Trees
COMS20017 (Algorithms and Data)

John Lapinskas, University of Bristol

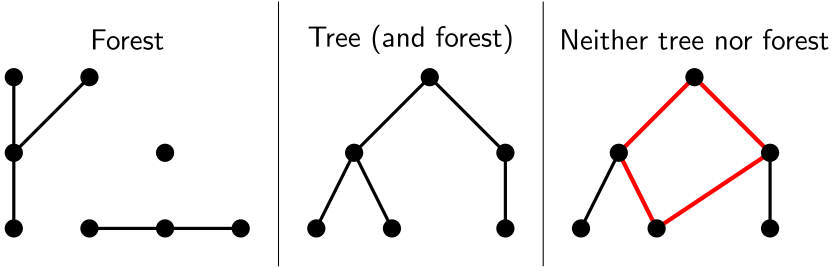
Trees

In COMS10007, you used (rooted) trees to model heaps, recursion, and the decisions of comparison-based sorting algorithms.

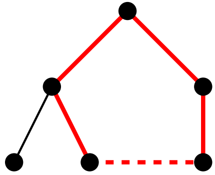
In this course, we will think of trees as examples of graphs.

We define a **forest** to be a graph which contains no cycles,
and a **tree** to be a **connected** graph with no cycles.

(So the components of a forest are trees, and all trees are forests!)

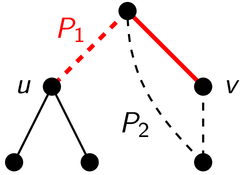


Notice how any edge we add to the tree from the last slide forms a cycle.



Notice how any edge we add to the tree from the last slide forms a cycle.

This is not a coincidence!

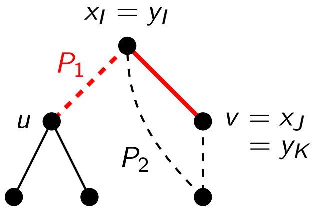


Lemma: If $T = (V, E)$ is a tree, then any pair of vertices $u, v \in V$ is joined by a **unique** path uTv in T.

Proof: T is connected, so there is a path $P_1 = x_0 \dots x_k$ from u to v. Suppose there is another path $P_2 = y_0 \dots y_k$ from u to v.

Notice how any edge we add to the tree from the last slide forms a cycle.

This is not a coincidence!



Lemma: If $T = (V, E)$ is a tree, then any pair of vertices $u, v \in V$ is joined by a **unique** path uTv in T.

Proof: T is connected, so there is a path $P_1 = x_0 \dots x_k$ from u to v. Suppose there is another path $P_2 = y_0 \dots y_k$ from u to v.

Then P_1 and P_2 must diverge from each other and come back together. Let $I = \min\{i : x_i \neq y_i\} - 1$ be the point of divergence. Let $J = \min\{i > I : x_i \in \{y_I, \dots, y_k\}\}$ be the point of remerging. Let K be the corresponding point on P_2, so $y_K = x_J$.

Then $x_I x_{I+1} \dots x_J y_{K-1} y_{K-2} \dots y_I$ is a cycle, so T is not a tree. $\square$

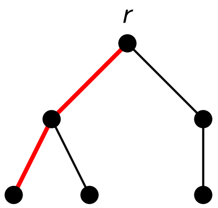
A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. □

Lemma 2: Any n-vertex tree has $n-1$ edges.

Proof: We start by showing how to turn a tree $T=(V,E)$ into a **rooted tree**, like those you worked with last year.

Let $r \in V$ be arbitrary — this will be the **root**. Every vertex $v \neq r$ has a unique path P_v from r to v by the lemma. Direct its edges from r to v.



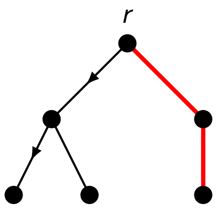
A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. □

Lemma 2: Any n-vertex tree has $n-1$ edges.

Proof: We start by showing how to turn a tree $T = (V, E)$ into a **rooted tree**, like those you worked with last year.

Let $r \in V$ be arbitrary — this will be the **root**. Every vertex $v \neq r$ has a unique path P_v from r to v by the lemma. Direct its edges from r to v.



