5.6 Oriented Trees and the Matrix–Tree Theorem

A famous problem that goes back to Euler asks for what graphs G there is a closed walk that uses every edge exactly once. (There is also a version for non-closed walks.) Such a walk is called an *Eulerian tour* (also known as an *Eulerian cycle*). A graph which has an Eulerian tour is called an *Eulerian graph*. Euler's famous theorem (the first real theorem of graph theory) states that G is Eulerian if and only if it is connected (except for isolated vertices) and every vertex has even degree. Here we will be concerned with the analogous theorem for directed graphs D. We want to know not just whether an Eulerian tour exists, but how many there are. We reduce this problem to that of counting certain subtrees of D called *oriented trees*. We will prove an elegant determinantal formula for this number, and from it derive a determinantal formula, known as the *Matrix–Tree Theorem*, for the number of spanning trees of any (undirected) graph. An application of the enumeration of Eulerian tours is given to the enumeration of de Bruijn sequences. For the case of undirected graphs no analogous formula is known for the number of Eulerian tours, explaining why we consider only the directed case.

We will use the terminology and notation associated with directed graphs introduced at the beginning of Section 4.7. Let $D=(V, E, \varphi)$ be a digraph with vertex set $V=\{v_1, \ldots, v_p\}$ and edge set $E=\{e_1, \ldots, e_q\}$. We say that D is *connected* if it is connected as an undirected graph. A *tour* in D is a sequence $e_1, e_2, \ldots, e_r$ of *distinct* edges such that the final vertex of e_i is the initial vertex of e_{i+1} for all $1 \le i \le r-1$, and the final vertex of e_r is the initial vertex of e_1. A tour is *Eulerian* if every edge of D occurs at least once (and hence exactly once). A digraph that has no isolated vertices and contains an Eulerian tour is called an *Eulerian digraph*. Clearly an Eulerian digraph is connected. (Even more strongly, there is a directed path between any pair of vertices.) The *outdegree* of a vertex v, denoted outdeg(v), is the number of edges of G with initial vertex v. Similarly the *indegree* of v, denoted indeg(v), is the number of edges of D with final vertex v. A loop (edge of the form (v, v)) contributes one to both the indegree and outdegree. A digraph is *balanced* if indeg(v) = outdeg(v) for all vertices v.

5.6.1 Theorem. *A digraph D without isolated vertices is Eulerian if and only if it is connected and balanced.*

Proof. Assume D is Eulerian, and let $e_1, \ldots, e_q$ be an Eulerian tour. As we move along the tour, whenever we enter a vertex v we must exit it, except that at the very end we enter the final vertex v of e_q without exiting it. However, at the beginning we exited v without having entered it. Hence every vertex is entered as often as it is exited and so must have the same outdegree as indegree. Therefore D is balanced, and as noted above D is clearly connected.

Now assume that D is balanced and connected. We may assume that D has at least one edge. We first claim that for any edge e of D, D has a tour (not necessarily Eulerian) for which $e=e_1$. If e_1 is a loop we are done. Otherwise we have entered

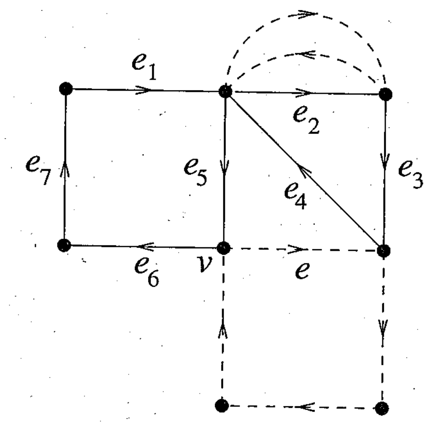


Figure 5-16. A nonmaximal tour in a balanced digraph.

the vertex fin(e_1) for the first time, so since D is balanced there is some exit edge e_2. Either fin(e_2) = init(e_1) and we are done, or else we have entered the vertex fin(e_2) once more than we have exited it. Since D is balanced there is a new edge e_3 with fin(e_2) = init(e_3). Continuing in this way, either we complete a tour or else we have entered the current vertex once more than we have exited it, in which case we can exit along a new edge. Since D has finitely many edges, eventually we must complete a tour. Thus D does have a tour for which $e=e_1$.

