Lecture 3: Trees, part I

September 8, 2020

Equivalent characterizations of trees

Recall that a tree is a connected graph that contains no subgraph isomorphic to a polygon P_n (in particular, it is a simple graph).

Proposition 1. For a simple graph G, the following are equivalent:

i) G is a tree.

ii) For every two distinct vertices x, y of G, there is a unique path from x to y (and this is simple).

iii) G is connected and it contains no closed paths.

We begin with the following easy

Remark. Given a closed walk $x_0, e_1, x_1, e_2 \ldots, e_n, x_n = x_0$ in a graph G, with $n > 0$, and such that $e_i \neq e_{i+1}$ for $1 \leq i \leq n$ (with the convention that $e_{n+1} = e_1$), we can find $0 \leq i < j \leq n$ such that

$$x_i, e_{i+1}, x_{i+1}, \ldots, e_j, x_j = x_i$$

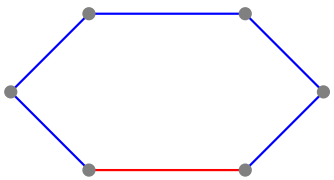
is a simple closed path.

Characterization of trees, cont'd

Remark, cont'd. Indeed, if we choose $0 \leq i < j \leq n$ such that $x_i = x_j$ and $j - i$ is minimal, then this satisfies the required property.

Proof of Proposition 1. ii)⇒i): existence of a path between any 2 distinct point implies that the graph is connected. Then, if we have a subgraph H of G isomorphic to some P_n, with $n \geq 3$, then we have vertices $x_1, \ldots, x_n$ of G such that

$$\{x_1, x_2\} \in E(G), \ldots, \{x_{n-1}, x_n\} \in E(G), \{x_n, x_1\} \in E(G).$$



$$n = 6.$$

We have two distinct simple paths between x_1 and x_n: $x_1, x_2, \ldots, x_n$ and x_1, x_n; this contradicts ii).

Characterization of trees, cont'd

i)⇒iii). The fact that a tree is connected is part of the definition. If G contains a closed path $x_0, x_1, \ldots, x_n = x_0$, then it follows from the remark before the proof that there are $0 \leq i < j \leq n$ such that $x_i, x_{i+1}, \ldots, x_j = x_i$ is a simple closed path. Therefore $x_i, x_{i+1}, \ldots, x_{j-1}$ are the vertices of a subgraph of G isomorphic to P_{j-i}, a contradiction.

iii)⇒ii): Let x and y be distinct vertices of G. Since G is connected, there is a walk $x = x_0, x_1, \ldots, x_{n-1}, x_n = y$ in G. If n is minimal, then this is a simple path: if we have $x_i = x_j$ for some $0 \leq i < j \leq n$, then $x_0, \ldots, x_i, x_{j+1}, \ldots, x_n$ is a shorter walk between x and y.

For uniqueness, we argue by induction on the shortest path between x and y. Suppose that we have two distinct paths $x = x_0, x_1, \ldots, x_n = y$ and $x = y_0, y_1, \ldots, y_m = y$. If $x_1 = y_1$, then $m > 1$ or $n > 1$ and there are two possibilities: either $x_1 \neq y$, in which case we are done by induction; or $x_1 = y$, in which case, if say $n > 1$, then $x_1, \ldots, x_n$ gives a closed path, a contradiction

Characterization of trees, cont'd

Suppose now that $x_1 \neq y_1$. Arguing similarly, we may assume that $x_{n-1} \neq y_{m-1}$. Consider the closed walk

$$x = w_0, w_1, \ldots, w_{n+m} = x,$$

where $w_i = x_i$ for $i \leq n$ and $w_i = y_{m+n-i}$ for $i \geq n$. Our assumptions imply that $\{w_i, w_{i+1}\} \neq \{w_{i+1}, w_{i+2}\}$ for $0 \leq i \leq m+n-1$ (with the convention that $w_{m+n+1} = w_1$). In this case, the remark before the proof of the proposition implies that we can find $0 \leq i < j \leq m+n$ such that $w_i, w_{i+1}, \ldots, w_j$ is a simple closed path, contradicting iii).