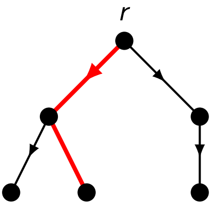
A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. □

Lemma 2: Any n-vertex tree has $n - 1$ edges.

Proof: We start by showing how to turn a tree $T = (V, E)$ into a **rooted tree**, like those you worked with last year.

Let $r \in V$ be arbitrary — this will be the **root**. Every vertex $v \neq r$ has a unique path P_v from r to v by the lemma. Direct its edges from r to v.



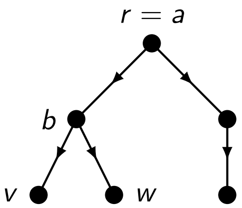
A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. □

Lemma 2: Any n-vertex tree has $n-1$ edges.

Proof: We start by showing how to turn a tree $T=(V,E)$ into a **rooted tree**, like those you worked with last year.

Let $r \in V$ be arbitrary — this will be the **root**. Every vertex $v \neq r$ has a unique path P_v from r to v by the lemma. Direct its edges from r to v.



Why are the directions consistent?

Suppose some path P_v directs $a \to b$.
And suppose b is also on another path P_w.

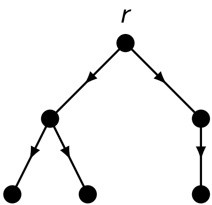
Then both P_v and P_w must start with P_b, since P_b is the **unique** path from r to b.
So P_w also directs $a \to b$. ✓

A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. □

Lemma 2: Any n-vertex tree $T = (V, E)$ has $n - 1$ edges.

Proof idea: Take an arbitrary root $r \in V$. For all vertices v, let P_v be the unique path from r to v. Direct T's edges along these paths. ✓



Because these paths are unique, every vertex other than r has in-degree 1, and r has in-degree 0.

So by the directed handshake lemma: $|E| = \sum_{v \in V} d^-(v) = n - 1$. □

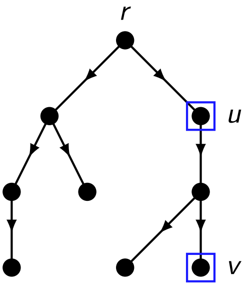
Bonus: We also just defined rooted trees in terms of graphs.

A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. □

Lemma 2: Any n-vertex tree has $n-1$ edges. □

We **root** a tree $T = (V, E)$ at $r \in V$ as follows. For all vertices $v \neq r$, let P_v be the unique path from r to v. Then direct each P_v from r to v.



In a **rooted tree** with root r:

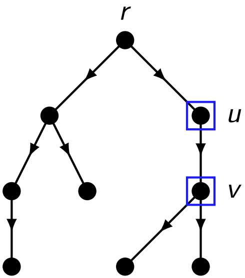
- u is an **ancestor** of v (or v is a **descendant** of u) if u is on P_v.

A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. □

Lemma 2: Any n-vertex tree has $n-1$ edges. □

We **root** a tree $T=(V,E)$ at $r \in V$ as follows. For all vertices $v \neq r$, let P_v be the unique path from r to v. Then direct each P_v from r to v.



In a **rooted tree** with root r:

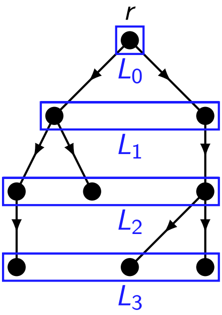
- u is an **ancestor** of v (or v is a **descendant** of u) if u is on P_v.
- u is the **parent** of v (or v is a **child** of u) if $u \in N^-(v)$.

A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. □

Lemma 2: Any n-vertex tree has $n-1$ edges. □

We **root** a tree $T = (V, E)$ at $r \in V$ as follows. For all vertices $v \neq r$, let P_v be the unique path from r to v. Then direct each P_v from r to v.