Now let $e_1, \ldots, e_r$ be a tour C of maximum length. We must show that $r=q$, the number of edges of D. Assume to the contrary that $r<q$. Since in moving along C every vertex is entered as often as it is exited (with init(e_1) exited at the beginning and entered at the end), when we remove the edges of C from D we obtain a digraph H that is still balanced, though it need not be connected. However, since D is connected, at least one connected component H_1 of H contains at least one edge and has a vertex v in common with C. Since H_1 is balanced, there is an edge e of H_1 with initial vertex v. See Figure 5-16, where the edges of a tour C are drawn as solid lines, and the remaining edges as dotted lines. The argument of the previous paragraph shows that H_1 has a tour C' of positive length beginning with the edge e. But then when moving along C, when we reach v we can take the "detour" C' before continuing with C. This gives a tour of length longer than r, a contradiction. Hence $r=q$, and the theorem is proved. □

Our primary goal is to count the number of Eulerian tours of a connected balanced digraph. A key concept in doing so is that of an oriented tree. An *oriented tree* with root v is a (finite) digraph T with v as one of its vertices, such that there is a unique directed path from any vertex u to v. In other words, for every vertex u there is a unique sequence of edges $e_1, \ldots, e_r$ such that (a) init(e_1) = u,

(b) $\text{fin}(e_r) = v$, and (c) $\text{fin}(e_i) = \text{init}(e_{i+1})$ for $1 \le i \le r - 1$. It is easy to see that this means that the underlying undirected graph (i.e., "erase" all the arrows from the edges of T) is a tree, and that all arrows in T "point toward" v. There is a surprising connection between Eulerian tours and oriented trees, given by the next result.

5.6.2 Theorem. *Let D be a connected balanced digraph with vertex set V. Fix an edge e of D, and let $v = \text{init}(e)$. Let $\tau(D, v)$ denote the number of oriented (spanning) subtrees of D with root v, and let $\epsilon(D, e)$ denote the number of Eulerian tours of D starting with the edge e. Then*

$$\epsilon(D, e) = \tau(D, v) \prod_{u \in V} (\text{outdeg}(u) - 1)!. \tag{5.82}$$

Proof. Let $e = e_1, e_2, \ldots, e_q$ be an Eulerian tour E in D. For each vertex $u \neq v$, let $e(u)$ be the *last exit* from u in the tour, i.e., let $e(u) = e_j$ where $\text{init}(e_j) = u$ and $\text{init}(e_k) \neq u$ for any $k > j$.

Claim 1. The vertices of D, together with the edges $e(u)$ for all vertices $u \neq v$, form an oriented subtree of D with root v.

Proof of Claim 1. This is a straightforward verification. Let T be the spanning subgraph of D with edges $e(u)$, $u \neq v$. Thus if $\#V = p$, then T has p vertices and $p - 1$ edges. We now make the following three observations:

(a) T does not have two edges f and f' satisfying $\text{init}(f) = \text{init}(f')$. This is clear, since both f and f' can't be last exits from the same vertex.
(b) T does not have an edge f with $\text{init}(f) = v$. This is clear, since by definition the edges of T consist only of last exits from vertices other than v, so no edge of T can exit from v.
(c) T does not have a (directed) cycle C. For suppose C were such a cycle. Let f be that edge of C which occurs after all the other edges of C in the Eulerian tour E. Let f' be the edge of C satisfying $\text{fin}(f) = \text{init}(f')$ $(= u$, say). We can't have $u = v$ by (b). Thus when we enter u via f, we must exit u. We can't exit u via f', since f occurs after f', in E. (Note that we cannot have $f = f'$, since then f would be a loop and therefore not a last exit.) Hence f' is not the last exit from u, contradicting the definition of T.

It is easy to see that conditions (a)–(c) imply that T is an oriented tree with root v, proving the claim.

Claim 2. We claim that the following converse to Claim 1 is true. Given a connected balanced digraph D and a vertex v, let T be an oriented (spanning) subtree of D with root v. Then we can construct an Eulerian tour Δ as follows. Choose an edge e_1 with $\text{init}(e_1) = v$. Then continue to choose any edge possible to continue the tour, except we never choose an edge f of T unless we have to, i.e., unless it's the only remaining edge exiting the vertex at which we stand. Then we never get stuck until all edges are used, so we have constructed an Eulerian tour Δ. Moreover, the set of last exits of Δ from vertices $u \neq v$ of D coincides with the set of edges of the oriented tree T.

Proof of Claim 2. Since D is balanced, the only way to get stuck is to end up at v with no further exits available, but with an edge still unused. Suppose this is the case. At least one unused edge must be a last exit edge, i.e., an edge of T. Let u be a vertex of T closest to v in T such that the unique edge f of T with $\text{init}(f) = u$ is not in the tour. Let $y = \text{fin}(f)$. Suppose $y \neq v$. Since we enter y as often as we leave it, we don't use the last exit from y. Thus $y = v$. But then we can leave v, a contradiction. This proves Claim 2.

