Math 530 Spring 2022, Lecture 13: Trees

website: `https://www.cip.ifi.lmu.de/~grinberg/t/22s`

1. Trees and arborescences

Trees are particularly nice graphs. Among other things, they can be characterized as

- the minimal connected graphs on a given set of vertices, or
- the maximal acyclic (= having no cycles) graphs on a given set of vertices, or
- in many other ways.

Arborescences are their closest analogue for digraphs.

In this chapter, we will discuss the theory of trees and some of their applications. Further applications are usually covered in courses in theoretical computer science, but their notion of a tree is somewhat different from ours.

1.1. Some general properties of components and cycles

1.1.1. Backtrack-free walks revisited

Before we start with trees, let us recall and prove some more facts about general multigraphs. Recall the notion of a "backtrack-free walk" that already had a brief appearance in Lecture 4:

Definition 1.1.1. Let G be a multigraph. A **backtrack-free walk** of G means a walk $\mathbf{w}$ such that no two adjacent edges of $\mathbf{w}$ are identical.

Here are a few properties of this notion:

Proposition 1.1.2. Let G be a multigraph. Let $\mathbf{w}$ be a backtrack-free walk of G. Then, $\mathbf{w}$ either is a path or contains a cycle.

Proof. We did this for simple graphs in Lecture 4 (Proposition 1.2.4). More or less the same argument works for multigraphs. ("More or less" because the definition of a cycle in a multigraph is slightly different from that in a simple graph; but the proof is easy to adapt.) □

Theorem 1.1.3. Let G be a multigraph. Let u and v be two vertices of G. Assume that there are two distinct backtrack-free walks from u to v in G. Then, G has a cycle.

Proof. We did this for simple graphs in Lecture 4 (Claim 1 in the proof of Theorem 1.2.7). More or less the same argument works for multigraphs. □

1.1.2. Counting components

Next, we shall derive a few properties of the number of components of a graph. Again, we have already done most of the hard work, and we can now derive corollaries. First, we give this number a name:

Definition 1.1.4. Let G be a multigraph. Then, $\operatorname{conn} G$ means the number of components of G. (Some authors also call this number $b_0(G)$. This notation comes from algebraic topology, where it stands for the 0-th Betti number. This makes sense, because we can regard a multigraph G as a topological space. But we won't need this.)

So a multigraph G satisfies $\operatorname{conn} G = 1$ if and only if G is connected. Moreover, $\operatorname{conn} G = 0$ if and only if G has no vertices.

Recall the following:

Theorem 1.1.5. Let G be a multigraph. Let e be an edge of G. Then:

- **(a)** If e is an edge of some cycle of G, then the components of $G \setminus e$ are precisely the components of G. (Keep in mind that the components are sets of vertices. It is these sets that we are talking about here, not the induced subgraphs on these sets.)
- **(b)** If e appears in no cycle of G (in other words, there exists no cycle of G such that e is an edge of this cycle), then the graph $G \setminus e$ has one more component than G.

Proof. For simple graphs, we proved this in Lecture 5 (Theorem 1.2.2). More or less the same proof works for multigraphs. See Lecture 8 for details. □

Corollary 1.1.6. Let G be a multigraph. Let e be an edge of G. Then:

- **(a)** If e is an edge of some cycle of G, then $\operatorname{conn}(G \setminus e) = \operatorname{conn} G$.
- **(b)** If e appears in no cycle of G, then $\operatorname{conn}(G \setminus e) = \operatorname{conn} G + 1$.
- **(c)** In either case, we have $\operatorname{conn}(G \setminus e) \leq \operatorname{conn} G + 1$.

Proof. Part **(a)** follows from Theorem 1.1.5 **(a)**. Part **(b)** follows from Theorem 1.1.5 **(b)**. Part **(c)** follows by combining parts **(a)** and **(b)**. □

Corollary 1.1.7. Let $G = (V, E, \varphi)$ be a multigraph. Then, $\operatorname{conn} G \geq |V| - |E|$.

Proof. We induct on $|E|$:

Base case: If $|E| = 0$, then $\operatorname{conn} G = |V|$ (since $|E| = 0$ means that the graph G has no edges, and thus no two distinct vertices are path-connected); but this rewrites as $\operatorname{conn} G = |V| - |E|$ (since $|E| = 0$). Thus, Corollary 1.1.7 is proved for $|E| = 0$.

Induction step: Let $k \in \mathbb{N}$. Assume (as the induction hypothesis) that Corollary 1.1.7 holds for $|E| = k$. We must now show that it also holds for $|E| = k + 1$.