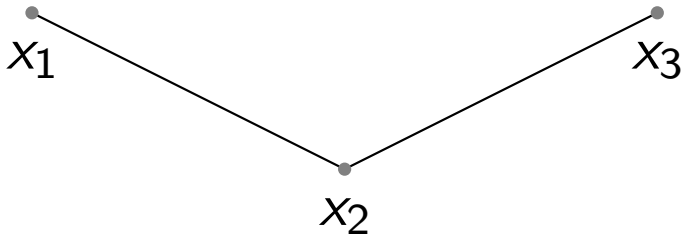
This completes the proof of Proposition 1.

Vertices of degree 1

We next discuss some general properties of trees. From now on, we assume that all graphs we encounter are finite.

Lemma 1. If G is a tree with $\#V(G) > 1$, then there is $x \in V(G)$ with $\deg(x) = 1$.

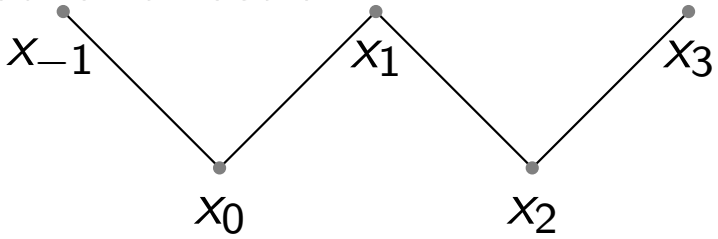
Proof. Let $x_1 \in V(G)$ be arbitrary. Since G is connected and $\#V(G) > 1$, we have $\deg(x_1) \geq 1$. Pick any edge incident to x_1 and x_2 be the other end of this edge. If $\deg(x_2) = 1$, then we are done. Otherwise, pick another edge incident to x_2 and repeat this construction.



Since G is simple and contains no closed paths, all vertices we encounter are distinct. Therefore the process has to stop at a vertex of degree 1. This completes the proof.

Vertices of degree 1, cont'd

Remark. The argument in the previous proof can be refined to show that if G is a tree and $\#V(G) > 1$, then G has at least 2 vertices of degree 1. Indeed, with the previous notation, this is clear if $\deg(x_1) = 1$. If this is not the case, there is a second edge incident to x_1 and we take x_0 to be the other end of this edge. We can then continue the argument in the other direction.



The process then ends at a second vertex of degree 1.

Removing a degree 1 vertex

Let G be a connected graph and $x \in V(G)$ with $\deg(x) = 1$. We consider the subgraph G' of G spanned by $V(G) \setminus \{x\}$. Therefore we have

$$V(G') = V(G) \setminus \{x\}, \quad E(G') = E(G) \setminus \{\text{edges incident to x}\}.$$

Lemma 2. With the above notation, the graph G' is connected.

Proof. Indeed, let $y_1, y_2 \in V(G) \setminus \{x\}$, with $y_1 \neq y_2$. Since G is connected, we have a walk $y_1 - y_2$ in G. We may assume that in this walk no two consecutive edges are equal (by simply removing such pairs). In this case, all vertices on this walk between y_1 and y_2 have degree ≥ 2. Therefore this is a walk in G'.

Removing a degree 1 vertex, cont'd

Corollary. If G is a tree, $x \in V(G)$ is a vertex of degree 1, and G' is defined as on the previous slide, then G' is a tree.

Proof. By Lemma 2, we know that G' is connected. On the other hand, since G is a simple graph that contains no polygons as subgraphs, it is clear that any subgraph has the same property. Therefore G' is a tree.

A characterization of trees

We next characterize trees via the numbers of vertices and edges.

Proposition 2. If G is a connected graph on n vertices, with d edges, then

$$d \geq n - 1.$$

Moreover, equality holds if and only if G is a tree.

Proof. We argue by induction on $n \geq 1$. If $n = 1$, then everything is clear, hence from now on we assume that $n \geq 2$.

Claim. We may assume that there is $x \in V(G)$ with $\deg(x) = 1$.

If G is a tree, we know this by Lemma 1. In general, recall that we proved in the first lecture that

$$\sum_{x \in V(G)} \deg(x) = 2d. \tag{1}$$

Hence if $\deg(x) \geq 2$ for every $x \in V(G)$, then $d \geq n$. It follows that for all the assertions in the theorem, we may assume that we have $x \in V(G)$ with $\deg(x) = 1$.