We have shown that every Eulerian tour Δ beginning with the edge e has associated with it a last-exit oriented subtree $T = T(\Delta)$ with root $v = \text{init}(e)$. Conversely, we have also shown that given an oriented subtree T with root v, we can obtain all Eulerian tours Δ beginning with e and satisfying $T = T(\Delta)$ by choosing for each vertex $u \neq v$ the order in which the edges from u, except the edge of T, appear in Δ. Thus for each vertex u we have $(\text{outdeg}(u) - 1)!$ choices, so for each T we have $\prod_u (\text{outdeg}(u) - 1)!$ choices. Since there are $\tau(D, v)$ choices for T, the proof follows. □

5.6.3 Corollary. *Let D be a connected balanced digraph, and let v be a vertex of D. Then the number $\tau(D, v)$ of oriented subtrees with root v is independent of v.*

Proof. Let e be an edge with initial vertex v. By equation (5.82), we need to show that the number $\epsilon(G, e)$ of Eulerian tours beginning with e is independent of e. But $e_1 e_2 \cdots e_q$ is an Eulerian tour if and only if $e_i e_{i+1} \cdots e_q e_1 e_2 \cdots e_{i-1}$ is also an Eulerian tour, and the proof follows. □

In order for Theorem 5.6.2 to be of use, we need a formula for $\tau(G, v)$. To this end, define the *Laplacian matrix* $\mathbf{L} = \mathbf{L}(D)$ of a directed graph D with vertex set $V = \{v_1, \ldots, v_p\}$ to be the $p \times p$ matrix

$$\mathbf{L}_{ij} = \begin{cases} -m_{ij} & \text{if } i \neq j \text{ and there are } m_{ij} \text{ edges with} \\ & \quad \text{initial vertex } v_i \text{ and final vertex } v_j \\ \text{outdeg}(v_i) - m_{ii} & \text{if } i = j. \end{cases}$$

Note that the diagonal entry $\text{outdeg}(v_i) - m_{ii}$ is just the number of *nonloop* edges of D with initial vertex v_i. Hence the Laplacian matrix $\mathbf{L}(D)$ is independent of the loops of D. Note also that if every vertex of D has the same outdegree d, then the adjacency matrix $\mathbf{A}$ (defined in Section 4.7) and Laplacian matrix $\mathbf{L}$ of D are related by $\mathbf{L}=d\mathbf{I}-\mathbf{A}$, where $\mathbf{I}$ denotes the $p \times p$ identity matrix. In particular, if $\mathbf{A}$ has eigenvalues $\lambda_1, \ldots, \lambda_p$, then $\mathbf{L}$ has eigenvalues $d-\lambda_1, \ldots, d-\lambda_p$.

5.6.4 Theorem. *Let D be a loopless digraph with vertex set $V=\{v_1, \ldots, v_p\}$, and let $1 \le k \le p$. Let $\mathbf{L}$ be the Laplacian matrix of D, and define $\mathbf{L_0}$ to be $\mathbf{L}$ with the k-th row and column deleted. Then*

$$\det \mathbf{L_0} = \tau(D, v_k). \tag{5.83}$$

Proof. Induction on q, the number of edges of D. First note that the theorem is true if D is not connected, since clearly $\tau(D, v_k)=0$, while if D_1 is the component of D containing v_k and D_2 is the rest of D, then $\det \mathbf{L_0}(D)=\det \mathbf{L_0}(D_1)\cdot \det \mathbf{L}(D_2)=0$. The least number of edges that D can have is $p-1$ (since D is connected). Suppose then that D has $p-1$ edges, so that as an undirected graph D is a tree. If D is not an oriented tree with root v_k, then some vertex $v_i \neq v_k$ of D has outdegree 0. Then $\mathbf{L_0}$ has a zero row, so $\det \mathbf{L_0}=0=\tau(D, v_k)$. If on the other hand D is an oriented tree with root v_k, then there is an ordering of the set $V-\{v_k\}$ so that $\mathbf{L_0}$ is upper triangular with 1's on the main diagonal. Hence $\det \mathbf{L_0}=1=\tau(D, v_k)$.