So let us consider a multigraph $G = (V, E, \varphi)$ with $|E| = k + 1$. Thus, $|E| - 1 = k$. Pick any edge $e \in E$ (such an edge exists, since $|E| = k + 1 \geq 1 > 0$). Then, the multigraph $G \setminus e$ has edge set $E \setminus \{e\}$ and therefore has $|E \setminus \{e\}| = |E| - 1 = k$ many edges. Hence, by the induction hypothesis, we have

$$\operatorname{conn}(G \setminus e) \geq |V| - |E \setminus \{e\}|$$

(since $G \setminus e$ is a multigraph with vertex set V and edge set $E \setminus \{e\}$). However, Corollary 1.1.6 **(c)** yields $\operatorname{conn}(G \setminus e) \leq \operatorname{conn} G + 1$. Thus,

$$\operatorname{conn} G \geq \underbrace{\operatorname{conn}(G \setminus e)}_{\geq |V| - |E \setminus \{e\}|} - 1 \geq |V| - \underbrace{|E \setminus \{e\}|}_{= |E| - 1} - 1 = |V| - (|E| - 1) - 1 = |V| - |E|.$$

This completes the induction step. Thus, Corollary 1.1.7 is proven. $\square$

Corollary 1.1.8. Let $G = (V, E, \varphi)$ be a multigraph that has no cycles. Then, $\operatorname{conn} G = |V| - |E|$.

Proof. Replay the proof of Corollary 1.1.7, with just a few changes: Instead of applying Corollary 1.1.6 **(c)**, apply Corollary 1.1.6 **(b)** (this is allowed because G has no cycles and thus e appears in no cycle of G). The induction hypothesis can be used because when G has no cycles, $G \setminus e$ has no cycles either. All $\leq$ and $\geq$ signs in the above proof now can be replaced by $=$ signs (since Corollary 1.1.6 **(b)** claims an equality, not an inequality). The result is therefore $\operatorname{conn} G = |V| - |E|$. $\square$

Corollary 1.1.9. Let $G = (V, E, \varphi)$ be a multigraph that has at least one cycle. Then, $\operatorname{conn} G \geq |V| - |E| + 1$.

Proof. Pick an edge $e \in E$ that belongs to some cycle (such an edge exists, since G has at least one cycle). Then, Corollary 1.1.6 **(a)** yields $\operatorname{conn}(G \setminus e) = \operatorname{conn} G$. However, Corollary 1.1.7 (applied to $G \setminus e$ and $E \setminus \{e\}$ instead of G and E) yields

$$\operatorname{conn}(G \setminus e) \geq |V| - \underbrace{|E \setminus \{e\}|}_{= |E| - 1} = |V| - (|E| - 1) = |V| - |E| + 1.$$

Since $\operatorname{conn}(G \setminus e) = \operatorname{conn} G$, this rewrites as $\operatorname{conn} G \geq |V| - |E| + 1$. $\square$

We summarize what we have proved into one convenient theorem:

Theorem 1.1.10. Let $G = (V, E, \varphi)$ be a multigraph. Then:

(a) We always have $\operatorname{conn} G \geq |V| - |E|$.

(b) We have $\operatorname{conn} G = |V| - |E|$ if and only if G has no cycles.

Proof. **(a)** This is Corollary 1.1.7.

(b) $\Longleftarrow$: This is Corollary 1.1.8.

$\Longrightarrow$: Assume that $\operatorname{conn} G = |V| - |E|$. If G had any cycles, then Corollary 1.1.9 would yield $\operatorname{conn} G \geq |V| - |E| + 1 > |V| - |E|$, which would contradict $\operatorname{conn} G = |V| - |E|$. So G has no cycles. This proves the "$\Longrightarrow$" direction of Theorem 1.1.10. □

Remark 1.1.11. Let $G = (V, E, \varphi)$ be a multigraph. Does the number

$$\operatorname{conn} G - (|V| - |E|)$$

have anything to do with how many cycles G has? We know that it is 0 if G has no cycles. More generally, could it just be the number of cycles of G ? (Let's say we count reversals and cyclic rotations of a cycle as being the same cycle.)

Unfortunately, the answer is still no. For example, a complete graph K_n has many more than $1 - \left(n - \dbinom{n}{2}\right)$ many cycles. However, there is still some subtler connection. The number $\operatorname{conn} G - (|V| - |E|)$ is known as the **circuit rank** or the **cyclomatic number** of G, and is the dimension of a certain vector space that, in some way, consists of cycles.

