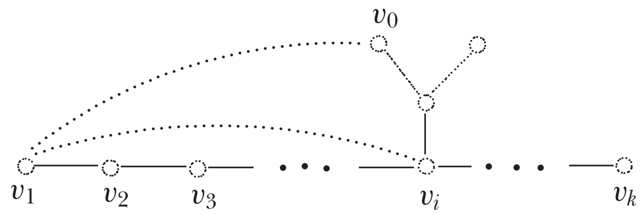


Figure 4.3. If $P = (v_1, v_2, \ldots, v_k)$ is a longest path in a tree, then v_1 can only be adjacent to v_2

Proof: This is obvious if a tree has exactly two vertices. So suppose that T is a tree of order $n \geq 3$ and let P be a path of maximum length in T, say $P = (v_1, v_2, \ldots, v_k)$. We claim that v_1 and v_k are leaves in T. To prove that v_1 is a leaf (the proof for v_k is similar), we must show that v_1 is adjacent only to v_2. Certainly, v_1 can't be adjacent to any other vertex on P for this would create a cycle in T. Nor could v_1 be adjacent to any other vertex, say v_0, not on P for then $(v_0, v_1, v_2, \ldots, v_k)$ would be a path longer than P, contradicting our assumption that P has maximum length. (See Figure 4.3.) ■

Notice that if u is a leaf of the tree T, then the graph $T - u$ obtained by removing u from T is connected and contains no cycles, that is, $T - u$ is also a tree. This fact allows us to "inductively" prove many important theorems about trees. In addition, if a new vertex v is added to T and joined to a single vertex of T, then the resulting graph T' is also connected and contains no cycles; so T' is a tree.

Observe that the tree T_1 of Figure 4.2 has order 7 and size 6 while the tree T_2 has order 9 and size 8. The tree T_3 has order 17 and size 16. In each of these examples, the size of the tree is one less than its order. This is true of every tree.

Theorem 4.3: *Every tree with n vertices has* $n - 1$ *edges.*

Idea of Proof: Figure 4.4 shows all trees of order 6 or less. In each tree the size m is one less than its order n. The vertices of two trees of order 6 are labeled with their degrees. Suppose that T is a tree of order 7 and v is a leaf of T. The vertex v is adjacent to one vertex in T, say u. Since $T - v$ is a tree of order 6 (necessarily one of those

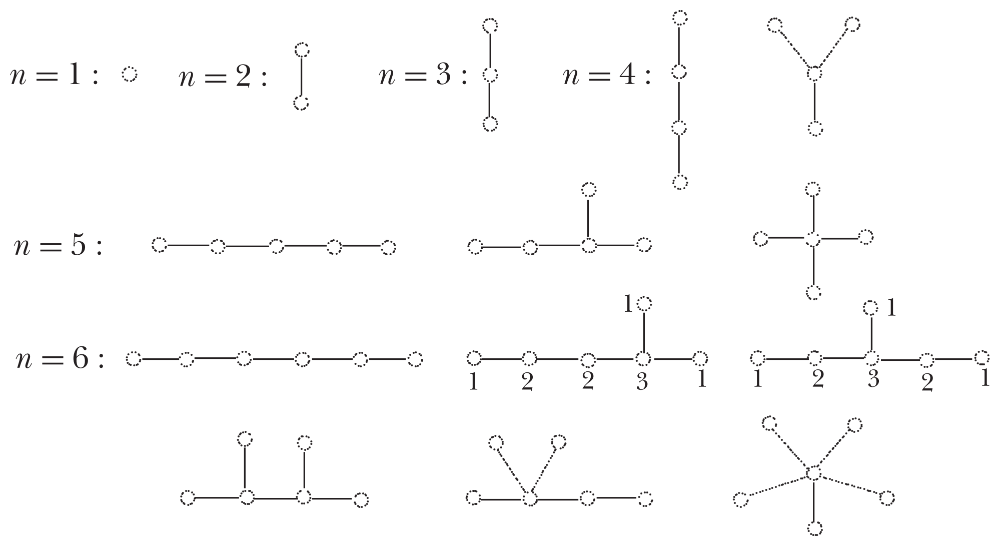


Figure 4.4. The trees with six or fewer vertices

shown in Figure 4.4) and so has five edges, the edges of T consist of these five edges together with the edge uv, or six edges in all.

Repeating this argument shows that every tree with eight vertices has seven edges, every tree with nine vertices has eight edges and so on. ■

From this theorem and the First Theorem of Graph Theory, we have the following.

Corollary 4.4: *If T is a tree of order n and size m whose vertices are $v_1, v_2, \ldots, v_n$, then*

$$\deg v_1 + \deg v_2 + \cdots + \deg v_n = 2m = 2(n-1) = 2n-2.$$

If we know the number of vertices of degree 3 or more in a tree, then the number of leaves can be determined.