Now suppose that D has $q > p-1$ edges, and assume the theorem for digraphs with at most $q-1$ edges. We may assume that no edge f of D has initial vertex v_k, since such an edge belongs to no oriented tree with root v_k and also makes no contribution to $\mathbf{L_0}$. It then follows, since D has at least p edges, that there exists a vertex $u \neq v_k$ of D of outdegree at least two. Let e be an edge with $\mathrm{init}(e)=u$. Let D_1 be D with the edge e removed. Let D_2 be D with all edges e' removed such that $\mathrm{init}(e)=\mathrm{init}(e')$ and $e' \neq e$. (Note that D_2 is strictly smaller than D, since $\mathrm{outdeg}(u) \ge 2$.) By induction, we have $\det \mathbf{L_0}(D_1)=\tau(D_1, v_k)$ and $\det \mathbf{L_0}(D_2)=\tau(D_2, v_k)$. Clearly $\tau(D, v_k)=\tau(D_1, v_k)+\tau(D_2, v_k)$, since in an oriented tree T with root v_k there is exactly one edge whose initial vertex coincides with that of e. On the other hand, it follows immediately from the multilinearity of the determinant that

$$\det \mathbf{L_0}(D) = \det \mathbf{L_0}(D_1) + \det \mathbf{L_0}(D_2).$$

From this the proof follows by induction. □

The operation of removing a row and column from $\mathbf{L}(D)$ may seem somewhat contrived. In the case when D is balanced (so $\tau(D, v)$ is independent of v), we would prefer a description of $\tau(D, v)$ directly in terms of $\mathbf{L}(D)$. Such a description will follow from the next lemma.

5.6.5 Lemma. *Let $\mathbf{M}$ be a $p \times p$ matrix (with entries in a field) such that the sum of the entries in every row and column is 0. Let $\mathbf{M_0}$ be the matrix obtained*

from $\mathbf{M}$ by removing the i-th row and j-th column. Then the coefficient of x in the characteristic polynomial $\det(\mathbf{M}-x\mathbf{I})$ of $\mathbf{M}$ is equal to $(-1)^{i+j+1}p\cdot\det(\mathbf{M_0})$. (Moreover, the constant term of $\det(\mathbf{M}-x\mathbf{I})$ is 0.)

Proof. The constant term of $\det(\mathbf{M}-x\mathbf{I})$ is $\det(\mathbf{M})$, which is 0 because the rows of $\mathbf{M}$ sum to 0.

For definiteness we prove the rest of the lemma only for removing the last row and column, though the proof works just as well for any row and column. Add all the rows of $\mathbf{M}-x\mathbf{I}$ except the last row to the last row. This doesn't effect the determinant, and will change the entries of the last row all to $-x$ (since the rows of $\mathbf{M}$ sum to 0). Factor out $-x$ from the last row, yielding a matrix $\mathbf{N}(x)$ satisfying $\det(\mathbf{M}-x\mathbf{I})=-x\det \mathbf{N}(x)$. Hence the coefficient of x in $\det(\mathbf{M}-x\mathbf{I})$ is given by $-\det \mathbf{N}(0)$. Now add all the columns of $\mathbf{N}(0)$ except the last column to the last column. This does not effect $\det \mathbf{N}(0)$. Because the columns of $\mathbf{M}$ sum to 0, the last column of $\mathbf{N}(0)$ becomes the column vector $[0, 0, \ldots, 0, p]^t$. Expanding the determinant by the last column shows that $\det \mathbf{N}(0)=p\cdot\det \mathbf{M_0}$, and the proof follows. □

Suppose that the eigenvalues of the matrix $\mathbf{M}$ of Lemma 5.6.5 are equal to $\mu_1, \ldots, \mu_p$ with $\mu_p=0$. Since $\det(\mathbf{M}-x\mathbf{I})=-x\prod_{j=1}^{p-1}(\mu_j-x)$, we see that

$$(-1)^{i+j+1}p\cdot\det \mathbf{M_0} = -\mu_1\cdots\mu_{p-1}. \tag{5.84}$$

This equation allows Theorem 5.6.4, in the case of balanced digraphs, to be restated as follows.

5.6.6 Corollary. *Let D be a balanced digraph with p vertices and with Laplacian matrix $\mathbf{L}$. Suppose that the eigenvalues of $\mathbf{L}$ are $\mu_1, \ldots, \mu_p$ with $\mu_p=0$. Then for any vertex v of D,*

$$\tau(D, v) = \frac{1}{p}\mu_1\cdots\mu_{p-1}.$$

Combining Theorems 5.6.2 and 5.6.4 yields a formula for the number of Eulerian tours in a balanced digraph.