1.2. Forests and trees

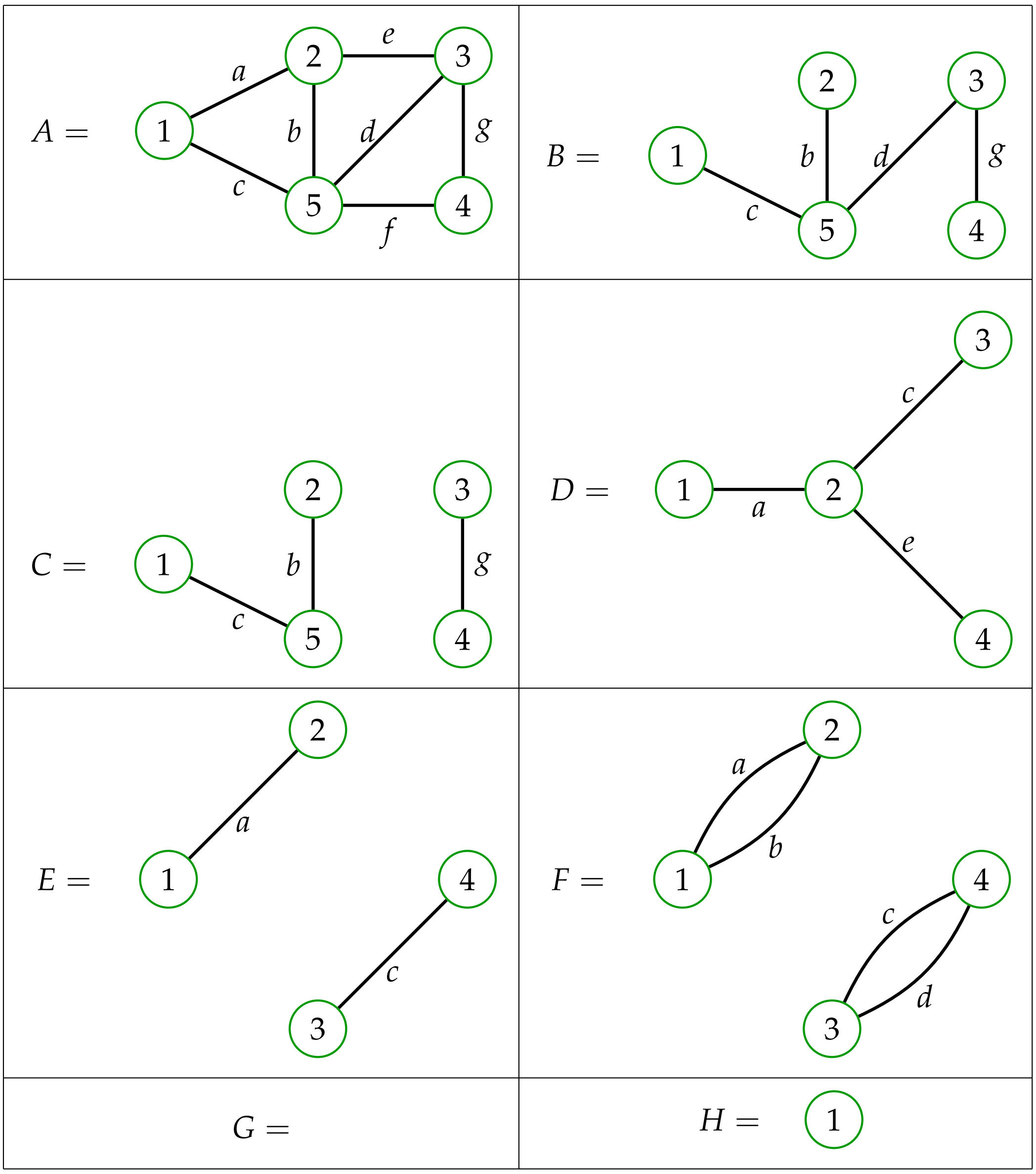
We now come to two of the heroes of this chapter:

Definition 1.2.1. A **forest** is a multigraph with no cycles.

(In particular, a forest therefore cannot contain two distinct parallel edges. It also cannot contain loops.)

Definition 1.2.2. A **tree** is a connected forest.

Example 1.2.3. Consider the following multigraphs:



.

(Yes, G is an empty graph with no vertices.) Which of them are forests, and which are trees?

- The graph A is not a forest, since it has a cycle (actually, several cycles). Thus, A is not a tree either.
- The graph B is a tree.
- The graph C is a forest, but not a tree, since it is not connected.
- The graph D is a tree.
- The graph E is a forest, but not a tree.