Corollary 4.5: *Suppose that T is a tree of order $n \geq 2$ such that Δ is the maximum degree of T and n_i is the number of vertices of degree i for $i = 1, 2, \ldots, \Delta$. Then the number n_1 of leaves in T is*

$$n_1 = 2 + n_3 + 2n_4 + 3n_5 + \cdots + (\Delta - 2)n_\Delta. \tag{4.1}$$

Proof: By the First Theorem of Graph Theory,

$$1n_1 + 2n_2 + 3n_3 + 4n_4 + 5n_5 + \cdots + \Delta n_\Delta = 2n - 2.$$

Since $n = n_1 + n_2 + n_3 + \cdots + n_\Delta$, it follows that

$$1n_1 + 2n_2 + 3n_3 + 4n_4 + 5n_5 + \cdots + \Delta n_\Delta = 2(n_1 + n_2 + \cdots + n_\Delta) - 2. \tag{4.2}$$

Solving equation (4.2) for n_1 gives us the formula in (4.1). ■

Let's now see an example of how Corollary 4.5 can be used to tell us how many leaves a tree contains if we know the number of vertices of every other degree.

Example 4.6: *Suppose that T is a tree with two vertices of degree 3, three vertices of degree 4 and one vertex of degree 6. No other vertices of T have degree 3 or more. How many leaves does T have?*

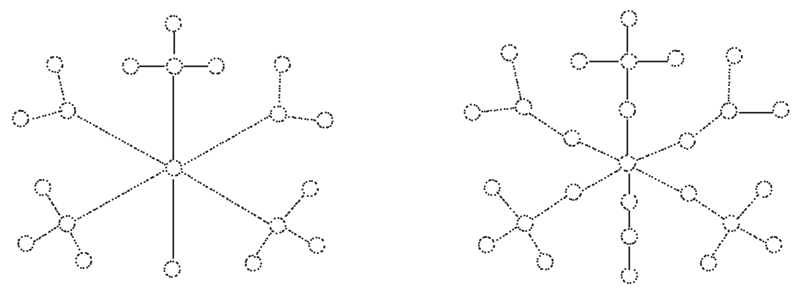
SOLUTION:

Let n_i be the number of vertices of degree i in T, where $1 \le i \le 6$. According to Corollary 4.5, the number n_1 of leaves in T is

$$n_1 = 2 + n_3 + 2n_4 + 3n_5 + 4n_6 = 2 + 2 + 2 \cdot 3 + 3 \cdot 0 + 4 \cdot 1 = 14.$$

Although there is no way to determine precisely which tree T we may be referring to, one possibility for T is shown in Figure 4.5a. You might notice that the formula (4.1) for the number n_1 of leaves in T does not involve the number n_2 of vertices of degree 2. In fact, the number of leaves in a tree is not affected by the number of vertices of degree 2 in a tree. While the tree T in Figure 4.5a has no vertex of degree 2, the tree in Figure 4.5b has seven vertices of degree 2 and satisfies all degree conditions stated in Example 4.6. ♦

We saw that Figure 4.4 shows all of the different (nonisomorphic) trees of order 6 or less. The two trees in this figure having six vertices that include one vertex of degree 3, two vertices of degree 2 and three leaves, are not isomorphic because, for example, the vertices of degree 2 are adjacent in one tree but not in the other. This illustrates the fact that



(a) (b)

Figure 4.5. Trees satisfying the degree conditions of Example 4.6.

even for trees, knowing the exact number of vertices of each degree does not necessarily tell us precisely which graph we might be referring to.

According to Theorem 4.2 and Corollary 4.4, if T is a tree of order $n \geq 2$ and $d_1, d_2, \ldots, d_n$ are the degrees of its vertices, then $\sum_{i=1}^{n} d_i = 2n-2$ and at least two of the degrees $d_1, d_2, \ldots, d_n$ are 1. How about the converse? For example, if you are given six positive integers that sum to 10, must there be a tree T that has these six integers as the degrees of its vertices? The number 10 can be written as the sum of six positive integers (say from largest to smallest) in five ways:

$$10 = 2+2+2+2+1+1, \qquad 10 = 3+2+2+1+1+1,$$

$$10 = 3+3+1+1+1+1, \qquad 10 = 4+2+1+1+1+1,$$

$$10 = 5+1+1+1+1+1.$$

For each sum, there is a tree of order 6 (in Figure 4.4) having these degrees for its vertices. In fact, the first sum listed is the first tree of order 6 listed in Figure 4.4.