In a **rooted tree** with root r:

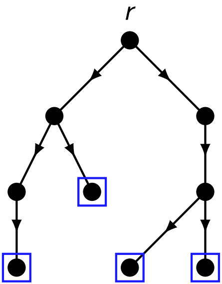
- u is an **ancestor** of v (or v is a **descendant** of u) if u is on P_v.
- u is the **parent** of v (or v is a **child** of u) if $u \in N^-(v)$.
- The first **level** L_0 of T is $\{r\}$, and $L_{i+1} = N^+(L_i)$.
- The **depth** of T is $\max\{i \colon L_i \neq \emptyset\}$, e.g. this tree has depth 3.

A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. □

Lemma 2: Any n-vertex tree has $n-1$ edges. □

We **root** a tree $T=(V,E)$ at $r \in V$ as follows. For all vertices $v \neq r$, let P_v be the unique path from r to v. Then direct each P_v from r to v.



In a **rooted tree** with root r:

- u is an **ancestor** of v (or v is a **descendant** of u) if u is on P_v.
- u is the **parent** of v (or v is a **child** of u) if $u \in N^-(v)$.
- The first **level** L_0 of T is $\{r\}$, and $L_{i+1} = N^+(L_i)$.
- The **depth** of T is $\max\{i : L_i \neq \emptyset\}$, e.g. this tree has depth 3.

In a **rooted** tree: A **leaf** is a vertex with no children.
In a **non-rooted** tree: A **leaf** is a degree-1 vertex.

These definitions agree except for a rooted tree whose root has one child.

A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. $\square$

Lemma 2: Any n-vertex tree has $n-1$ edges. $\square$

A **leaf** is a degree-1 vertex.

Lemma 3: Any n-vertex tree $T=(V,E)$ with $n \geq 2$ has at least 2 leaves.

Proof: Let x be the number of leaves in T.

By the handshaking lemma, $|E| = \frac{1}{2}\sum_{v \in V} d(v)$. Also, $|E| = n-1$.

Since T is connected and $n \geq 2$, every vertex has degree at least 1.

So all non-leaves have degree at least 2, and $\sum_{v \in V} d(v) \geq 2(n-x)+x$.

Plugging this in gives $|E| = n-1 = \frac{1}{2}\sum_{v \in V} d(v) \geq n - \frac{x}{2}$.

Solving for x gives $x \geq 2$, so we're done! $\square$

The Fundamental Lemma of Trees

A **tree** is a connected graph with no cycles.

Lemma 1: Any pair of vertices in a tree is joined by a **unique** path. □

Lemma 2: Any n-vertex tree has $n-1$ edges. □

When you're actually working with trees, it's good to have one single result that tells you that all the "obvious" things are true. This is that result.

Lemma: The following are equivalent for an n-vertex graph $T = (V, E)$:

(A) T is connected and has no cycles, i.e. is a tree;

(B) T has $n-1$ edges and is connected;

(C) T has $n-1$ edges and has no cycles;

(D) T has a unique path between any pair of vertices.

We've already proved (A) $\Rightarrow$ (D) (Lemma 1)...

as well as (A) $\Rightarrow$ (B) and (A) $\Rightarrow$ (C) (Lemma 2).

Lemma: The following are equivalent for an n-vertex graph $T = (V, E)$:

(A) T is connected and has no cycles, i.e. is a tree;

(B) T has $n - 1$ edges and is connected;

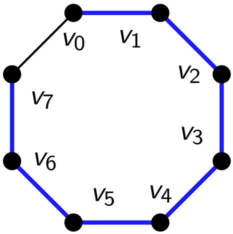
(C) T has $n - 1$ edges and has no cycles;

(D) T has a unique path between any pair of vertices.

(A) $\Rightarrow$ (B), (C) and (D): ✓

(D) $\Rightarrow$ (A): T has a path between any pair of vertices, so it's connected.

And on any cycle $v_0 \dots v_k$, there are two different paths from v_0 to v_k:



- the path $v_0 \dots v_k$; and

Lemma: The following are equivalent for an n-vertex graph $T = (V, E)$:

(A) T is connected and has no cycles, i.e. is a tree;

(B) T has $n - 1$ edges and is connected;

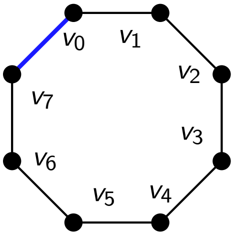
(C) T has $n - 1$ edges and has no cycles;

(D) T has a unique path between any pair of vertices.

