# Chapter 16

# Trees

## 16.1 Basic properties of trees

Trees in mathematics are graphs of a certain kind. In a sense, trees are the simplest interesting graphs, in that they have a very simple structure, but possess a rich variety of nontrivial properties. Trees have innumerable applications throughout computer science.

**Definition 16.1.1.** *A* tree *is a connected graph with no cycles. A vertex of degree one in a tree is called a* leaf*.*

An extensive theory of trees has been developed, and we give the tip of the iceberg below: Four additional characterizations that each could have been used to define trees.

**Theorem 16.1.2.** *Given a graph $G = (V, E)$, the following conditions are equivalent:*

*(a) $G$ is a connected graph with no cycles. (Thus $G$ is a tree by the above definition.)*

*(b) For every two vertices $u, v \in V$, there exists exactly one path from $u$ to $v$.*

*(c) $G$ is connected, and removing any edge from $G$ disconnects it. (Thus $G$ is a* minimal connected graph*.)*

*(d) $G$ has no cycles, and adding any edge to $G$ gives rise to a cycle. (Thus $G$ is a* maximal acyclic graph*.)*

*(e) $G$ is connected and $|E| = |V| - 1$.*

*Proof.* We will prove for each of the conditions (b)–(e) in turn that it is equivalent to condition (a). This implies the equivalence of all the conditions. The proof proceeds by induction on the number of vertices $|V|$ in $G$, and we relate a tree with $n + 1$ vertices to a tree with $n$ vertices in the inductive step by "tearing off" a leaf. We begin by proving two lemmas that will be useful in this process.

**Lemma 16.1.3.** *Each tree with at least 2 vertices contains at least 2 leaves.*

*Proof.* Given a tree $T = (V, E)$, consider a path $P$ of maximum length in $T$. We claim that the two end-points of $P$ are leaves of $T$. Indeed, assume for the sake of contradiction that an end-vertex $u$ of $P$ has degree greater than 1 in $T$. Thus there exists an edge $\{u, u'\} \in E$ that is not part of $P$. If $u'$ belongs to $P$ then $T$ contains a cycle. Otherwise we can extend $P$ by the edge $\{u, u'\}$ and $P$ is not a longest path in $T$. This contradiction proves the lemma. □

**Lemma 16.1.4.** *Given a graph $G = (V, E)$ and a leaf $v \in V$ that is incident to an edge $e = \{v, v'\} \in E$, the graph $G$ is a tree if and only if $G' = (V \setminus \{v\}, E \setminus \{e\})$ is a tree.*

*Proof.* Assume that $G$ is a tree and consider two vertices $u, w \in V \setminus \{v\}$. $u$ and $w$ are connected by a path $P$ in $G$. Every vertex of $P$ other than $u$ and $w$ has degree at least 2, and thus $v$ cannot be a vertex of $P$. Therefore $P$ is a path in $G'$, which proves that $G'$ is connected. Since $G$ does not contain a cycle, $G'$ cannot contain a cycle and is thus a tree.

For the other direction, assume that $G'$ is a tree. Since a cycle only contains vertices with degree at least 2, a cycle in $G$ must also be a cycle in $G'$. Therefore there are no cycles in $G$. Also, any two vertices of $G$ other than $v$ can be connected by the same path as in $G'$, and $v$ can be connected to any vertex $u$ in $G$ by a path that consists of the edge $e$ and a path in $G'$ between $v'$ and $u$. Thus $G$ is a tree. □

We are now ready to employ induction to prove that condition (a) implies each of (b)–(e). For the induction basis, all five conditions hold for the graph with a single vertex. Consider a graph $G = (V, E)$ with $|V| = n \geq 2$ and assume that (a) holds for $G$. By Lemma 16.1.3, $G$ has a leaf $v \in V$ that is incident to an edge $e = \{v, v'\} \in E$. By Lemma 16.1.4, condition (a) holds for $G'$. The inductive hypothesis states that condition (a) implies conditions (b)–(e) for the graph $G' = (V \setminus \{v\}, E \setminus \{e\})$. We now need to prove that conditions (b)–(e) also hold for $G$.

Condition (b) holds for $G$ by a similar argument to the one employed in the proof of Lemma 16.1.4. Condition (c) holds for $G$ since removing any edge other than $e$ disconnects $G$ by the induction hypothesis, and removing $e$ disconnects the vertex $v$ from the rest of the graph. Condition (d) holds since $G$ cannot have cycles by an argument similar to the proof of Lemma 16.1.4; adding an edge that is not incident to $v$ creates a cycle by the inductive hypothesis, and adding an edge $\{v, u\}$, for some $u \in V \setminus \{v\}$ creates a cycle that consists of the edge $\{v, u\}$, the path from $u$ to $v'$, and the edge $e$. Finally, condition (e) holds since $G$ is obtained from $G'$ by adding one vertex and one edge.

We now prove that each of (b)–(d) imply (a). Conditions (b) and (c) on $G$ each imply connectedness of $G$. By contrapositive, assume that $G$ contains a cycle. Then taking two distinct vertices $u, w$ on the cycle, there are two paths from $u$ to $w$ along the cycle, which implies $(b) \Rightarrow (a)$. Furthermore, removing one edge of the cycle does not disconnect $G$, which implies $(c) \Rightarrow (a)$. Condition (d) implies that $G$ does not contain a cycle. By contrapositive, assume that $G$ is disconnected. Then there are two vertices $u$ and $w$ that have no path connecting them and we can add the edge $\{u, w\}$ to $G$ without creating a cycle. This implies $(d) \Rightarrow (a)$.

To prove $(e) \Rightarrow (a)$ we use induction on the number of vertices of $G$. The induction basis is the graph with one vertex and the claim trivially holds. For the induction hypothesis, assume that the claim holds for all graphs with $|V| - 1$ vertices. For the inductive step, assume that condition (e) holds for $G$ and hence $|E| = |V| - 1$. Therefore the sum of the degrees of the vertices of $G$ is $2|V| - 2$, and thus there is some vertex $v \in V$ of degree 1. The graph $G' = (V \setminus \{v\}, E \setminus \{e\})$ is connected and satisfies $|E \setminus \{e\}| = |V \setminus \{v\}| - 1$. By the induction hypothesis, $G'$ is a tree. Lemma 16.1.4 now implies that $G$ is a tree, which completes the proof. □

## 16.2 Spanning trees

One of the reasons that trees are so pervasive is that every connected graph $G$ contains a subgraph that is a tree on all of the vertices of $G$. Such a subgraph is called a *spanning tree* of $G$.

**Definition 16.2.1.** *Consider a graph $G = (V, E)$. A tree of the form $(V, E')$, where $E' \subseteq E$ is called a spanning tree of $G$.*

**Proposition 16.2.2.** *Every connected graph $G = (V, E)$ contains a spanning tree.*

*Proof.* Consider a tree subgraph $T = (V', E')$ of $G$ with the largest number of vertices. Suppose for the sake of contradiction that $V' \neq V$, and thus there exists $v \in V \setminus V'$. Take an arbitrary vertex $u \in V'$ and consider a path $P$ between $v$ and $u$. Let $u'$ be the first vertex along $P$ that belongs to $V'$, and let $v'$ be the vertex that immediately precedes $u'$ in $P$. Consider the graph $T' = (V' \cup \{v'\}, E' \cup \{v', u'\})$. Lemma 16.1.4 implies that $T'$ is a tree in $G$ with a greater number of vertices than $T$, which is a contradiction. □