Theorem 4.7: *If $d_1, d_2, \ldots, d_n$ are $n \geq 2$ positive integers whose sum is $2n-2$, then these integers are the degrees of the vertices of some tree T of order n.*

Idea of Proof: We have already seen that the theorem is true for $n = 6$. How can we use this fact to show that the theorem is true for $n = 7$? Let $d_1, d_2, \ldots, d_7$ be seven positive integers whose sum is $2n-2=2(7)-2=12$, that is, $d_1+d_2+\cdots+d_7=12$. At least one of these numbers, say d_1, is 2 or more. On the other hand, if all seven of these numbers were 2 or more, then the sum would be at least 14; so at least one number is 1, say $d_7=1$. Now, ignoring d_7 and decreasing d_1 by 1, we arrive at six positive integers that sum to 10, namely $(d_1-1)+d_2+\cdots+d_6=10$. But since we know the theorem is true for $n=6$, there is a tree T' of order 6 with six vertices $v_1, v_2, \ldots, v_6$ with respective degrees $d_1-1, d_2, \ldots, d_6$. Now appending the vertex v_7 to T' and joining it to v_1 produces a tree T whose seven vertices have the desired property.

For example, consider the seven numbers 3, 3, 2, 1, 1, 1, 1 which sum to 12. We create a tree T with seven vertices whose degrees are these seven numbers. Reducing the largest number by 1 and deleting the least number gives us the six integers 2, 3, 2, 1, 1, 1, which sum to 10. Locating a tree in Figure 4.4 having $n=6$ whose vertices have these degrees (there are two of them) and attaching a leaf to any vertex of degree 2 produces a desired tree T.

Now repeating this process for $n=8$, $n=9$ and so on establishes the theorem. ■

Example 4.8: *The sum of the 17 integers* 5, 4, 3, 3, 3, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 *is* 32. *Since* $32=2\cdot 17-2$, *it follows by Theorem 4.7 that there is a tree of order* 17, *the degrees of whose vertices are these* 17 *integers. An example of such a tree* T *is given in Figure 4.6a.*

One must be careful not to read too much into Theorem 4.7. Not every graph whose vertices have the degrees mentioned in Example 4.8 must be a tree. For example, the vertices of the graph G in Figure 4.6b have the same degrees as the vertices of T but G is not a tree.

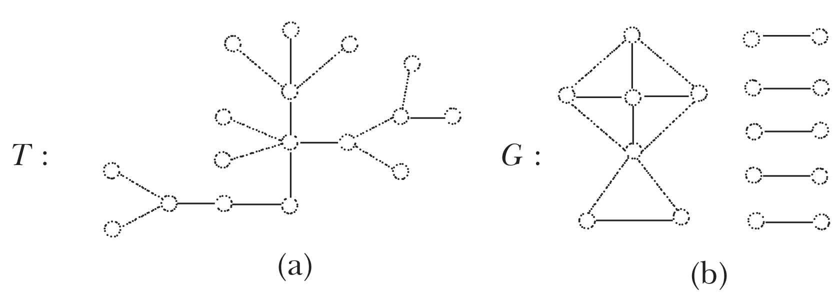


Figure 4.6. Two graphs of the same order whose vertices have the same degrees.

TREES AND SATURATED HYDROCARBONS

The British mathematician Arthur Cayley (1821–1895) was one of the great mathematicians of the nineteenth century. In fact, Cayley wrote 967 research papers and is one of the most prolific mathematicians of all time. Paul Erdős wrote more than 1525 papers, Euler 886 and Augustin-Louis Cauchy 789. As we saw in Chapter 3, many papers of Erdős were coauthored. The papers of Cayley and Euler are noted for their volume. While Cayley was best known for his work in modern algebra, he also spent several years as a lawyer. Although Cayley's connections with graph theory were few in number, they were nevertheless important.

In 1857, while trying to solve a problem in differential calculus, Cayley was led to the problem of counting all *rooted trees*—trees in which a particular vertex has been designated as the *root*. In 1875 Cayley would encounter another tree-counting problem when he studied the counting of special kinds of molecules called saturated hydrocarbons (also referred to as alkanes or paraffins). These molecules consist of carbon atoms and hydrogen atoms. In a saturated hydrocarbon, every carbon atom has valency 4 and every hydrogen atom has valency 1. The chemical formula for each such molecule is C_nH_{2n+2}, indicating that each molecule has n carbon atoms and $2n+2$ hydrogen atoms for some positive integer n. For $n = 1, 2, 3$, these molecules are called methane, ethane and propane, respectively. The diagrams for these three molecules, as shown in Figure 4.7, are actually trees, where each "carbon vertex" has degree 4 and each "hydrogen vertex" is a leaf. Not only is the valency of an atom in

