Notes onTrees

Definition 7.16: A (**free**) **tree** is an undirected simple graph G = (V, E) that is connected and has no non-null simple cycles. If $v \in V$ has degree(v) = 1, then v is a **leaf** node; if degree(v) >1, then v is a **branch** node.

Theorem: an undirected simple graph G = (V, E) is a tree if and only if:
(1) each pair of distinct vertices is connected by a *single* simple path, or
(2) G is connected and the removal of any edge creates a disconnected graph, or
(3) G has no simple cycles and adding any new edge creates a simple cycle, or
(4) G is connected, has e edges and v vertices, and e = v–1.

Proof of (1):
First suppose that condition (1) is true, and each pair of distinct vertices is connected by a single simple path. Then obviously G is connected. To see that G has no non-null simple cycle, the proof is by contradiction. Suppose there is such a cycle, say $v_1, v_2, \dots, v_n, v_1$ ($n \geq 2$). Since this is a simple path, no edges are repeated and therefore there are two paths between v_1 and v_2, namely v_1, v_2 and $v_1, v_n, v_{n-1}, \dots, v_2$. This is a contradiction, and so there can be no non-null simple cycles, and G is a tree.

Conversely, suppose that G is a tree. but has some pair of vertices u and v connected by two (or more) distinct simple paths — say $u, x_1, x_2, \dots, x_p, v$ and $u, y_1, y_2, \dots, y_q, v$. These two paths may agree for some of the initial steps, but will eventually differ since they are different paths. Hence $x_1 = y_1, x_2 = y_2, \dots, x_m = y_m$, and $x_{m+1} \neq y_{m+1}$ for some $m \geq 0$. Then the path □ = $x_m, x_{m+1}, \dots, x_p, v, y_q, y_{q-1}, \dots, y_{m+1}, y_m$ is a cycle ($y_m = x_m$). This may not be a simple cycle, but since $x_{m+1} \neq y_{m+1}$ it must contain a simple cycle. If □ is not a simple cycle let i and j be the largest values, $m+1 < i \leq p$ and $m+1 < j \leq q$ so that $x_i = y_j$. Then $x_m, x_{m+1}, \dots, x_i, y_{j+1}, y_{j+2}, \dots, y_{m+1}, y_m$ is a simple subcycle of □ and is non-null since $x_{m+1} \neq y_{m+1}$. But G has no non-null simple cycles, so this is a contradiction. Hence it is impossible for G to have a pair of nodes connected by two distinct simple paths.

Proof of (2):
First suppose that condition (2) is true, and removal of any edge disconnects G. We prove G has no simple cycles by contradiction. Suppose G contains some simple cycle, say $v_1, v_2, \dots, v_n, v_1$ ($n \geq 2$). But then removal of edge (v_1, v_2) does not disconnect G since in any path that uses it, (v_1, v_2) can be replaced by $v_1, v_n, v_{n-1}, \dots, v_2$ and a path still exists so G is still connected, a contradiction.

Conversely, suppose that G is a tree. For any edge (v_1, v_2), by condition (1), this is the only path between v_1 and v_2. Therefore when this edge is removed, G becomes disconnected.

Proof of (4):
First suppose that condition (4) is true. We prove that G has no simple cycles by contradiction. So suppose that G has a simple cycle, and either this cycle is elementary or it contains an elementary subcycle -- let this elementary cycle be C(v.v) and have length p. An elementary cycle of length p contains p vertices. Consider the other v–p vertices of G. Each of these vertices is reachable from v, and therefore there must be at least one edge not appearing in C to each of them. This yields at least p + (v–p) = v edges, a contradiction. Hence G cannot contain a simple cycle and is a tree.

Conversely, suppose that G is a tree. The proof that e = v-1 is by induction on the number of edges.
Basis case: e = 0 -- with no edges, we must have v = 1 for G to be connected.
Induction step:
Assume that e = v-1 for any tree with n or fewer edges (the inductive hypothesis), and let G be a tree with e = n+1 edges -- we need to show G has e+1 = n+2 vertices. Pick any edge (x_1, x_2) of G and consider the graph G' = (V, E–(x_1, x_2)) with that edge removed. Since G is a tree, by property (1) there is no path from

x_1 to x_2 after this edge is removed. Therefore, G' is disconnected and there are two connected components, call them G_1 and G_2, one containing x_1 and the other containing x_2. Each of these connected components is a subgraph of G and so can have no simple cycles and is therefore a tree. Suppose that G_i has e_i edges and v_i vertices, for i = 1,2. Then since G_1 and G_2 arose by the deletion of a single edge from G, it must be that $e-1 = e_1+e_2$, and $v = v_1+v_2$. Now since G_1 and G_2 are both trees with n or fewer edges, by the induction hypothesis, $e_1 = v_1-1$ and $e_2 = v_2-1$. Therefore $e = e_1+e_2+1 = v_1-1 + v_2-1 + 1 = v_1+v_2-1 = v-1$ and the induction is extended, completing the induction proof.

Definition: A (non-oriented) **rooted tree** consists of a pair (G,r) where G = (V, E) is a tree and $r \in V$ (the **root**). The **level** of a node in a rooted tree is the length of the simple path from the root to the node. The **height** of a tree is the maximum level of any of its vertices. If v is a branch node of a rooted tree, then all those adjacent nodes whose level is 1 greater than that of v are called **child** nodes of v and v is called their **parent** node; if two nodes have the same parent they are called **siblings**. If node v lies on the path from the root to node u, then v is an **ancestor** of u and u is a **descendant** of v.

Grassman & Tremblay regard rooted trees as *directed* graphs. This differs from the definition given immediately above, but it's primarily a matter of interpretation. In a tree, there is a unique path from the root to any other node -- Grassman & Tremblay regard all these paths to be oriented away from the root. Commonly we understand path questions in rooted trees to emanate from the root so an explicit path orientation can be omitted. Therefore the non-oriented definition of a rooted tree stated above is common with many other authors.