j)})}.
\]

Combining Eq. (5) with Eq. (3) yields

\[
w_1(T_i) = C \det(L^+_{(i,i)}),
\]

where \(C\) is a ratio between those equations’ denominators and as such does not depend on \(i\). Completing the proof, therefore, requires showing that \(C = 1\), which is arguably the most challenging step.

We start by demonstrating that \(C\) is independent not only of the particular vertex \(i\) in question, but also of graph \(G\) itself. To this end, let \(i_1\) be any vertex in \(G\) and modify this graph by letting \(a_{i_1,j} = 1\) for every \(j\). This modification creates edges directed away from \(i_1\) toward vertices that are not out-neighbors of \(i_1\) in \(G\), and may alter the weights of edges leading from \(i_1\) to its out-neighbors in \(G\). If \(G^{(1)}\) denotes the resulting graph, then clearly the set \(T_{i_1}\) of spanning trees rooted at \(i_1\) is the same in both \(G\) and \(G^{(1)}\) (even though, for any \(i \neq i_1\), \(T_i\) may differ between graphs \(G\) and \(G^{(1)}\)). In addition, \(w_1(T_{i_1})\) and \(L^+_{(i_1)}\) remain invariant as well, and consequently so does the value of \(C\), by Eq. (6). Thus, the constant \(C\) when considering \(G\) is identical to \(C\) when considering \(G^{(1)}\).

For \(n\) denoting the number of vertices in \(G\), the above procedure can be iterated for the remaining vertices \(i_2, \ldots, i_n\). The first of these iterations targets the out-neighborhood of \(i_2\) in \(G^{(1)}\) and yields graph \(G^{(2)}\), the second one targets the out-neighborhood of \(i_3\) in \(G^{(2)}\) and yields \(G^{(3)}\), and so on. For \(1 < \ell \leq n\), we again have that both \(w_1(T_{i_\ell})\) and \(L^+_{(i_\ell)}\) are unaltered by the manipulation of the out-neighborhood of \(i_\ell\) in \(G^{(\ell-1)}\) to give rise to \(G^{(\ell)}\). All along, therefore, the value of \(C\) remains unchanged.

Graph \(G^{(n)}\) has antiparallel edges interconnecting all pairs of distinct vertices, so its adjacency matrix is that of a complete undirected graph with unit weights on all edges. For any vertex \(i\), therefore, in \(G^{(n)}\) we have \(w_1(T_i) = |T_i| = n^{n-2}\), the former equality following from the unit weights and the latter from Cayley’s formula for the complete undirected graph. On the other hand, matrix \(L^+\) is fully symmetric for graph \(G^{(n)}\), so its \((i, i)\) minor (which in this case does not depend on \(i\)) can be easily computed by diagonalizing \(L^+_{(i,i)}\). Indeed, we can write \(L^+_{(i,i)} = n\mathbb{I}_{n-1} - B_{n-1}\), where \(\mathbb{I}_{n-1}\) is the \((n-1) \times (n-1)\) identity matrix and \(B_{n-1}\) is the \((n-1) \times (n-1)\) matrix having all entries equal to 1. Matrix \(B_{n-1}\) is diagonalizable to \(\text{diag}(n-1, 0, \ldots, 0)\), whence it follows that the diagonal form of \(L^+_{(i,i)}\) is \(\text{diag}(1, n, \ldots, n)\). Therefore, \(\det(L^+_{(i,i)}) = n^{n-2}\), and consequently, \(C = 1\).

**Acknowledgments**

This work has been partially funded by CAPES, CNPq, and FAPERJ BBP grants.
References