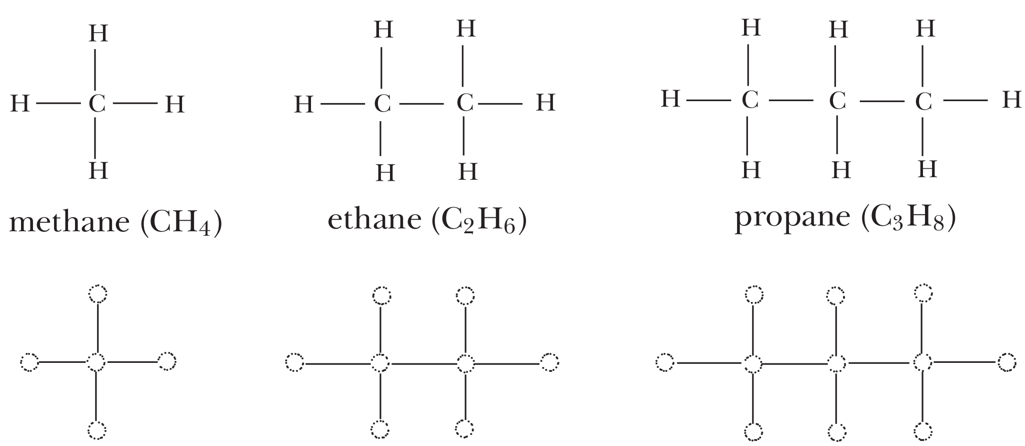


Figure 4.7. The molecules C_nH_{2n+2} for $n = 1, 2, 3$.

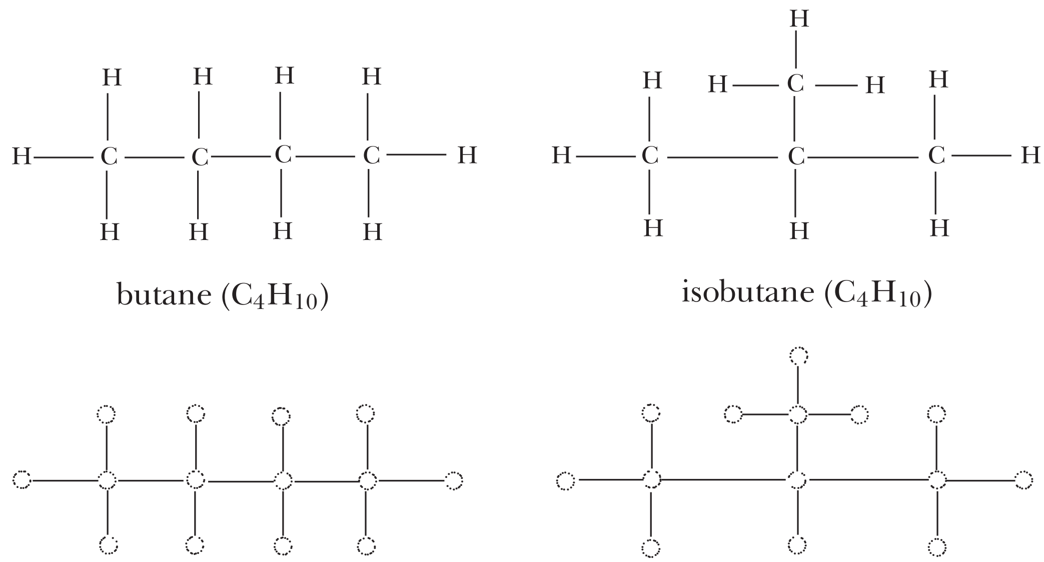


Figure 4.8. The molecules C_4H_{10}.

a molecule the same as the degree of the vertex in the tree representing the molecule, there were times when some mathematicians referred to the degree of a vertex in a graph as the valency of that vertex.

The saturated hydrocarbons butane and isobutane (see Figure 4.8) have the same formula, C_4H_{10}, indicating that each molecule has four carbon atoms and ten hydrogen atoms. Nevertheless, they have different properties. These are examples of what are called isomers, which are molecules with different properties having the same formula.

While there are only two isomers whose formula is C_4H_{10}, determining the number i_n of isomers having the formula C_nH_{2n+2} for $n \geq 5$ is considerably more complicated. Cayley was able to compute the numbers i_n correctly for $1 \leq n \leq 11$:

n	1	2	3	4	5	6	7	8	9	10	11
i_n	1	1	1	2	3	5	9	18	35	75	159

Counting the number of saturated hydrocarbons is the same as counting the number of certain kinds of nonisomorphic trees. If the hydrogen atoms are removed from the tree diagram of a saturated hydrocarbon, another tree is obtained—a so-called carbon tree, in which all vertices have degree 4 or less. So the number of saturated hydrocarbons with n carbon atoms equals the number of trees with n vertices where the degree of each vertex is 4 or less. From the trees drawn in Figure 4.4, we see that the number t_n of such trees with n vertices is that given below.

n	1	2	3	4	5	6
t_n	1	1	1	2	3	5

This, of course, agrees with the first portion of the table for i_n discovered by Cayley.

