# Trees

**Equivalent forms of the definition of a tree:**

(1) A ***tree*** is a connected graph that contains no cycles.

(2) A tree is a connected graph that contains no circuits.

(3) A tree is a graph with exactly one path between any two vertices.

**Definitions**

Vertices of a tree are called ***nodes***.

Edges of a tree are called ***branches***.

A tree is ***rooted*** if one of its vertices is specified as the ***root.***

Any node of a given tree can be specified as the root. Example: Fig.1

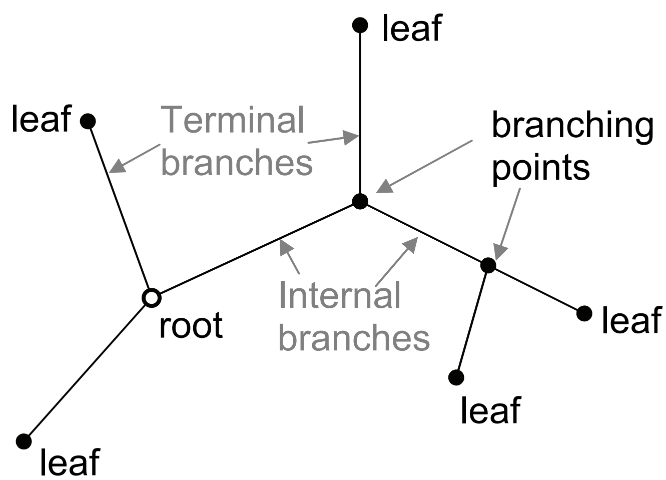


Figure 1. Rooted tree of degree 5.

Formally speaking, an isolated node is a tree. If a rooted tree has more than two nodes, then it necessarily (prove!) has:

- ***leafs***, or ***terminal nodes*** (i.e., nodes of degree 1), and
- ***internal nodes***, or ***branching points*** (degree >1).

The number of leaves in a tree is called the ***degree*** of the tree.
The number of nodes in a tree is called the ***size*** of the tree.

Therefore, the notion of a degree for trees differs from the notion of a degree of a vertex in a graph (which is the number of incident edges). Sometimes (not in this course) ***the degree*** of a node in a tree is understood as the degree of the subtree rooted at this node.

A ***pruned tree*** (also called a ***subtree***) rooted at a given node $x$ is the tree rooted at $x$ that results from deletion of the incident with $x$ branch leading to the original root. *Prove that a subtree is a tree*.

For example, a subtree rooted at $x$ in Figure 2 A is obtained by deleting the branch $xy$ and selecting the component containing $x$.

The ***level*** (or ***depth***) of a node is its path length to the root.

There is a convention for picturing rooted trees (Figure 2). The root goes on the top, branches are drawn down starting from the root, all branches have the same vertical dimension that separates levels of the tree. In ***sorted trees***, nodes at each level are sorted from left to right in the order of depths of their subtrees.

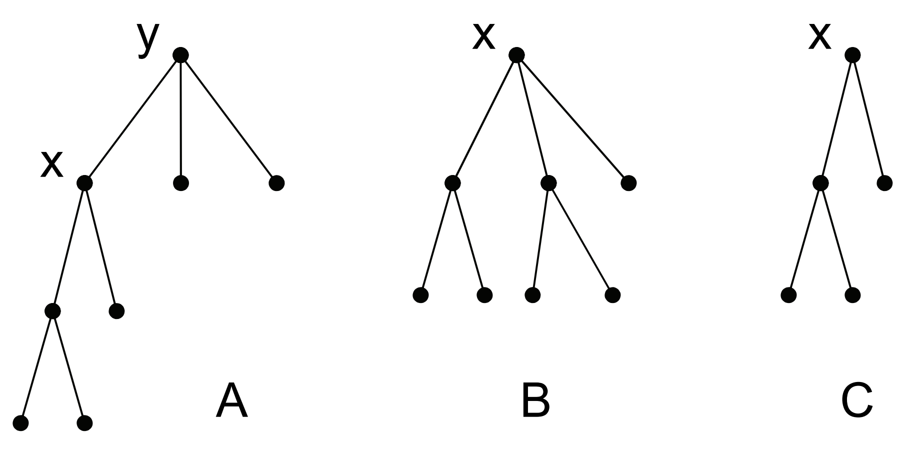


Figure 2. **A:** Same tree as in Figure 1 represented according to the conventional rules. **B:** The same tree with a different selection of the root, labeled “x” in A, B. **C:** The subtree rooted at x of the tree shown in A.

