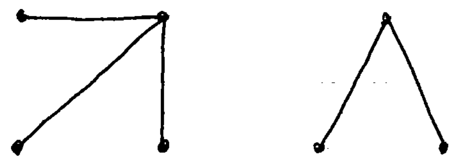
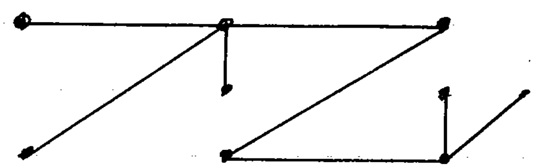
B.5 TREES

A GRAPH is CALLED ACYCLIC (ALSO A FOREST) IF IT CONTAINS NO CYCLES.



A GRAPH WHICH IS BOTH ACYCLIC AND CONNECTED IS CALLED A TREE.



THEOREM (TREENESS)

LET $G=(V,E)$ BE A GRAPH. THE FOLLOWING STATEMENTS ARE EQUIVALENT.

(1) G is A TREE

(2) ANY TWO VERTICES ARE JOINED BY A UNIQUE PATH.

(3) G is CONNECTED, BUT IF ANY EDGE is REMOVED, IT BECOMES DISCONNECTED.

(4) G is CONNECTED, AND $|E|=|V|-1$

(5) G is ACYCLIC, AND $|E|=|V|-1$

(6) G is ACYCLIC, BUT IF ANY EDGE is ADDED, A UNIQUE CYCLE is CREATED.

THE COMPLETE PROOF is ON P. 1085.

WE PROVE HERE THAT (1) AND (3) imply (5).

PROOF:

LET T BE A TREE, SO THAT T IS ACYCLIC. LET $n = |V(T)|$ AND $m = |E(T)|$. WE USE INDUCTION ON m TO SHOW $m = n - 1$.

BASE: $m = 1 \Rightarrow \overset{T}{\bullet\!\!-\!\!-\!\!\bullet} \Rightarrow n = 2 \Rightarrow m = n - 1$.

STRONG INDUCTION:

LET $m > 1$ AND ASSUME FOR ALL TREES T' WITH $|E(T')| < m$, THAT $|E(T')| = |V(T')| - 1$.

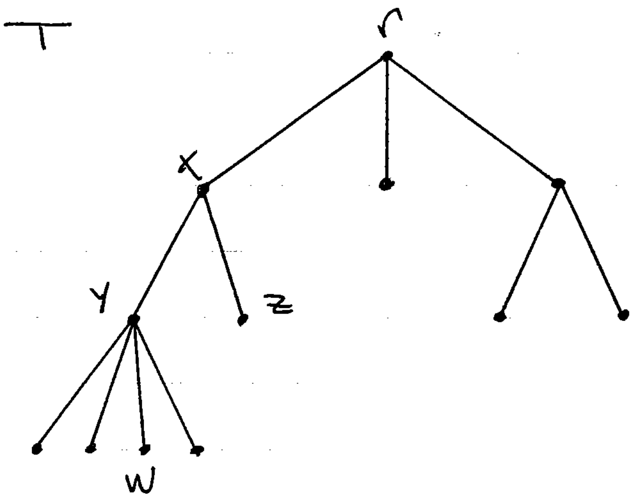
NOW CHOOSE ANY EDGE $e \in E(T)$ AND REMOVE IT FROM T. BY (3) THE RESULTING GRAPH IS DISCONNECTED, AND CONSISTS OF TWO DISJOINT TREES T_1, T_2, EACH WITH FEWER THAN m EDGES. THUS

$$
\begin{aligned}
|E(T)| &= |E(T_1)| + |E(T_2)| + 1 \\
&= (|V(T_1)| - 1) + (|V(T_2)| - 1) + 1 \quad \text{(BY IND. HYP.)} \\
&= |V(T_1)| + |V(T_2)| - 1 \\
&= |V(T)| - 1
\end{aligned}
$$

$\therefore$ RESULT HOLDS FOR ALL TREES BY INDUCTION. ///

A <u>ROOTED TREE</u> IS A TREE IN WHICH ONE VERTEX HAS BEEN DISTINGUISHED AS THE <u>ROOT</u>. WE OFTEN REFER TO VERTICES AS <u>NODES</u> IN THIS CASE. THE TERM <u>FREE TREE</u> IS SOMETIMES USED TO DISTINGUISH AN ORDINARY (NON-ROOTED) TREE FROM A ROOTED TREE.

<u>EX</u> T



THE ROOT IS USUALLY DRAWN AT THE TOP.

X IS AN <u>ANCESTOR</u> OF W, AND W A <u>DESCENDANT</u> OF X

IN THIS EXAMPLE X IS THE <u>PARENT</u> OF Y, Y IS A <u>CHILD</u> OF X, Z IS A <u>SIBLING</u> OF Y. A NODE WITH NO CHILDREN SUCH AS W IS CALLED A <u>LEAF</u> (OR <u>EXTERNAL NODE</u>). A NON-LEAF NODE IS CALLED AN <u>INTERNAL NODE</u>.