CAYLEY'S TREE FORMULA

In 1889 Cayley turned his attention to yet another counting problem, one that involved counting labeled trees. Suppose that T' and T'' are two trees of order n, each of whose vertices are labeled with the positive integers $1, 2, \ldots, n$. Then each tree has $n - 1$ edges. The labeled trees T' and T'' are considered different if they don't contain the same edges. For example, the vertices of the three trees T_1, T_2 and T_3 of order 3 in Figure 4.9 are labeled with 1, 2, 3. While these three trees are isomorphic, as labeled trees they are all different. For example, T_1 is the only tree not containing the edge 12.

Figure 4.10 shows the 16 labeled trees with four vertices. Let's see why there are 16 such trees. First, there are only two nonisomorphic trees with four vertices, namely the "path graph" and the "star graph". There are exactly four ways to label the vertices of the star graph with the numbers 1, 2, 3, 4 as the labeling is completely determined by the label given to the central vertex. There are $4 \cdot 3 \cdot 2 \cdot 1 = 4!$ ways to label the vertices of the path graph from one end of the path to the other but each labeling is equivalent to the reverse labeling starting from the other end of the path, so there are only $4!/2 = 12$ distinct ways to label the vertices of the path graph. Consequently, there are $4 + 12 = 16$ ways to label all trees of order 4.

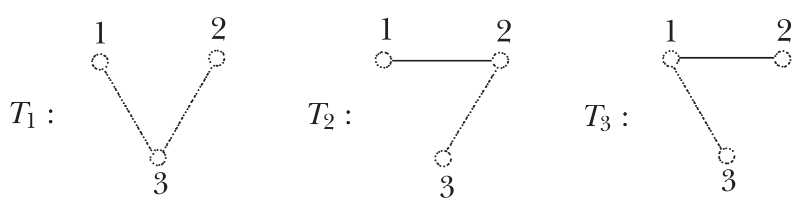


Figure 4.9. The three different trees of order 3.

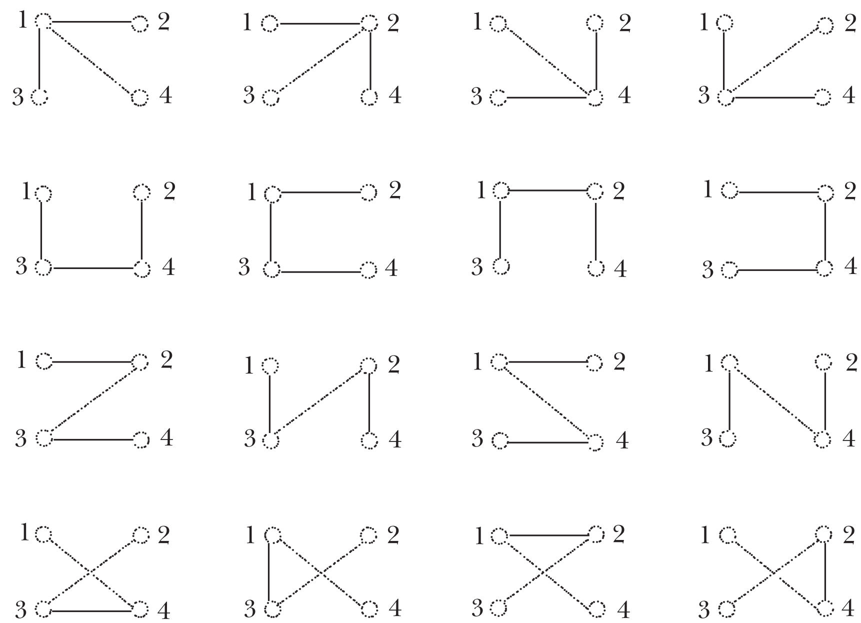


Figure 4.10. Labeled trees with four vertices.

There are three nonisomorphic trees with five vertices, namely the star graph, the path graph and the “broomstick graph”. As above, there are five ways to label the star graph and $5!/2 = 60$ ways to label the path graph. For the broomstick graph, there are five choices for the vertex of degree 3 and four choices for the vertex of degree 2. Since there are three choices for the leaf adjacent to the vertex of degree 2, there are $5 \cdot 4 \cdot 3 = 60$ ways to label this graph in all, or $5 + 60 + 60 = 125$ labeled trees of order 5.

It is an exercise to show that there are 1296 ways to label the trees of order 6. Observing that $3 = 3^1$, $16 = 4^2$, $125 = 5^3$ and $1296 = 6^4$, it is impossible not to think that the number of ways to label the trees of order n is n^{n-2}, as Cayley thought. While Cayley was the one who stated the general formula for the number of labeled trees of a given order and, using the same set of labels, illustrated it for trees of order 6, he did not give a complete proof.

Cayley’s Tree Formula

There are n^{n-2} different labeled trees whose vertices are labeled with the same set of n labels.

Figure 4.11. Two trees whose vertices have degrees 3, 2, 2, 1, 1, 1.