**Theorem:**

A connected graph of $n$ vertices is a tree iff it has $n$-1 edges.

*Proof*

The theorem for $n \leq 2$ can be verified on a case-by-case basis. Assuming that this is done, consider the case $n > 2$.

(→) Suppose that a connected graph $G$ of $n > 2$ vertices is not a tree. Then it has a cycle including $m$ vertices and $m$ edges, with $2 < m \leq n$. Suppose that $m < n$. Because G is connected, there is a path from the $m$-cycle to any vertex that is not in the cycle. Therefore, starting with the $m$-cycle, one can reconstruct $G$ by adding edges and vertices one by one while keeping the graph connected. During this process, one must add at least one new edge for each new vertex in order to connect it. Therefore, the total number of edges in $G$ cannot be less than $n$. Therefore, if the total number of edges in a connected graph $G$ of $n$ vertices is $n$-1, then $G$ is a tree.

(←) Consider a tree $T$ of $n > 1$ nodes. W.l.o.g., select a root in $T$. Then for every other node in $T$ there is a unique path to the root and a unique incident branch that belongs to this path. And vice versa, every branch in $T$ has a unique incident node for which this branch is the first step to the root. Therefore, the number of branches in $T$ is $n$-1. QED.

**Corollaries:**

- Any edge added to a tree must produce a cycle.

*Proof:* according to the Theorem, the number of edges in a tree is uniquely determined by the number of vertices. Adding one edge destroys this relation, therefore, the resultant graph is not a tree and must contain a cycle, according to the definition of a tree.

- A tree with $n > 1$ nodes has at least 2 leaves.

*Proof*: according to the Theorem, the sum of degrees of a node in a tree of $n$ nodes is $2(n-1)$. If a tree of $n$ nodes had no leaves or only one leaf, then its sum of node degrees would be greater or equal than $2n-1 > 2(n-1)$. Therefore, it must have at least two leaves.

*Outline of an alternative proof (not based on the Theorem)*:
From an arbitrarily selected branch, start growing a path (extending both ends) until the process terminates – necessarily, at two distinct leaves. Otherwise the tree would contain a cycle.

**Example application of the tree concept:** paraffin isomers (Figure 3) are trees in which every internal node has degree 4. If we delete all hydrogens (whose number and positions are determined by the carbon skeleton), then we have all possible trees in which every internal node has degree ≤ 4.