(A) $\Rightarrow$ (B), (C) and (D): ✓

(D) $\Rightarrow$ (A): T has a path between any pair of vertices, so it's connected.

And on any cycle $v_0 \dots v_k$, there are two different paths from v_0 to v_k:



- the path $v_0 \dots v_k$; and
- the edge $v_0 v_k$.

Lemma: The following are equivalent for an n-vertex graph $T = (V, E)$:

(A) T is connected and has no cycles, i.e. is a tree;

(B) T has $n - 1$ edges and is connected;

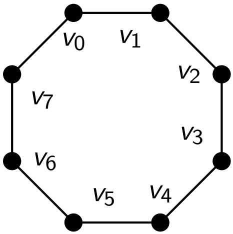
(C) T has $n - 1$ edges and has no cycles;

(D) T has a unique path between any pair of vertices.

(A) $\Rightarrow$ (B), (C) and (D): ✓

(D) $\Rightarrow$ (A): T has a path between any pair of vertices, so it's connected.

And on any cycle $v_0 \ldots v_k$, there are two different paths from v_0 to v_k:



- the path $v_0 \ldots v_k$; and
- the edge $v_0 v_k$.

So T has no cycles. ✓

Lemma: The following are equivalent for an n-vertex graph $T = (V, E)$:

(A) T is connected and has no cycles, i.e. is a tree;

(B) T has $n - 1$ edges and is connected;

(C) T has $n - 1$ edges and has no cycles;

(D) T has a unique path between any pair of vertices.

(A) $\Rightarrow$ (B), (C) and (D): ✓ **(D) $\Rightarrow$ (A):** ✓

(C) $\Rightarrow$ (A): Suppose T has no cycles and components $C_1, \dots, C_r$.

Each of these components has no cycles, and is connected, so it's a tree. So by (A) $\Rightarrow$ (B) (or Lemma 2), each C_i has $|V(C_i)| - 1$ edges.

Every edge of T is in some C_i, so $|E| = \sum_i (|V(C_i)| - 1) = n - r$. But we know $|E| = n - 1$, so we must have $r = 1$. ✓

Lemma: The following are equivalent for an n-vertex graph $T = (V, E)$:

(A) T is connected and has no cycles, i.e. is a tree;

(B) T has $n - 1$ edges and is connected;

(C) T has $n - 1$ edges and has no cycles;

(D) T has a unique path between any pair of vertices.

(A) $\Rightarrow$ (B), (C) and (D): ✓ **(C) and (D) $\Rightarrow$ (A):** ✓

(B) $\Rightarrow$ (A): We will need to use:

Claim: If $T = (V, E)$ is connected, and $e \in E$ is on a cycle, then $T - e$ is connected.

Proof from Claim: Suppose T is not a tree, so it has a cycle.

We form a new graph T' by repeatedly removing edges from cycles in T (in arbitrary order) until no more cycles remain.

Then T' has no cycles, and the Claim implies it's connected, so it's a tree. So by (A) $\Rightarrow$ (B) (or Lemma 2), T' has $n - 1$ edges.

So T must have had **more than** $n - 1$ edges — a contradiction. □

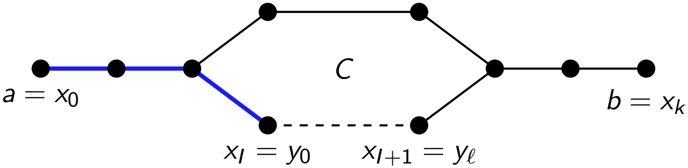
Claim: If $T = (V, E)$ is connected, and $e \in E$ is on a cycle, then $T - e$ is connected.

For all $a, b \in V$, we must find a path from a to b in $T - e$.

Let $P = x_0 \dots x_k$ be a path from a to b in T.

If e is not in P: Then P is the path we want. ✓

If e is in P: Write $e = \{x_I, x_{I+1}\}$. Let $C = y_0 \dots y_\ell$ be a cycle in T containing e — without loss of generality we can take $y_0 = x_I$ and $y_\ell = x_{I+1}$.