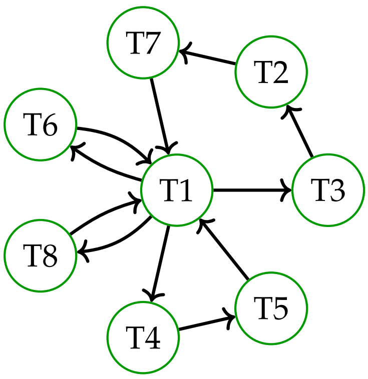
- The graph F is not a forest, since it has cycles.
- The graph G (which has no vertices and no edges) is a forest, but not a tree, since it is not connected (recall: a graph is connected if it has 1 component; but G has 0 components).
- The graph H is a tree.

Trees can be described in many ways:

Theorem 1.2.4 (The tree equivalence theorem)**.** Let $G = (V, E, \varphi)$ be a multigraph. Then, the following eight statements are equivalent:

- **Statement T1:** The multigraph G is a tree.
- **Statement T2:** The multigraph G has no loops, and we have $V \neq \varnothing$, and for each $u \in V$ and $v \in V$, there is a **unique** path from u to v.
- **Statement T3:** We have $V \neq \varnothing$, and for each $u \in V$ and $v \in V$, there is a **unique** backtrack-free walk from u to v.
- **Statement T4:** The multigraph G is connected, and we have $|E| = |V| - 1$.
- **Statement T5:** The multigraph G is connected, and we have $|E| < |V|$.
- **Statement T6:** We have $V \neq \varnothing$, and the graph G is a forest, but adding any new edge to G creates a cycle.
- **Statement T7:** The multigraph G is connected, but removing any edge from G yields a disconnected (i.e., non-connected) graph.
- **Statement T8:** The multigraph G is a forest, and we have $|E| \geq |V| - 1$ and $V \neq \varnothing$.

Proof. We shall prove the following implications:



.

In this digraph, an arc from Ti to Tj stands for the implication T$i \Longrightarrow$Tj. Since this digraph is strongly connected (i.e., you can travel from Statement Ti to Statement Tj along its arcs for any i, j), this will prove the theorem. So let us prove the implications.

Proof of T1$\Longrightarrow$T3: Assume that Statement T1 holds. Thus, G is a tree. Therefore, G is connected, so that $V \neq \varnothing$. We must prove that for each $u \in V$ and $v \in V$, there is a **unique** backtrack-free walk from u to v. The existence of such a walk is clear (since G is connected, so there is a path from u to v). Thus, we only need to show that it is unique. But this is easy: If there were two distinct backtrack-free walks from u to v (for some $u \in V$ and $v \in V$), then Theorem 1.1.3 would show that G has a cycle, and thus G could not be a forest, let alone a tree. Thus, the backtrack-free walk from u to v is unique. So we have proved Statement T3. The implication T1$\Longrightarrow$T3 is thus proved.

Proof of T3$\Longrightarrow$T2: Assume that Statement T3 holds. We must prove that Statement T2 holds. First, G has no loops, because if there was a loop e with endpoint u, then the two walks (u) and (u, e, u) would be two distinct backtrack-free walks from u to u. It remains to prove that for each each $u \in V$ and $v \in V$, there is a **unique** path from u to v. However, the existence of a walk from u to v always implies the existence of a path from u to v (by Corollary 1.1.5 from Lecture 8). Moreover, the uniqueness of a backtrack-free walk from u to v implies the uniqueness of a path from u to v (since any path is a backtrack-free walk). Thus, Statement T2 follows from Statement T3.

Proof of T2$\Longrightarrow$T7: Assume that Statement T2 holds. Then, G is connected. Now, let us remove any edge e from G. Let u and v be the endpoints of e. Then, $u \neq v$ (since G has no loops). There cannot be a path from u to v in the graph $G \setminus e$ (because if there was such a path, then it would also be a path from u to v in the graph G, and this path would be distinct from the path (u, e, v); thus, the graph G would have at least two paths from u to v; but this would contradict the uniqueness part of Statement T2). Hence, the graph $G \setminus e$ is disconnected. So we have shown that G is connected, but removing any edge from G yields a disconnected graph. In other words, Statement T7 holds.

Proof of T7$\Longrightarrow$T1: Assume that Statement T7 holds. We must show that G is a tree. Since G is connected (by Statement T7), it suffices to show that G is a forest, i.e., that G has no cycles. However, if G had any cycle, then we could pick any edge e of this cycle, and then we would know that $G \setminus e$ is still connected (since Corollary 1.1.6 **(a)** would yield $\operatorname{conn}(G \setminus e) = \operatorname{conn} G = 1$), and this would contradict Statement T7. Thus, G has no cycles, hence is a forest. This proves Statement T1.

