Lecture 6

Trees and forests

This section of the notes introduces an important family of graphs—trees and forests—and also serves as an introduction to inductive proofs on graphs.

Reading:
The material in today's lecture comes from Section 1.2 of

> Dieter Jungnickel (2013), *Graphs, Networks and Algorithms*, 4th edition (available online via `SpringerLink`),

though the discussion there includes a lot of material about counting trees that we'll handle in a different way.

6.1 Basic definitions

We begin with a flurry of definitions.

Definition 6.1. *A graph $G(V, E)$ is **acyclic** if it doesn't include any cycles.*

Another way to say a graph is acyclic is to say that it contains no subgraphs isomorphic to one of the cycle graphs.

Definition 6.2. *A **tree** is a connected, acyclic graph.*

Definition 6.3. *A **forest** is a graph whose connected components are trees.*

Trees play an important role in many applications: see Figure 6.1 for examples.

6.1.1 Leaves and internal nodes

Trees have two sorts of vertices: *leaves* (sometimes also called *leaf nodes*) and *internal nodes*: these terms are defined more carefully below and are illustrated in Figure 6.2.

Definition 6.4. *A vertex $v \in V$ in a tree $T(V, E)$ is called a **leaf** or **leaf node** if* $\deg(v) = 1$ *and it is called an **internal node** if* $\deg(v) > 1$.

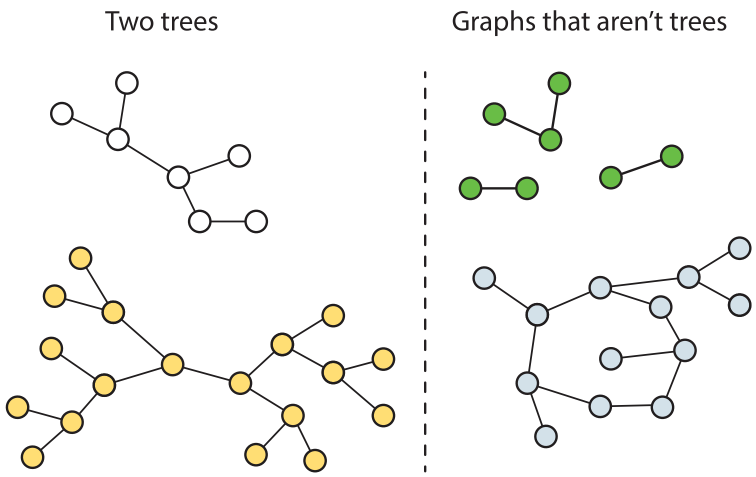

Figure 6.1: *The two graphs at left (white and yellow vertices) are trees, but the two at right aren't: the one at upper right (with green vertices) has multiple connected components (and so it isn't connected) while the one at lower right (blue vertices) contains a cycle. The graph at upper right is, however, a forest as each of its connected components is a tree.*

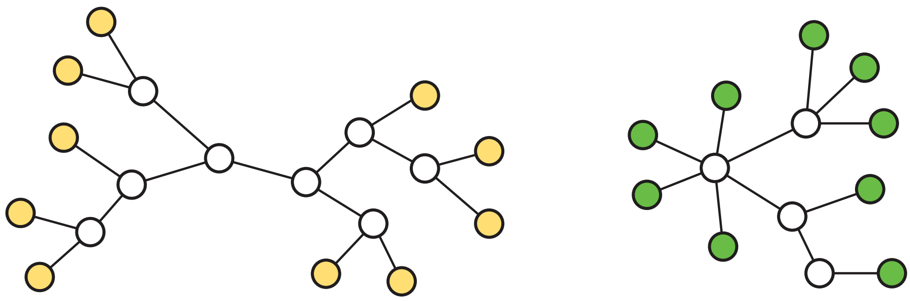

Figure 6.2: *In the two trees above the internal nodes are white, while the leaf nodes are coloured green or yellow.*

6.1.2 Kinds of trees

Definition 6.5. *A **binary tree** is a tree in which every internal node has degree three.*

Definition 6.6. *A **rooted tree** is a tree with a distinguished leaf node called the* root node*.*

Warning to the reader: The definition of rooted tree above is common among biologists, who use trees to represent evolutionary lineages (see Darwin's sketch at right in Figure 6.3). Other researchers, especially computer scientists, use the same term to mean something slightly different.