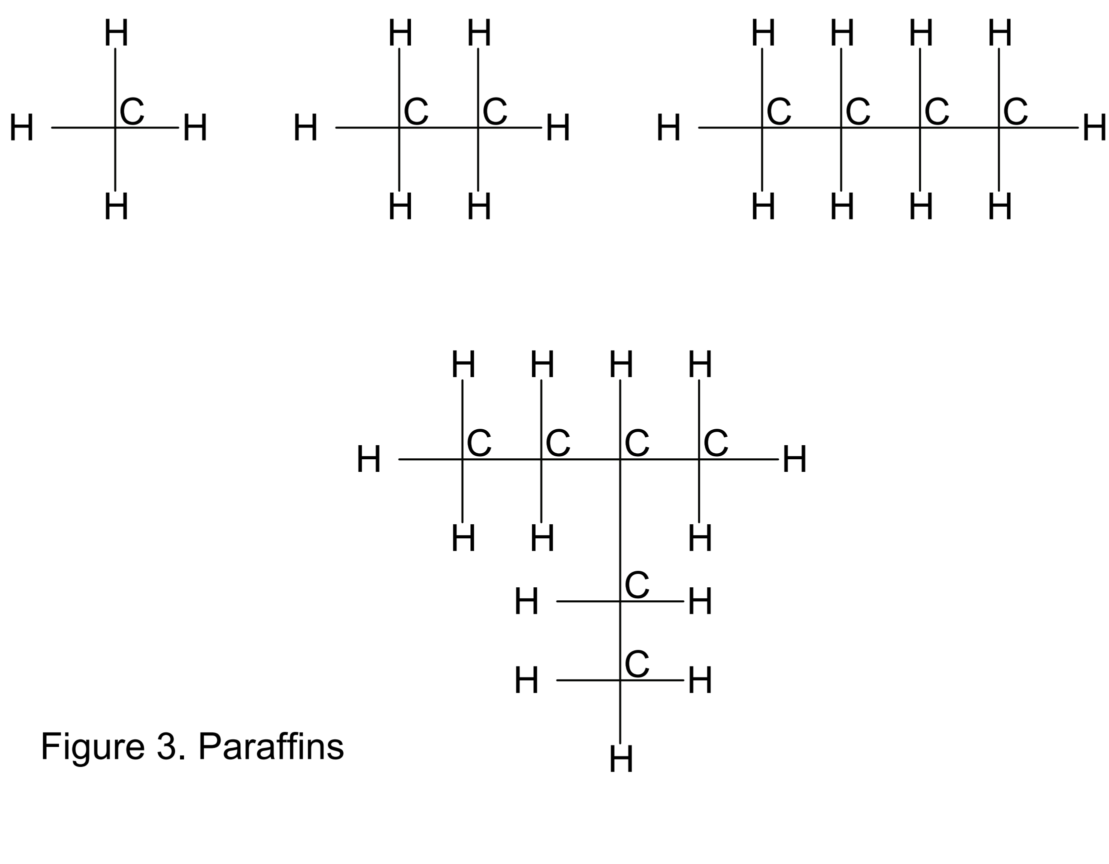

Figure 3. Paraffins

Trees are graphs. Recall the notion of isomorphism for graphs. How many isomorphism classes of trees of $n$ nodes are there?

**Table 1.** Numbers of (unlabeled) trees of size $n$ that are not isomorphic to each other.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Number of trees | 1 | 1 | 1 | 2 | 3 | 6 | 11 | 23 |

**Definitions:**

A ***labeled tree*** has unique labels attached to each node.

Two labeled trees with the same set of labels are ***isomorphic*** iff they have the same adjacency matrix.

**Examples:**

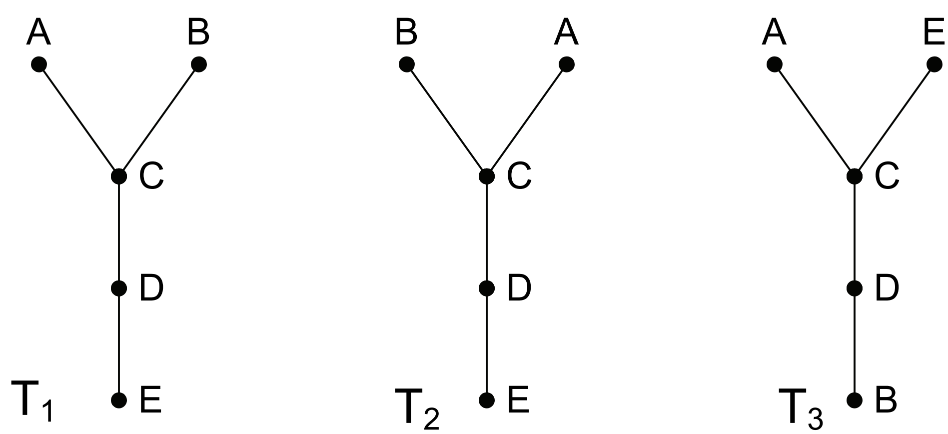


Figure 4. Labeled trees $T_1$ and $T_2$ are isomorphic, while $T_1$ and $T_3$ are not.

**Definitions for rooted trees:**

***Children***, or ***daughters***, of a given node are adjacent nodes of a higher level (higher depth).

***Parent*** is the adjacent node of a lower level.

***Siblings*** are children of the same parent.

The ***height*** of a tree is the length of the path from the root node to its furthest leaf.

**Definitions for a rooted tree represented as a digraph:**

***Directed edge*** is the arc from the parent to the child.

If a path exists from node $p$ to node $q$, where node $p$ is closer to the root than $q$, then $p$ is an ***ancestor*** of $q$, and $q$ is a ***descendant*** of $p$.