A characterization of trees, cont'd

Consider the subgraph G' of G spanned by $V(G) \setminus \{x\}$. Note that G' is a graph on $n-1$ vertices and with $d-1$ edges.
Since $n \geq 2$, it follows from Lemma 2 that G' is connected. We can thus apply the induction hypothesis for G' to conclude

$$d - 1 \geq n - 2, \tag{2}$$

which implies the first assertion in the theorem.

Moreover, we see that we have equality for G if and only if we have equality for G'. Hence the proposition follows if we show that G is a tree if and only if G' is the tree. The "only if" part we have already seen (the corollary to Lemma 2). The "if" part is the content of the next lemma.

A characterization of trees, cont'd

Lemma 3. Let G' be a connected graph and let $y \in V(G')$. Suppose that G is a graph such that $V(G) = V(G') \sqcup \{x\}$ and $E(G) = E(G') \sqcup \{e\}$, where e is an edge that joins x and y. If G' is a tree, then G is a tree.

Proof. Note first that since G' is connected, G is connected as well. Indeed, y can be connected by a walk in G' (hence in G) with any vertex in G'. Since x is joined with y by e, it follows that G is connected. If P is a subgraph of G that is a polygon, then in fact it must be a subgraph of G', contradicting the fact that G' is a tree. This shows that G is a tree.

Existence of spanning trees

Let's work in groups on the following problem. Recall that for a graph G, a *spanning subgraph* is a subgraph G' of G with $V(G') = V(G)$.

Problem. Show that if G is a connected graph, then a subgraph T of G is a minimal element in

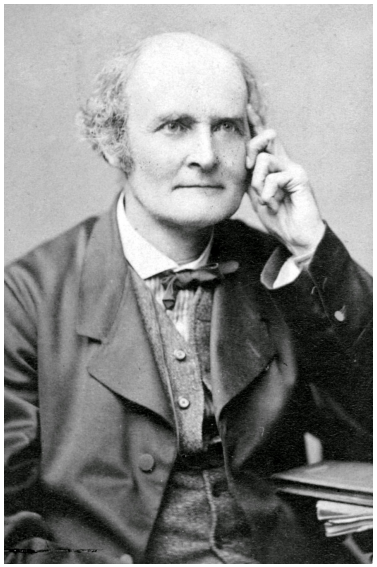
$$\{G' \text{ spanning subgraph of } G \mid G' \text{ is connected}\}$$

if and only if T is a spanning tree of G (that is, a spanning subgraph of G which is a tree). Deduce that every connected graph has a spanning tree.

The number of spanning trees of K_n

Goal: give a formula for the number of trees on a set of n vertices. Note: we want to count all trees, not up to graph isomorphism.
Equivalently: we want to count the number of spanning trees on the complete graph K_n.

Theorem (Cayley's formula). The number of trees on a set of $n \geq 2$ vertices is n^{n-2}.



Cayley: British mathematician, 1821-1895

Worked in algebra, writing almost 1,000 papers.

Worked for 14 years as a lawyer, before becoming a professor in Cambridge.

The above formula was actually first proved by Borchardt in 1860 and then extended by Cayley in 1889.

First proof of Cayley's formula

There are many known proofs of this theorem. We will discuss two proofs.

The first proof is based on a counting argument. In this proof we will make use of the multinomial coefficients $\binom{n}{m_1,\ldots,m_k}$. These are defined by the multinomial formula:

$$(x_1 + \ldots + x_k)^n = \sum_{\substack{m_1,\ldots,m_k \geq 0,\\ m_1+\ldots+m_k=n}} \binom{n}{m_1,\ldots,m_k} x_1^{m_1} \cdots x_k^{m_k}.$$

Exercise. Show that

$$\binom{n}{m_1,\ldots,m_k} = \frac{n!}{m_1! \cdots m_k!}$$

by induction on $k \geq 2$.

First proof of Cayley's formula, cont'd

We will need the following recursion formula for multinomial coefficients, extending the familiar one for binomial coefficients.