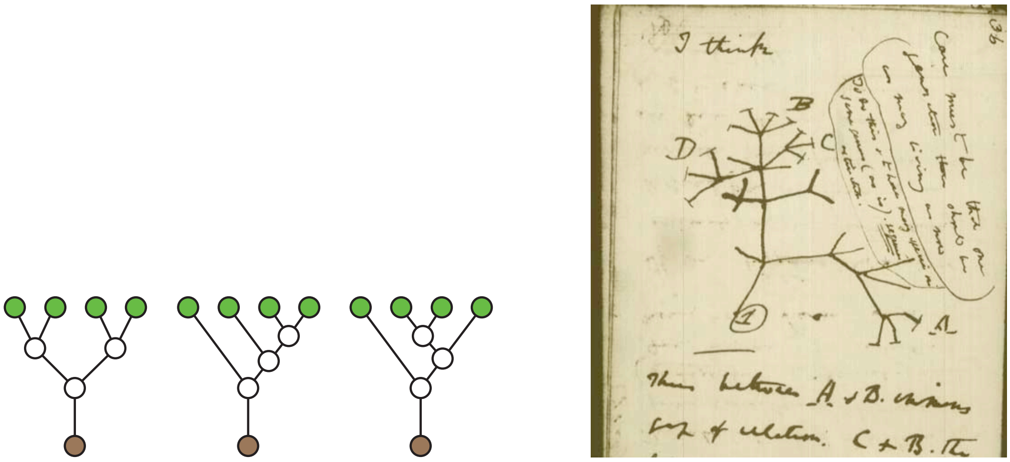

Figure 6.3: *At left are three examples of rooted binary trees. In all cases the root node is brown, the leaves are green and the internal nodes are white. At right is a page from one of Darwin's notebooks, showing the first known sketch of an evolutionary tree: here the nodes represent species and the edges indicate evolutionary descent.*

6.2 Three useful lemmas and a proposition

Lemma 6.7 (Minimal $|E|$ in a connected graph)**.** *A connected graph on n vertices has at least $(n-1)$ edges.*

Lemma 6.8 (Maximal $|E|$ in an acyclic graph)**.** *An acyclic graph on n vertices has at most $(n-1)$ edges.*

Definition 6.9. *A vertex v is said to be **isolated** if it has no neighbours. Equivalently, v is isolated if $\deg(v) = 0$.*

Lemma 6.10 (Vertices of degree 1)**.** *If a graph $G(V, E)$ has $n \geq 2$ vertices, none of which are isolated, and $(n-1)$ edges then G has at least two vertices of degree 1.*

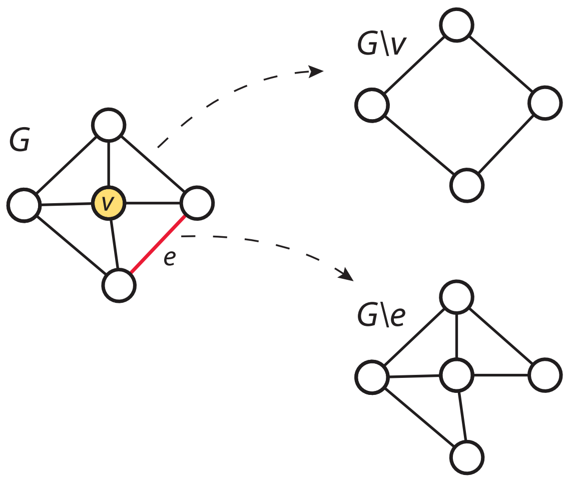

Figure 6.4: *A graph $G(V, E)$ and the subgraphs $G\backslash v$ formed by deleting the yellow vertex v and $G\backslash e$ formed by deleting the red edge e.*