Proof of T1$\Longrightarrow$T6: Assume that Statement T1 holds. Thus, G is a tree. We must show that adding any new edge to G creates a cycle (since all other parts of Statement T6 are clear).

Indeed, let us add a new edge f to G. Let u and v be the endpoints of f. The graph G is connected, so there is already a path from u to v in G. Combining this path with the edge f, we obtain a cycle. Thus, the graph obtained from G by adding the new edge f has a cycle. This completes our proof that Statement T6 holds.

Proof of T6$\Longrightarrow$T1: Assume that Statement T6 holds. Thus, G is a forest. We must only show that G is connected.

Assume the contrary. Thus, there exist two vertices u and v of G that are not path-connected in G. Hence, adding a new edge f with endpoints u and v to the graph G cannot create a new cycle (because any such cycle would have to contain f (otherwise, it would already be a cycle of G, but G has no cycles), and then we could remove f from it to obtain a path from u to v in G; but such a path cannot exist, since u and v are not path-connected in G). This contradicts Statement T6.

So we have shown that G is connected, and thus G is a tree. This proves Statement T1.

Proof of T1$\Longrightarrow$T8: Assume that Statement T1 holds. So G is a tree. Clearly, G is then a forest. We must show that $|E| \geq |V| - 1$.

Theorem 1.1.10 **(a)** yields $\operatorname{conn} G \geq |V| - |E|$. But we have $\operatorname{conn} G = 1$ because G is connected. Thus, $1 = \operatorname{conn} G \geq |V| - |E|$. In other words, $|E| \geq |V| - 1$. This proves Statement T8.

Proof of T8$\Longrightarrow$T1: Assume that Statement T8 holds. Thus, G is a forest. We must only show that G is connected. However, G is a forest, and thus has no cycles. Hence, Theorem 1.1.10 **(b)** yields $\operatorname{conn} G = |V| - |E| \leq 1$ (since Statement 8 yields $|E| \geq |V| - 1$). On the other hand, $\operatorname{conn} G \geq 1$ (since $V \neq \varnothing$). Combining these two inequalities, we obtain $\operatorname{conn} G = 1$. In other words, G is connected. This yields Statement T1 (since G is a forest).

Proof of T1$\Longrightarrow$T4: Assume that Statement T1 holds. Then, G is a tree, hence a connected forest. Therefore, G has no cycles (by the definition of a forest). Theorem 1.1.10 **(b)** therefore yields $\operatorname{conn} G = |V| - |E|$. Thus, $|V| - |E| = \operatorname{conn} G = 1$ (since G is connected), so that $|E| = |V| - 1$. Thus, Statement T4 is proved.

Proof of T4$\Longrightarrow$T5: The implication T4$\Longrightarrow$T5 is obvious.

Proof of T5$\Longrightarrow$T1: Assume that Statement T5 holds. Thus, the multigraph G is connected, and we have $|E| < |V|$. Thus, $|E| \leq |V| - 1$. In other words, $1 \leq |V| - |E|$. Since G is connected, we have $\operatorname{conn} G = 1 \leq |V| - |E|$. However, Theorem 1.1.10 **(a)** yields $\operatorname{conn} G \geq |V| - |E|$. Combining these two inequalities, we obtain $\operatorname{conn} G = |V| - |E|$. Thus, Theorem 1.1.10 **(b)** shows that G has no cycles. In other words, G is a forest. Hence, G is a tree (since G is connected). This proves Statement T1.

We have now proved all necessary implications to conclude that all eight statements T1, T2, ..., T8 are equivalent. Theorem 1.2.4 is thus proved. $\square$

We also observe the following connection between trees and forests:

Proposition 1.2.5. Let G be a multigraph, and let $C_1, C_2, \ldots, C_k$ be its components. Then, G is a forest if and only if all the induced subgraphs $G[C_1], G[C_2], \ldots, G[C_k]$ are trees.

Proof. $\Longrightarrow$: Assume that G is a forest. Thus, G has no cycles. Hence, the induced subgraphs $G[C_1], G[C_2], \ldots, G[C_k]$ have no cycles either (since a cycle in any of them would be a cycle of G); in other words, they are forests. But they are furthermore connected (since the induced subgraph on a component is always connected). Hence, they are connected forests, i.e., trees.

$\Longleftarrow$: Assume that the induced subgraphs $G[C_1], G[C_2], \ldots, G[C_k]$ are trees. Hence, none of them has a cycle. Thus, G has no cycles either (since a cycle of G would have to be fully contained in one of these induced subgraphs[1]). In other words, G is a forest. $\square$

[1] Indeed, if it wasn't, then it would contain vertices from different components. But this is impossible, since there are no walks between vertices in different components.