5.6.7 Corollary. *Let D be a connected balanced digraph with p vertices. Let e be an edge of D. Then the number $\epsilon(D, e)$ of Eulerian tours of D with first edge e is given by*

$$\epsilon(D, e) = (\det \mathbf{L_0}(D))\prod_{u\in V}(\mathrm{outdeg}(u)-1)!.$$

Equivalently (using Corollary 5.6.6), if $\mathbf{L}(D)$ has eigenvalues $\mu_1, \ldots, \mu_p$ with

$\mu_p = 0$, *then*

$$\epsilon(D, e) = \frac{1}{p}\mu_1 \cdots \mu_{p-1} \prod_{u \in V}(\text{outdeg}(u) - 1)!.$$

Let us consider an important special case of Corollary 5.6.7. The *Laplacian matrix* $\mathbf{L} = \mathbf{L}(G)$ of the *undirected* graph G with vertex set $V = \{v_1, \ldots, v_p\}$ is the $p \times p$ matrix

$$\mathbf{L}_{ij} = \begin{cases} -m_{ij} & \text{if } i \neq j \text{ and there are } m_{ij} \text{ edges between} \\ & \quad \text{vertices } v_i \text{ and } v_j \\ \deg(v_i) - m_{ii} & \text{if } i = j, \end{cases}$$

where $\deg(v_i)$ denotes the degree (number of incident edges) of v_i. Let $\hat{G}$ be the digraph obtained from G by replacing each edge $e = uv$ of G with a pair of directed edges $u \to v$ and $v \to u$. Clearly $\hat{G}$ is balanced, and $\hat{G}$ is connected whenever G is. Choose a vertex v of G. There is an obvious one-to-one correspondence between spanning trees T of G and oriented spanning trees $\hat{T}$ of $\hat{G}$ with root v, namely, direct each edge of T toward v. Moreover, $\mathbf{L}(G) = \mathbf{L}(\hat{G})$. Let $c(G)$ denote the number of spanning trees (or *complexity*) of G. Then as an immediate consequence of Theorem 5.6.4 we obtain the following determinantal formula for $c(G)$. This formula is known as the *Matrix–Tree Theorem.*

5.6.8 Theorem (The Matrix–Tree Theorem). *Let G be a finite connected p-vertex graph without loops, with Laplacian matrix $\mathbf{L} = \mathbf{L}(G)$. Let $1 \leq i \leq p$, and let $\mathbf{L}_0$ denote $\mathbf{L}$ with the i-th row and column removed. Then*

$$c(G) = \det \mathbf{L}_0.$$

Equivalently, if $\mathbf{L}$ has eigenvalues $\mu_1, \ldots, \mu_p$ with $\mu_p = 0$, then

$$c(G) = \frac{1}{p}\mu_1 \cdots \mu_{p-1}.$$

Let us look at some examples of the use of the results we have just proved.

5.6.9 Example. Let $G = K_p$, the complete graph on p vertices. We have $\mathbf{L}(K_p) = p\mathbf{I} - \mathbf{J}$, where $\mathbf{J}$ is the $p \times p$ matrix of all 1's, and $\mathbf{I}$ is the $p \times p$ identity matrix. Since $\mathbf{J}$ has rank one, $p - 1$ of its eigenvalues are equal to 0. Since tr $\mathbf{J} = p$, the other eigenvalue is equal to p. (Alternatively, the column vector of all 1's is an eigenvector with eigenvalue p.) Hence the eigenvalues of $p\mathbf{I} - \mathbf{J}$ are

p ($p - 1$ times) and 0 (once). By the Matrix–Tree Theorem we get

$$c(K_p) = \frac{1}{p}p^{p-1} = p^{p-2},$$

agreeing with the formula for $t(n)$ in Proposition 5.3.2.

5.6.10 Example. Let Γ be the group $(\mathbb{Z}/2\mathbb{Z})^n$ of n-tuples of 0's and 1's under componentwise addition modulo 2. Define a "scalar product" $\alpha \cdot \beta$ on Γ by

$$(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = \sum a_i b_i \in \mathbb{Z}/2\mathbb{Z}.$$

Note that since $(-1)^m$ depends only on the value of the integer m modulo 2, such expressions as $(-1)^{\alpha \cdot \beta + \gamma \cdot \delta}$ are well defined for $\alpha, \beta, \gamma, \delta \in \Gamma$ whether we interpret the addition in the exponent as taking place in $\mathbb{Z}/2\mathbb{Z}$ or in $\mathbb{Z}$. In particular, there continues to hold the law of exponents $(-1)^{\alpha+\beta} = (-1)^\alpha(-1)^\beta$. Let C_n be the graph whose vertices are the elements of Γ, with two vertices α and β connected by an edge whenever $\alpha + \beta$ has exactly one component equal to 1. Thus C_n may be regarded as the graph formed by the vertices and edges of an n-dimensional cube. Equivalently, C_n is the Hasse diagram of the boolean algebra B_n, regarded as a graph. Let V be the vector space of all functions $f : \Gamma \to \mathbb{Q}$. Define a linear transformation $\Phi : V \to V$ by

$$(\Phi f)(\alpha) = nf(\alpha) - \sum_\beta f(\beta),$$

where β ranges over all elements of Γ adjacent to α in C_n. Note that the matrix of Φ with respect to some ordering of the basis Γ of V is just the Laplacian matrix $\mathbf{L}(C_n)$ (with respect to the same ordering of the vertices of C_n). Now for each $\gamma \in \Gamma$ define a function $\chi_\gamma \in V$ by