Then $\boldsymbol{x_0 \dots x_I} y_1 \dots y_\ell x_{I+2} \dots x_k$ is a walk from a to b in $T - e$. Any walk from a to b contains a path from a to b (see quiz 2), so we're done. ✓

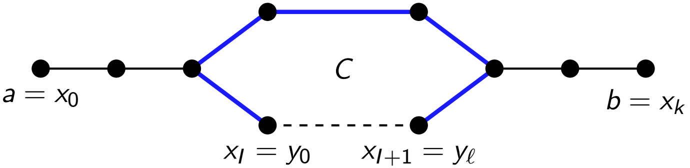
Claim: If $T = (V, E)$ is connected, and $e \in E$ is on a cycle, then $T - e$ is connected.

For all $a, b \in V$, we must find a path from a to b in $T - e$.

Let $P = x_0 \dots x_k$ be a path from a to b in T.

If e is not in P: Then P is the path we want. ✓

If e is in P: Write $e = \{x_I, x_{I+1}\}$. Let $C = y_0 \dots y_\ell$ be a cycle in T containing e — without loss of generality we can take $y_0 = x_I$ and $y_\ell = x_{I+1}$.



Then $x_0 \dots \boldsymbol{x_I y_1 \dots y_\ell} x_{I+2} \dots x_k$ is a walk from a to b in $T - e$. Any walk from a to b contains a path from a to b (see quiz 2), so we're done. ✓

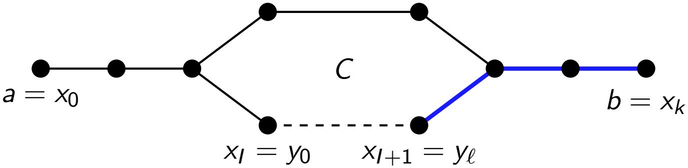
Claim: If $T = (V, E)$ is connected, and $e \in E$ is on a cycle, then $T - e$ is connected.

For all $a, b \in V$, we must find a path from a to b in $T - e$.

Let $P = x_0 \dots x_k$ be a path from a to b in T.

If e is not in P: Then P is the path we want. ✓

If e is in P: Write $e = \{x_l, x_{l+1}\}$. Let $C = y_0 \dots y_\ell$ be a cycle in T containing e — without loss of generality we can take $y_0 = x_l$ and $y_\ell = x_{l+1}$.



Then $x_0 \dots x_l y_1 \dots \boldsymbol{y_\ell x_{l+2} \dots x_k}$ is a walk from a to b in $T - e$. Any walk from a to b contains a path from a to b (see quiz 2), so we're done. ✓

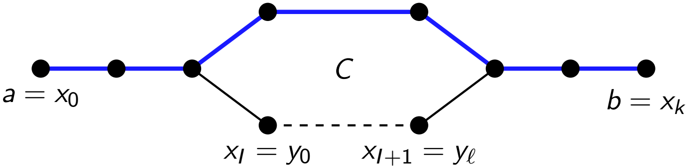
Claim: If $T = (V, E)$ is connected, and $e \in E$ is on a cycle, then $T - e$ is connected.

For all $a, b \in V$, we must find a path from a to b in $T - e$.

Let $P = x_0 \ldots x_k$ be a path from a to b in T.

If e is not in P: Then P is the path we want. ✓

If e is in P: Write $e = \{x_I, x_{I+1}\}$. Let $C = y_0 \ldots y_\ell$ be a cycle in T containing e — without loss of generality we can take $y_0 = x_I$ and $y_\ell = x_{I+1}$.



Then $x_0 \ldots x_I y_1 \ldots y_\ell x_{I+2} \ldots x_k$ is a walk from a to b in $T - e$. Any walk from a to b contains a **path** from a to b (see quiz 2), so we're done. ✓

Lemma: The following are equivalent for an n-vertex graph $T = (V, E)$:

(A) T is connected and has no cycles, i.e. is a tree;

(B) T has $n-1$ edges and is connected;

(C) T has $n-1$ edges and has no cycles;

(D) T has a unique path between any pair of vertices. $\square$

Our reward for proving this lemma is: we never have to think about basic tree properties in this level of detail ever again. (Except on the exam!)

And there was much rejoicing.