There is another result concerning labeled trees—this one dealing with the degrees of the vertices. According to Theorem 4.7, there is a tree of order 6 whose vertices have degrees 3, 2, 2, 1, 1, 1. In fact, there are two nonisomorphic trees whose vertices have degrees 3, 2, 2, 1, 1, 1, both shown in Figure 4.11.

A related result is the following.

Theorem 4.9: *If $d_1, d_2, \ldots, d_n$ are n positive integers whose sum is $2n-2$, then the number of labeled trees of order n where vertex 1 has degree d_1, vertex 2 has degree d_2 and so on is*

$$\frac{(n-2)!}{(d_1-1)!(d_2-1)!\cdots(d_n-1)!}.$$

For example, the degree sequence

$$5, 4, 3, 3, 3, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1$$

in Example 4.8 can be represented by

$$\frac{(17-2)!}{(5-1)!(4-1)!(3-1)!(3-1)!(3-1)!} = 1{,}135{,}134{,}000$$

different labeled trees, in which the vertex of degree 5 is labeled 1, the vertex of degree 4 is labeled 2, the three vertices of degree 3 are labeled 3, 4, 5, the two vertices of degree 2 are labeled 6 and 7 and the ten leaves are labeled with 8, 9, ..., 17. The degree sequence 3, 2, 2, 1, 1, 1 for the two trees in Figure 4.11 can therefore be represented by

$$\frac{(6-2)!}{(3-1)!(2-1)!(2-1)!(1-1)!(1-1)!(1-1)!} = \frac{24}{2} = 12$$

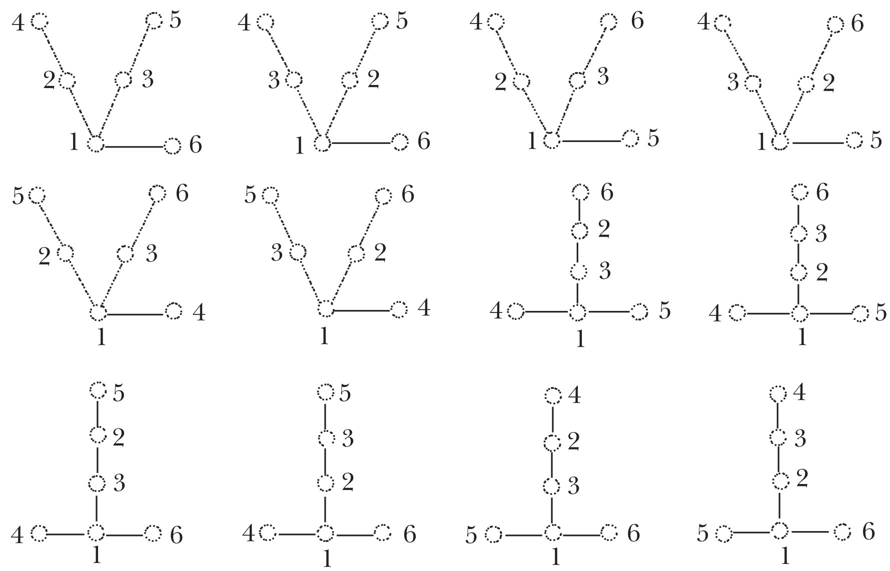


Figure 4.12. The 12 labeled trees with degree sequence $3, 2, 2, 1, 1, 1$, whose vertex of degree 3 is labeled 1, whose vertices of degree 2 are labeled $2, 3$ and whose leaves are labeled $4, 5, 6$.

labeled trees, where the vertex of degree 3 is labeled 1, the two vertices of degree 2 are labeled 2 and 3 and the leaves are labeled 4, 5, 6. These 12 trees are shown in Figure 4.12.

PRÜFER CODES

The first complete proof of Cayley's Tree Formula was given in 1918 by the German mathematician Heinz Prüfer (1896–1934). Although the verification of Prüfer's technique is a bit complicated to explain, the technique itself is not difficult to describe.

Consider the tree T of order 9 shown in Figure 4.13, whose vertices are labeled with $1, 2, \dots, 9$. Associated with the tree T is a sequence of length 7 (2 less than the order of T), each of whose terms is one of the labels $1, 2, \dots, 9$. This sequence is called the *Prüfer sequence* or the *Prüfer code* of the tree.

We now describe how to compute the Prüfer code of the tree T of Figure 4.13. First, let $T_0 = T$. In T_0, the five leaves have the labels 3, 4, 5, 6 and 8. The smallest label of a leaf of T_0 is therefore 3. The neighbor of this vertex is labeled 9. This is the first term of the Prüfer code of T.

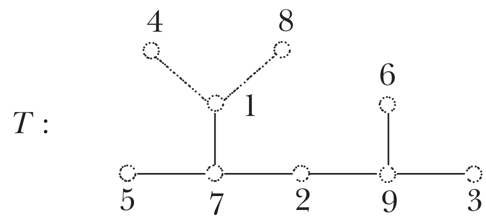


Figure 4.13. A labeled tree of order 9.