$$\chi_\gamma(\alpha) = (-1)^{\alpha \cdot \gamma}.$$

Then

$$(\Phi \chi_\gamma)(\alpha) = n(-1)^{\alpha \cdot \gamma} - \sum_\beta (-1)^{\beta \cdot \gamma},$$

with β as above. If γ has exactly k 1's, then for exactly $n - k$ values of β do we have $\beta \cdot \gamma = \alpha \cdot \gamma$, while for the remaining k values of β we have $\beta \cdot \gamma = \alpha \cdot \gamma + 1$. Hence

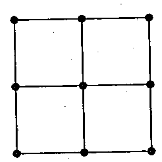
$$\begin{aligned}(\Phi \chi_\gamma)(\alpha) &= (n - [(n - k) - k])(-1)^{\alpha \cdot \gamma} \\ &= 2k\, \chi_\gamma(\alpha).\end{aligned}$$

It follows that χ_γ is an eigenvector of Φ with eigenvalue $2k$. It is easy to see that the χ_γ's are linearly independent, so we have found all 2^n eigenvalues of $\mathbf{L}$, viz., $2k$ is an eigenvalue of multiplicity $\binom{n}{k}$, $0 \le k \le n$. Hence from the Matrix–Tree Theorem there follows the remarkable result

$$\begin{aligned} c(C_n) &= \frac{1}{2^n}\prod_{k=1}^{n}(2k)^{\binom{n}{k}} \\ &= 2^{2^n-n-1}\prod_{k=1}^{n}k^{\binom{n}{k}}. \end{aligned} \tag{5.85}$$

A direct combinatorial proof of this formula is not known.

5.6.11 Example (The efficient mail carrier). A mail carrier has an itinerary of city blocks to which he must deliver mail. He wants to accomplish this by walking along each block twice, once in each direction, thus passing along houses on each side of the street. The blocks form the edges of a graph G, whose vertices are the intersections. The mail carrier wants simply to walk along an Eulerian tour in the digraph $\hat{G}$ defined after Corollary 5.6.7. Making the plausible assumption that the graph is connected, not only does an Eulerian tour always exist, but we can tell the mail carrier how many there are. Thus he will know how many different routes he can take to avoid boredom. For instance, suppose G is the 3×3 grid illustrated below:



This graph has 192 spanning trees. Hence the number of mail carrier routes beginning with a fixed edge (in a given direction) is $192 \cdot 1!^4 2!^4 3! = 18432$. The total number of routes is thus 18432 times twice the number of edges, viz., $18432 \times 24 = 442368$. Assuming the mail carrier delivered mail 250 days a year, it would be 1769 years before he would have to repeat a route!

5.6.12 Example (Binary de Bruijn sequences). A *binary sequence* is just a sequence of 0's and 1's. A *(binary) de Bruijn sequence* of degree n is a binary sequence $A = a_1a_2\cdots a_{2^n}$ such that every binary sequence $b_1\cdots b_n$ of length n occurs exactly once as a "circular factor" of A, i.e., as a sequence $a_ia_{i+1}\cdots a_{i+n-1}$, where the subscripts are taken modulo n if necessary. Note that there are exactly 2^n binary sequences of length n, so the only possible length of a de Bruijn sequence of degree n is 2^n. Clearly any conjugate (cyclic shift) $a_ia_{i+1}\cdots a_{2^n}a_1a_2\cdots a_{i-1}$ of a de Bruijn sequence $a_1a_2\cdots a_{2^n}$ is also a de Bruijn sequence, and we call two such sequences *equivalent*. This relation of equivalence is obviously an equivalence relation, and every equivalence class contains exactly one sequence beginning with n 0's. Up to equivalence, there is one de Bruijn sequence of degree two, namely, 0011. It's easy to check that there are two inequivalent de Bruijn sequences of degree three, namely, 00010111 and 00011101. However, it's not clear at this point whether de Bruijn sequences exist for all n. By a clever application of Theorems 5.6.2 and 5.6.4, we will not only show that such sequences exist for all positive integers n, but will also count them. It turns out that there are *lots* of them. For instance, the number of inequivalent de Bruijn sequences of degree eight is equal to

$$1329227995784915872903807060280344576.$$

Our method of enumerating de Bruijn sequence will be to set up a correspondence between them and Eulerian tours in a certain directed graph D_n, the *de Bruijn graph* of degree n. The graph D_n has 2^{n-1} vertices, which we will take to consist of the 2^{n-1} binary sequences of length $n-1$. A pair $(a_1a_2\cdots a_{n-1}, b_1b_2\cdots b_{n-1})$ of vertices forms an edge of D_n if and only if $a_2a_3\cdots a_{n-1} = b_1b_2\cdots b_{n-2}$, i.e., e is an edge if the last $n-2$ terms of $\text{init}(e)$ agree with the first $n-2$ terms of $\text{fin}(e)$. Thus every vertex has indegree two and outdegree two, so D_n is balanced. The number of edges of D_n is 2^n. Moreover, it's easy to see that D_n is connected (see Lemma 5.6.13). The graphs D_3 and D_4 are shown in Figure 5-17.