6.2.1 A festival of proofs by induction

Proofs by induction about graphs generally have three parts

- a **base case** that typically involves a graph with very few vertices or edges (often just one or two) and for which the result is obvious;
- an **inductive hypothesis** in which one assumes the result is true for all graphs with, say, n_0 or fewer vertices (or perhaps m_0 or fewer edges);
- an **inductive step** where one starts with a graph that satisfies the hypotheses of the theorem and has, say, $n_0 + 1$ vertices (or $m_0 + 1$ edges or whatever is appropriate) and then *reduces* the theorem as it applies to this larger graph to something involving smaller graphs (to which the inductive hypothesis applies), typically by *deleting* an edge or vertex.

6.2.2 Graph surgery

The proofs below accomplish their inductive steps by deleting either an edge or a vertex, so here I introduce some notation for these processes.

Definition 6.11. *If $G(V, E)$ is a graph and $v \in V$ is one of its vertices then $G\backslash v$ is defined to be the subgraph formed by deleting v and all the edges that are incident on v.*

Definition 6.12. *If $G(V, E)$ is a graph and $e \in E$ is one of its edges then $G\backslash e$ is defined to be the subgraph formed by deleting e.*

Both these definitions are illustrated in Figure 6.4.

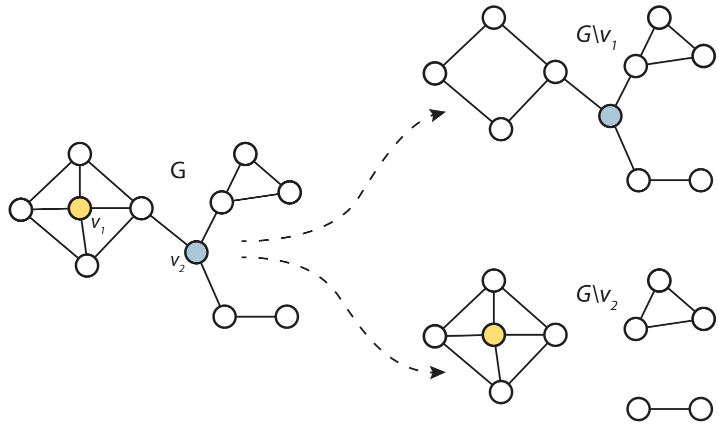

Figure 6.5: *In the inductive step of the proof of Lemma 6.7 we delete some arbitrary vertex $v \in V$ in a connected graph $G(V, E)$ to form the graph $G \backslash v$. The result may still be a connected graph, as in $G \backslash v_1$ at upper right, or may fall into several connected components, as in $G \backslash v_2$ at lower right.*

Proof of Lemma 6.7

We'll prove Lemma 6.7 by induction on the number of vertices. First let us rephrase the lemma in an equivalent way:

> If $G(V, E)$ is a connected graph on $|V| = n$ vertices, then $|E| \geq n - 1$.

Base case: There is only one graph with $|V| = 1$ and it is, by definition, connected and has $|E| = 0$, which satisfies the lemma. One could alternatively start from K_2, which is the only connected graph on two vertices and has $|E| = 1$.

Inductive hypothesis: Suppose that the lemma is true for all graphs $G(V, E)$ with $1 \leq |V| \leq n_0$, for some fixed n_0.

Inductive step: Now consider a connected graph $G(V, E)$ with $|V| = n_0 + 1$: the lemma we're trying to prove then says $|E| \geq n_0$. Choose some vertex $v \in V$ and delete it, forming the graph $G \backslash v$. We'll say that the new graph has vertex set $V' = V \backslash v$ and edge set E'. There are two possibilities (see Figure 6.5):

(i) $G \backslash v$ is still a connected graph;

(ii) $G \backslash v$ has $k \geq 2$ connected components: call these $G_1(V_1, E_1), \ldots, G_k(V_k, E_k)$.