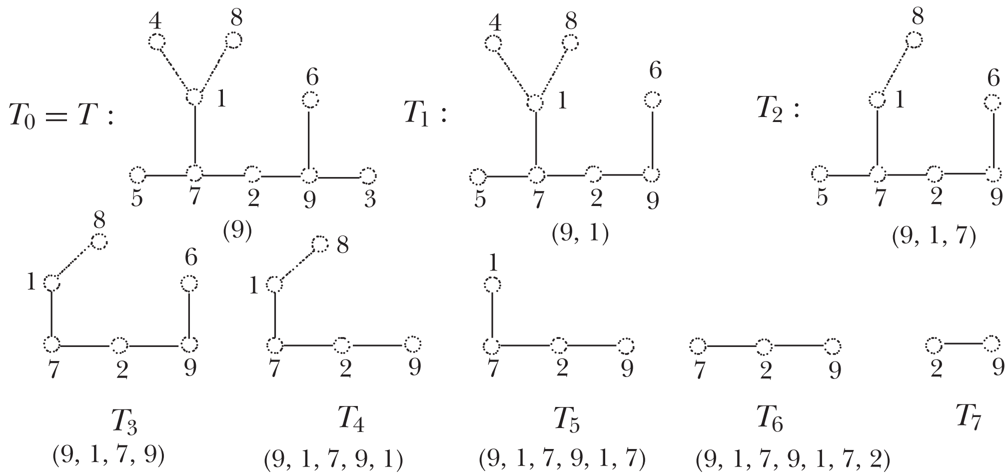


Figure 4.14. Constructing the Prüfer code for a tree.

This leaf is then deleted from T_0, producing a tree T_1 of order 8. The neighbor of the leaf of T_1 having the smallest label is the second term of the Prüfer code for T. Here T_1 has leaves with labels 4, 5, 6, 8 and the neighbor of leaf 4 is vertex 1. Thus the second term is 1. This is continued until we arrive at $T_7 = K_2$, resulting in the Prüfer code (9, 1, 7, 9, 1, 7, 2) (see Figure 4.14). In general, the Prüfer code of a tree of order n (whose vertices are labeled with $1, 2, \ldots, n$) is a sequence of length $n-2$, each of whose terms is an element of $\{1, 2, \ldots, n\}$.

We now consider the converse question. That is, for some integer $n \geq 3$, suppose that we have a sequence of length $n-2$, each of whose terms is an element of the set $\{1, 2, \ldots, n\}$. Which labeled tree T of order n has this sequence as its Prüfer code? For example, suppose that $n = 9$ and that $(9, 1, 7, 9, 1, 7, 2)$ is the given sequence of length 7. The smallest element of $\{1, 2, \ldots, 9\}$ not in this sequence is 3. We join vertex 3 to vertex 9 (the first element of the sequence). The first term of the sequence is then deleted, producing the reduced sequence $(1, 7, 9, 1, 7, 2)$. The element 3 is removed from the set $\{1, 2, \ldots, 9\}$,