Suppose that $E = e_1e_2\cdots e_{2^n}$ is an Eulerian tour in D_n. If $\text{fin}(e_i)$ is the binary sequence $a_{i1}a_{i2}\cdots a_{i,n-1}$, then replace e_i in E by the last bit $a_{i,n-1}$. It is easy to see that the resulting sequence $\beta(E) = a_{1,n-1}a_{2,n-1}\cdots a_{2^n,n-1}$ is a de Bruijn sequence, and conversely every de Bruijn sequence arises in this way. In particular, since D_n is balanced and connected, there exists at least one de Bruijn sequence. In order to count all such sequences, we need to compute $\det \mathbf{L}_0(D_n)$. One way to do this is by a clever but messy sequence of elementary row and column operations which transforms the determinant into triangular form. We will give instead an elegant computation of the eigenvalues of $\mathbf{L}(D_n)$ (and hence of $\det \mathbf{L}_0$) based on the following simple lemma.

5.6.13 Lemma. *Let u and v be any two vertices of D_n. Then there is a unique (directed) walk from u to v of length $n-1$.*

Proof. Suppose $u = a_1a_2\cdots a_{n-1}$ and $v = b_1b_2\cdots b_{n-1}$. Then the unique path of length $n-1$ from u to v has vertices

$$a_1a_2\cdots a_{n-1},\ a_2a_3\cdots a_{n-1}b_1,\ a_3a_4\cdots a_{n-1}b_1b_2,\ \ldots,\ a_{n-1}b_1\cdots b_{n-2},\ b_1b_2\cdots b_{n-1}.$$

□

5.6.14 Lemma. *The eigenvalues of $\mathbf{L}(D_n)$ are 0 (with multiplicity one) and 2 (with multiplicity $2^{n-1}-1$).*

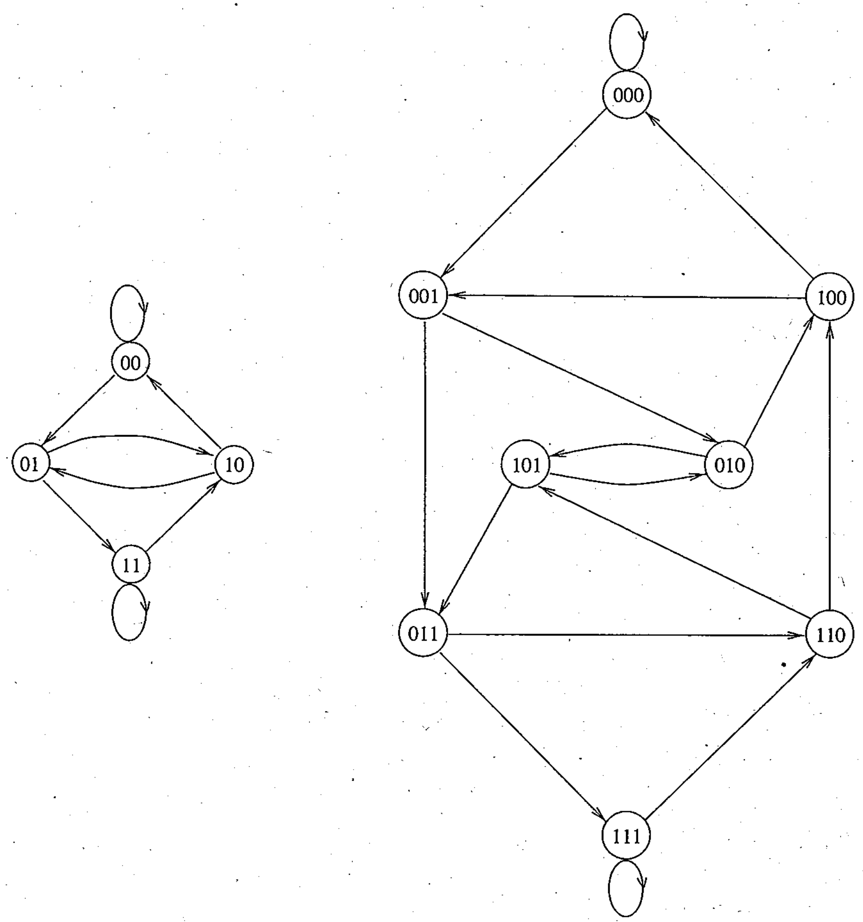


Figure 5-17. The de Bruijn graphs D_3 and D_4.