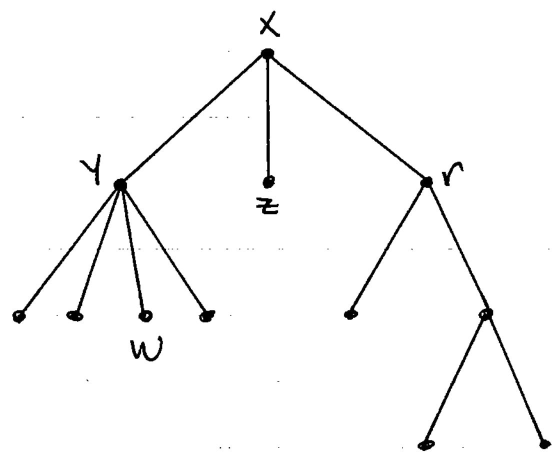
THE <u>DEPTH</u> OF A NODE IS ITS DISTANCE FROM THE ROOT, SO r, x, y, w HAVE DEPTHS 0, 1, 2, AND 3 RESPECTIVELY.

THE <u>HEIGHT</u> OF A ROOTED TREE IS THE MAXIMUM NODE DEPTH. THUS THE HEIGHT OF T ABOVE IS 3.

THE HEIGHT OF A NODE IS HEIGHT OF THE SUBTREE ROOTED AT THAT NODE. EQUIVALENTLY THE HEIGHT OF A NODE IS THE MAXIMUM DISTANCE TO A DESCENDANT LEAF.

NOTE THERE ARE $|V|$ DISTINCT ROOTED TREES ASSOCIATED WITH ONE FREE TREE.

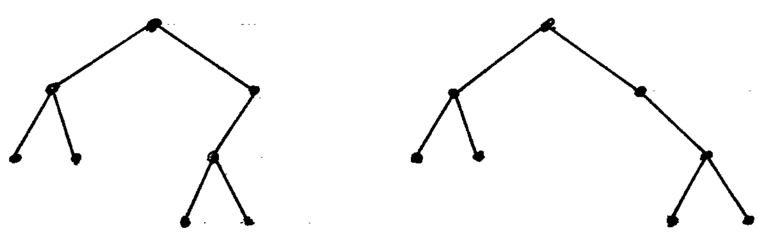
Ex.



THE SAME FREE TREE, BUT A DIFFERENT ROOTED TREE FROM THE PREVIOUS EXAMPLE.

A BINARY TREE IS A ROOTED TREE IN WHICH EACH NODE HAS AT MOST TWO CHILDREN, DESIGNATED AS EITHER THE LEFT CHILD OR RIGHT CHILD.

Ex.



THE SAME ROOTED TREE BUT DIFFERENT BINARY TREES.

RECURSIVE DEFINITION

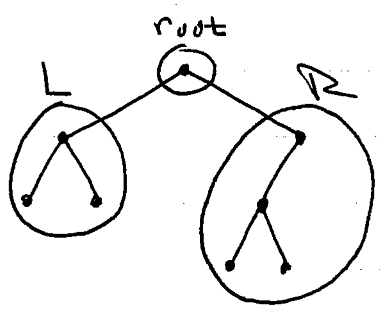
A BINARY TREE T IS A FINITE SET OF NODES WHICH IS EITHER

- EMPTY

OR

- COMPRISED OF 3 DISJOINT SETS OF NODES:
  (1) A ROOT NODE
  (2) A BINARY TREE CALLED THE LEFT SUBTREE $L(T)$
  (3) A BINARY TREE CALLED THE RIGHT SUBTREE $R(T)$

EX



OBSERVE IF $T = \emptyset$ THEN $|T| = 0$, OTHERWISE $|T| = |L| + |R| + 1$. THE HEIGHT OF A BINARY TREE IS DEFINED RECURSIVELY BY THE FORMULA