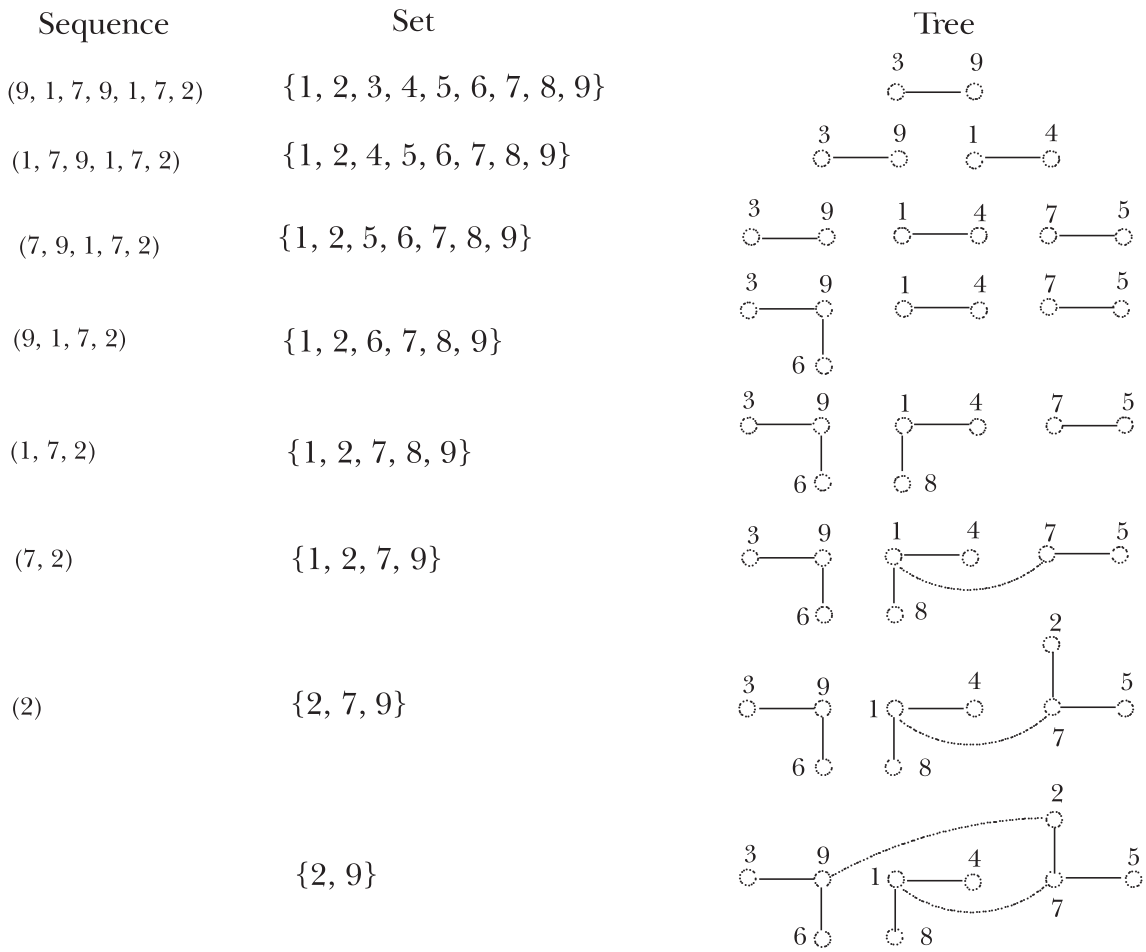
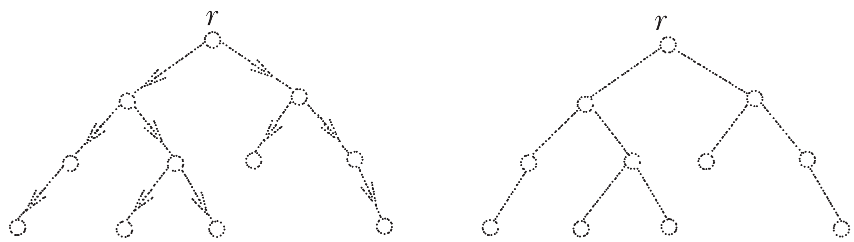


Figure 4.15. Constructing a labeled tree from a sequence.

giving us the set {1, 2, 4, 5, 6, 7, 8, 9}. The smallest element of this set not in the sequence is 4, which is joined to vertex 1 (the first element of the reduced sequence). We continue this procedure until no sequence remains and the final two elements of the set are 2 and 9. These two vertices are joined, producing the tree T of Figure 4.13. This procedure is illustrated in Figure 4.15. As we saw, the Prüfer code of this tree is (9, 1, 7, 9, 1, 7, 2).

What Heinz Prüfer was able to prove is that each labeled tree has a Prüfer code that no other labeled tree has and that every Prüfer code corresponds to a unique labeled tree. Since a Prüfer code has the appearance $(a_1, a_2, \ldots, a_{n-2})$ and each term is one of the elements $1, 2, \ldots, n$, the total number of Prüfer codes is obtained by computing the product $n \cdot n \cdot \ldots \cdot n$ (a total of $n-2$ terms), arriving at n^{n-2}. That is, there



(a) (b)

Figure 4.16. A rooted binary tree.

are n^{n-2} different labeled trees of order n, where the trees are labeled with the same set of n elements.

DECISION TREES

We saw that a rooted tree is a tree T in which some vertex r is designated as the root. Typically, each edge of T is directed away from r and so T becomes a directed tree. It is customary to draw T with r at the top and all arcs of T directed downward. In a directed rooted tree, every vertex v different from the root has exactly one arc incident with v that is directed into v. If every vertex u of a directed rooted tree T has at most two arcs directed away from u, then T is called a *binary tree*, such as the rooted tree in Figure 4.16a. Since it is understood that all arcs of a rooted tree are directed downward, rooted trees are often drawn without arrowheads. Thus the binary tree in Figure 4.16a can also be drawn as in Figure 4.16b.

Rooted trees can be used to represent a variety of situations. They can be especially useful when encountering problems in which solutions can be found by means of successive comparisons where each vertex v of the tree corresponds to a decision which when made directs us via an arc (v, u) to another decision to be made at the vertex u. We continue doing this until we reach a leaf, indicating that a solution to the problem has been found. For this reason, such a rooted tree is called a *decision tree*. There is one type of problem for which a procedure for finding a solution can be conveniently described by means of a decision tree.

Suppose that we have a collection of coins, all of which look the same and all but one of which are authentic coins. The remaining coin is a fake coin, however, and weighs slightly less than each of the authentic coins. What we have available is a balance scale (see Figure 4.17) consisting of