A proper rooted binary tree, or a ***binary tree***, is a rooted tree in which every internal node has degree 3, except that the root has degree 2.

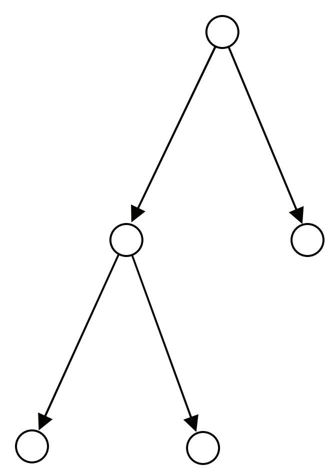

Figure 5. A binary tree represented as a digraph.

Therefore:

(a) A leaf is a node that has no children.
(b) The only node without parent is the root.
(c) In a binary tree, every internal node that is not the root has one parent and two children (called the ***left daughter*** and the ***right daughter*** in an ordered binary tree – see below).

An ***ordered tree*** is a rooted tree in which the order of branches (from left to right) is specified for each node. An ordered tree can be drawn on a plane without self-intersections with the root placed on the top and branches sorted from left to right by their order.

Two (unlabeled) ordered trees are isomorphic iff they are isomorphic graphs and the isomorphism preserves the root and the order of branches.

**Problem:** What is the number $C_n$ of ordered binary trees of degree $n$ that are not isomorphic to each other?

*Answer:* These numbers are known as Catalan numbers (after the Belgian mathematician Eugène Charles Catalan, 1814–1894) and appear in solutions of many other problems.

**Table 2.** Numbers of ordered binary trees of degree $n$, known as Catalan numbers, for $n$ = 1..8.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $C_n$ | 1 | 1 | 2 | 5 | 14 | 42 | 132 | 429 |

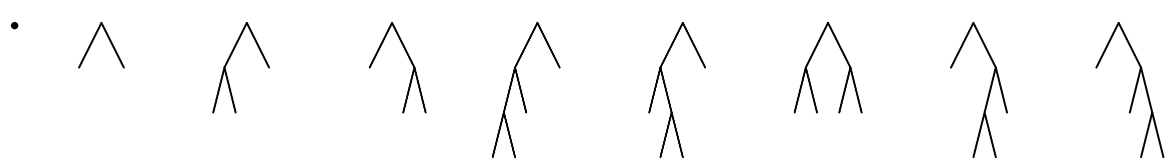

Figure 6. The first 9 ordered binary trees ($n$ = 1, 2, 3 and 4).

Catalan numbers are given by the formula

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

that follows from the following recurrence relation:

$$C_1 = 1, \qquad C_n = \sum_{k=1}^{n-1} C_k C_{n-k} . \qquad (*)$$

The rationale for this recurrence relation is that every ordered binary tree of degree $n > 1$ can be partitioned into 3 parts: the root, the left subtree and the right subtree. The two subtrees are ordered binary trees on their own. The sum of degrees of the two subtrees is $n$. The degree $k$ of the left subtree runs from 1 to $n$-1. Once this degree $k$ is specified, configurations of the two subtrees can be selected independently, and therefore (by the multiplication principle) the total number of possibilities for each pair ($n$, $k$) is the product $C_k C_{n-k}$ (Figure 7).

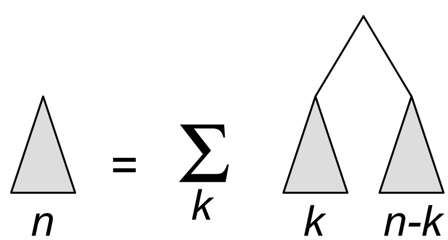


Figure 7. Diagram for derivation of (*). Each shaded triangle represents a set of possible ordered binary tree configurations.

**Exercise 1**

Derive a recurrence relation similar to (*) for the number of non-isomorphic *non-ordered* binary trees of degree $n$, and compute the first 8 numbers. Compare with Table 1. *Hint:* use Figure 7.

Clarification: here "non-ordered" means that two configurations of a tree obtained from each other by permutation of any pair of siblings are considered isomorphic. For example, among the last 5 configurations shown in Figure 6 there are only 2 non-isomorphic non-ordered configurations.