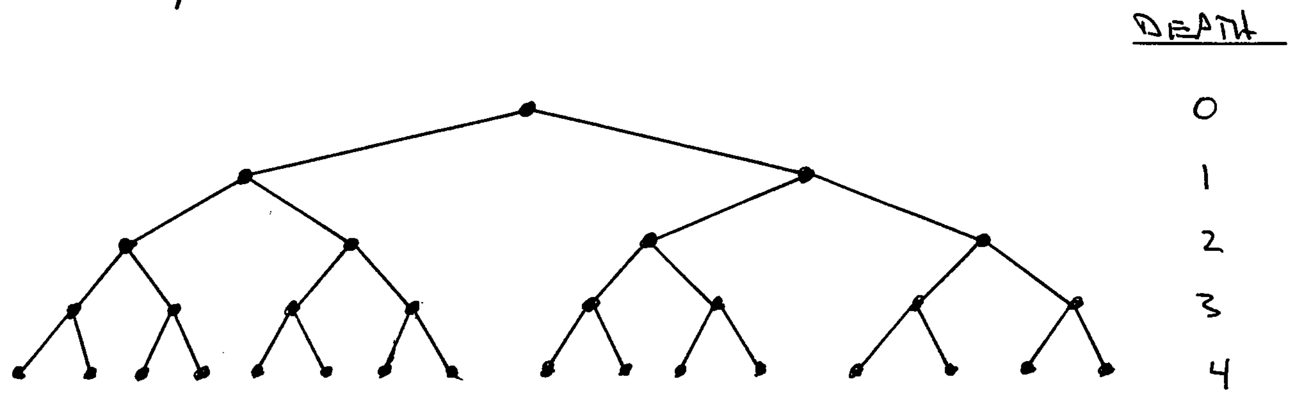
$$h(T) = \begin{cases} -\infty & |T| = 0 \\ 0 & |T| = 1 \\ 1 + \max(h(L), h(R)) & |T| > 1. \end{cases}$$

HINT ON PROBLEM B.5-4: $\left(h(T) \geq \lfloor \lg |T| \rfloor \right)$

USE (AND PROVE) THAT FOR ANY $k \in \mathbb{Z}^+$:

$$\lfloor \lg(2k+1) \rfloor = \lfloor \lg(2k) \rfloor.$$

A <u>COMPLETE BINARY TREE</u> (CBT) IS A BINARY TREE IN WHICH ALL LEAVES HAVE THE SAME DEPTH, AND ALL INTERNAL NODES HAVE EXACTLY TWO CHILDREN.



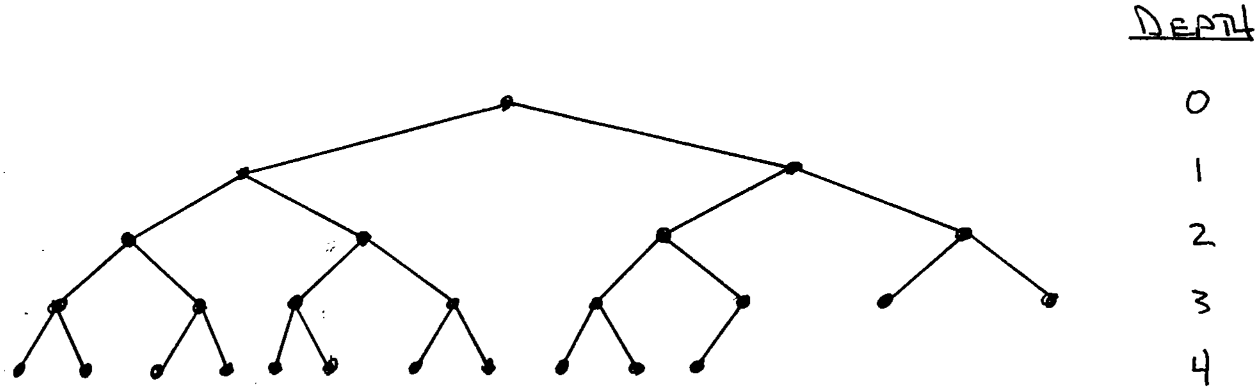
OBSERVE THE NUMBER OF NODES AT DEPTH d IS 2^d. THUS THE TOTAL NUMBER OF NODES IN A CBT OF HEIGHT h IS

$$n = \sum_{d=0}^{h} 2^d = \frac{2^{h+1}-1}{2-1} = 2^{h+1}-1$$

$\therefore$ THE HEIGHT OF A CBT ON n NODES IS

$$h = \lg(n+1) - 1$$

AN <u>ALMOST COMPLETE BINARY TREE</u> (ACBT) IS A BINARY TREE WHICH IS FILLED AT ALL LEVELS, EXCEPT POSSIBLE THE LAST, WHICH MAY BE PARTIALLY FILLED FROM LEFT TO RIGHT.



THE NUMBER OF NODES n IN AN ACBT OF HEIGHT h MUST SATISFY

$$2^h - 1 < n \le 2^{h+1} - 1$$

$$\therefore \quad 2^h \le n < 2^{h+1}$$

$$\therefore \quad h \le \lg n < h+1$$

$$\therefore \quad h = \lfloor \lg n \rfloor$$

<u>EXERCISE</u>

SHOW THAT IF $n = 2^k - 1$ FOR SOME INTEGER $k \ge 1$, THEN

$$\lfloor \lg n \rfloor = \lg(n+1) - 1,$$

HENCE OUR TWO FORMULAS FOR THE HEIGHT OF AN ACBT AND A CBT ON n NODES ARE CONSISTENT.