In the first case—where $G \backslash v$ is connected—we also know $|V'| = |V| - 1 = n_0$ and so the inductive hypothesis applies and tells us that $|E'| \geq (n_0 - 1)$. But as G was connected, the vertex v that we deleted must have had at least one neighbour, and hence at least one edge, so we have

$$|E| \;\geq\; |E'| + 1 \;\geq\; (n_0 - 1) + 1 \;\geq\; n_0$$

which is exactly the result we sought.

In the second case—where deleting v causes G to fall into $k \geq 2$ connected components—we can call the components $G_1(V_1, E_1), G_2(V_2, E_2), \cdots, G_k(V_k, E_k)$ with $n_j = |V_j|$. Then

$$\sum_{j=1}^{k} n_j = \sum_{j=1}^{k} |V_j| = |V'| = |V| - 1 = n_0.$$

Further, the j-th connected component is a connected graph on $n_j < n_0$ vertices and so the inductive hypothesis applies to each component separately, telling us that $|E_j| \geq n_j - 1$. But then we have

$$|E'| = \sum_{j=1}^{k} |E_j| \geq \sum_{j=1}^{k} (n_j - 1) \geq \left(\sum_{j=1}^{k} n_j \right) - k \geq n_0 - k. \tag{6.1}$$

And, as we know that the original graph G was connected, we also know that the deleted vertex v was connected by at least one edge to each of the k components of $G \backslash v$. Combining this observation with Eqn. (6.1) gives us

$$|E| \geq |E'| + k \geq (n_0 - k) + k \geq n_0,$$

which proves the lemma for the second case too.

Proof of Lemma 6.8

Once again, we'll do induction on the number of vertices. As above, we begin by rephrasing the lemma:

If $G(V, E)$ is an acyclic graph on $|V| = n$ vertices, then $|E| \leq n - 1$.

Base case: Either K_1 or K_2 could serve as the base case: both are acyclic graphs that have a maximum of $|V| - 1$ edges.

Inductive hypothesis: Suppose that Lemma 6.8 is true for all acyclic graphs with $|V| \leq n_0$, for some fixed n_0.

Inductive step: Consider an acyclic graph $G(V, E)$ with $|V| = n_0 + 1$: we want to prove that $|E| \leq n_0$. Choose an arbitrary edge $e = (a, b) \in E$ and delete it to form $G'(V, E') = G \backslash e$, which has the same vertex set as G, but a smaller edge set $E' = E \backslash e$.

First note that G' must have one more connected component than G does because a and b, the two vertices that appear in the deleted edge e, are connected in G, but cannot be connected in G'. If they were still connected, there would (by Prop. 5.17) be a path connecting them in G' that, when combined with e, would form a cycle in G, contradicting the assumption that G is acyclic. Thus we know that G' has $k \geq 2$ connected components that we can call $G_1(V_1, E_1), \ldots, G_k(V_k, E_k)$.

If we again define $n_j = |V_j|$, we know that $n_j \leq n_0$ for all j and so the inductive hypothesis applies to each component separately: $|E_j| \leq n_j - 1$. Adding these up yields

$$|E'| \;=\; \sum_{j=1}^{k} |E_j| \;\leq\; \sum_{j=1}^{k} (n_j - 1) \;\leq\; \left(\sum_{j=1}^{k} n_j \right) - k \;\leq\; (n_0 + 1) - k.$$

And then, as $|E| = |E'| + 1$ we have

$$|E| \;=\; |E'| + 1 \;\leq\; (n_0 + 1 - k) + 1 \;\leq\; n_0 + (2 - k) \;\leq\; n_0,$$

where the final inequality follows from the observation that G' has $k \geq 2$ connected components.

Proof of Lemma 6.10

The final Lemma in this section is somewhat technical: we'll use it in the proof of a theorem in Section 6.3. The lemma says that graphs $G(V, E)$ that have $|V| = n$ and $|E| = (n - 1)$ and have no isolated vertices must contain at least two vertices with degree one. The proof is by contradiction and uses the Handshaking Lemma.