Proof. Let $\mathbf{A}(D_n)$ denote the directed adjacency matrix of D_n, i.e., the rows and columns are indexed by the vertices, with

$$\mathbf{A}_{uv} = \begin{cases} 1 & \text{if } (u, v) \text{ is an edge,} \\ 0 & \text{otherwise.} \end{cases}$$

Now Lemma 5.6.13 is equivalent to the assertion that $\mathbf{A}^{n-1} = \mathbf{J}$, the $2^{n-1} \times 2^{n-1}$ matrix of all 1's. If the eigenvalues of $\mathbf{A}$ are $\lambda_1, \dots \lambda_{2^{n-1}}$, then the eigenvalues of $\mathbf{J} = \mathbf{A}^{n-1}$ are $\lambda_1^{n-1}, \dots, \lambda_{2^{n-1}}^{n-1}$. By Example 5.6.9, the eigenvalues of $\mathbf{J}$ are 2^{n-1} (once) and 0 ($2^{n-1} - 1$ times). Hence the eigenvalues of $\mathbf{A}$ are 2ζ (once, where ζ is an $(n-1)$-st root of unity to be determined), and 0 ($2^{n-1} - 1$ times). Since the trace of $\mathbf{A}$ is 2, it follows that $\zeta = 1$, and we have found all the eigenvalues of A.

Now $\mathbf{L}(D_n) = 2\mathbf{I} - \mathbf{A}(D_n)$. Hence the eigenvalues of $\mathbf{L}$ are $2 - \lambda_1, \dots, 2 - \lambda_{2^{n-1}}$, and the proof follows from the above determination of $\lambda_1, \dots, \lambda_{2^{n-1}}$. □

5.6.15 Corollary. *The number $B_0(n)$ of de Bruijn sequences of degree n beginning with n 0's is equal to $2^{2^{n-1}-n}$. The total number $B(n)$ of de Bruijn sequences of degree n is equal to $2^{2^{n-1}}$.*

Proof. By the above discussion, $B_0(n)$ is the number of Eulerian tours in D_n whose first edge is the loop at vertex $00 \cdots 0$. Moreover, the outdegree of every vertex of D_n is two. Hence by Corollary 5.6.7 and Theorem 5.6.14 we have

$$B_0(n) = \frac{1}{2^{n-1}} 2^{2^{n-1}-1} = 2^{2^{n-1}-n}.$$

Finally, $B(n)$ is obtained from $B_0(n)$ by multiplying by the number 2^n of edges, and the proof follows. □

Notes[1]

The compositional formula (Theorem 5.1.4) and the exponential formula (Corollary 5.1.6) had many precursors before blossoming into their present form. A purely formal formula for the coefficients of the composition of two exponential generating functions goes back to Faà di Bruno **[23][24]** in 1855 and 1857, and is known as *Faà di Bruno's formula.* For additional references on this formula, see **[2.3**, p. 137]. An early precursor of the exponential formula is due to Jacobi **[38]**. The idea of interpreting the coefficients of $e^{F(x)}$ combinatorially was considered in certain special cases by Touchard **[69]** and by Riddell and Uhlenbeck **[56]**. Touchard was concerned with properties of permutations and obtained our equation (5.30), from which he derived many consequences. Equation (5.30) was earlier obtained by Pólya [**50**, Sect. 13], but he was not interested in general combinatorial applications. It is also apparent from the work of Frobenius (see [**27**, bottom of p. 152 of GA]) and Hurwitz [**37**, §4] that they were aware of (5.30), even if they did not state it explicitly. Riddell and Uhlenbeck, on the other hand, were concerned with graphical enumeration and obtained our Example 5.2.1 and related results.

It was not until the early 1970s that a general combinatorial interpretation of $e^{F(x)}$ was developed independently by Foata and Schützenberger **[26]**, Bender and Goldman **[3.3]**, and Doubilet, Rota, and Stanley **[3.12]**. The approach most like the one taken here is that of Foata and Schützenberger. Doubilet, Rota, and Stanley use an incidence-algebra approach and prove a result (Theorem 5.1) equivalent to our Theorem 5.1.11. The most sophisticated combinatorial theory of power

[1] A reference such as **[*m.n*]** refers to reference *n* of the Notes section to Chapter *m*. A reference without a prefix refers to the reference list of this chapter (which follows these notes).