Imagine the vertices are numbered and arranged in order of increasing degree so $V = \{v_1, \ldots, v_n\}$ and $\deg(v_1) \leq \deg(v_2) \leq \cdots \leq \deg(v_n)$. The Handshaking Lemma then tells us that

$$\sum_{j=1}^{n} \deg(v_j) \;=\; 2|E| \;=\; 2(n-1) \;=\; 2n - 2. \tag{6.2}$$

As there are no isolated vertices, we also know that $\deg(v_j) \geq 1$ for all j. Now assume—aiming for a contradiction—that there is at most a single vertex with degree one. That is, assume $\deg(v_1) \geq 1$, but $\deg(v_j) \geq 2 \;\; \forall j \geq 2$. Then

$$\begin{aligned}
\sum_{j=1}^{n} \deg(v_j) &= \deg(v_1) + \sum_{j=2}^{n} \deg(v_j) \\
&\geq 1 + \sum_{j=2}^{n} 2 \\
&\geq 1 + (n-1) \times 2 \\
&\geq 2n - 1.
\end{aligned}$$

This contradicts Eqn. (6.2), which says that the sum of degrees is $2n - 2$. Thus it must be true that two or more vertices have degree one, which is the result we sought.

6.3 A theorem about trees

The lemmas of the previous section make it possible to give several nice characterisations of a tree and the theorem below, which has a form that one often finds in Discrete Maths or Algebra books, shows that they're all equivalent.

Theorem 6.13 (Jungnickel's Theorem 1.2.8)**.** *For a graph $G(V,E)$ on $|V| = n$ vertices, any two of the following imply the third:*

(a) G is connected.

(b) G is acyclic.

(c) G has $(n-1)$ edges.

6.3.1 Proof of the theorem

The theorem above is really three separate propositions bundled into one statement: we'll prove them in turn.

(a) and (b) $\implies$ (c)

On the one hand, our lemma about the minimal number of edges in a connected graph (Lemma 6.7) says that property (a) implies that $|E| \geq (n-1)$. On the other hand our lemma about the maximal number of edges in an acyclic graph (Lemma 6.8) says $|E| \leq (n-1)$. The only possibility compatible with both these inequalities is $|E| = (n-1)$.

(a) and (c) $\implies$ (b)

To prove this by contradiction, assume that it's possible to have a connected graph $G(V,E)$ that has $(n-1)$ edges and contains a cycle. Choose some edge e that's part of the cycle and delete it to form $H = G\backslash e$. H is then a connected graph (removing an edge from a cycle does not change the number of connected components) with only $n-2$ edges, which contradicts our earlier result (Lemma 6.7) about the minimal number of edges in a connected graph.

(b) and (c) $\implies$ (a)

We'll prove—by induction on $n = |V|$—that an acyclic graph with $|V| - 1$ edges must be connected.

Base case: There is only one graph with $|V| = 1$. It's acyclic, has $|V| - 1 = 0$ edges and is connected.

Inductive hypothesis: Suppose that all acyclic graphs with $1 \leq |V| \leq n_0$ vertices and $|E| = |V| - 1$ edges are connected.

Inductive step: Now consider an acyclic graph $G(V,E)$ with $|V| = n_0 + 1$ and $|E| = n_0$: we need to prove that it's connected. First, notice that such a graph cannot have any isolated vertices, for suppose there was some vertex v with $\deg(v) = 0$. We could then delete v to produce $H = G\backslash v$, which would be an acyclic graph with n_0 vertices and n_0 edges, contradicting our lemma (Lemma 6.8) about the maximal number of edges in an acyclic graph.

Thus G contains no isolated vertices and so, by the technical lemma from the previous section (Lemma 6.10), we know that it has at least two vertices of degree one. Say that one of these two is $u \in V$ and delete it to make $G'(V', E') = G \backslash u$. Then G' is still acyclic, because G is, and deleting vertices can't create cycles. Furthermore G' has $|V'| = |V| - 1 = n_0$ vertices and $|E'| = |E| - 1 = n_0 - 1$ edges. This means that the inductive hypothesis applies and we can conclude that G' is connected. But if G' is connected, so is G and we are finished.