Lemma 4. For every $n \geq 2$, we have

$$\binom{n}{m_1, \ldots, m_k} = \sum_{i=1}^{k} \binom{n-1}{m_1, \ldots, m_i - 1, \ldots m_k},$$

with the convention that if $m_i = 0$, then the corresponding term on the right-hand side is 0.

Proof. The formula follows by writing

$$(x_1 + \ldots + x_k)^n = (x_1 + \ldots + x_k)^{n-1} \cdot (x_1 + \ldots + x_k)$$

and using the definition of the multinomial coefficients in terms of the multinomial formula.

First proof of Cayley's formula, cont'd

We now begin the first proof of Cayley's formula. In fact, we will prove more. Suppose $n \geq 2$ and let us label the vertices $1, \ldots, n$. We denote by $t(n; d_1, \ldots, d_n)$ the number of trees with vertex set $\{1, \ldots, n\}$ such that the vertex i has degree d_i. Note: since $n \geq 2$, this number is 0 unless $d_i \geq 1$ for all i. The key point will be to show that if $n \geq 3$ and $d_1, \ldots, d_n \geq 1$, then

$$t(n; d_1, \ldots, d_n) = \binom{n-2}{d_1 - 1, \ldots, d_n - 1}. \tag{3}$$

Note that

$$\sum_{i=1}^{n} (d_i - 1) = 2\#E(G) - n = 2\big(\#V(G) - 1\big) - n = n - 2,$$

hence the above multinomial coefficient is well-defined.

First proof of Cayley's formula, cont'd

Assuming this formula, it follows that the number of trees G with $V(G) = \{1, \ldots, n\}$ is

$$\sum_{d_1,\ldots,d_n} \binom{n-2}{d_1-1,\ldots,d_n-1},$$

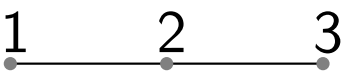
where the sum is over all $d_1, \ldots, d_n \geq 1$ such that $\sum_i (d_i - 1) = n - 2$. By the multinomial formula, this sum is thus equal to $(1 + \ldots + 1)^{n-2} = n^{n-2}$.

This proves Cayley's formula for $n \geq 3$ and for $n = 2$ the formula is trivially true. Hence in order to prove the theorem, it is enough to prove the formula for $t(n; d_1, \ldots, d_n)$.

We argue by induction on $n \geq 3$.

First proof of Cayley's formula, cont'd

If $n = 3$, then up to a permutation of the vertices, every tree on $\{1, 2, 3\}$ is of the form



hence $t(3; d_1, d_2, d_3) = 0$ unless 2 of the d_i are equal to 1 and the other one is equal to 2. It is clear that the same holds for the corresponding multinomial coefficient $\binom{1}{d_1-1, d_2-1, d_3-1}$.

Arguing by induction, we see that it is enough to show that $t(n; d_1, \ldots, d_n)$ satisfies the same recursion formula as given by Lemma 4 for the corresponding multinomial coefficients.

Since $t(n; d_1, \ldots, d_n)$ does not change if we permute the d_i, we may and will assume that $d_1 \geq \ldots \geq d_n = 1$ ($d_n = 1$ since $\sum_i d_i = 2n - 2$).

First proof of Cayley's formula, cont'd

For a tree on $1, \ldots, n$ with $\deg(n) = 1$, there is a unique i, with $1 \leq i \leq n-1$, that is joined by an edge with n. By deleting n and this edge, we have a tree (see the corollary to Lemma 2) in which $1, \ldots, n-1$ have degrees $d_1, \ldots, d_i - 1, \ldots, d_{n-1}$.

Moreover, Lemma 3 says that conversely, starting from such a tree, by adding the vertex n and an edge joining i and n, we get a tree with vertices $1, \ldots, n$, which have degrees $d_1, \ldots, d_n$. Hence

$$t(n; d_1, \ldots, d_{n-1}, 1) = \sum_{i=1}^{n-1} t(n-1; d_1, \ldots, d_i - 1, \ldots, d_{n-1}).$$

and Lemma 4 and the induction hypothesis give $t(n; d_1, \ldots, d_{n-1}, 1)$

$$= \sum_{i=1}^{n-1} \binom{n-3}{d_1 - 1, \ldots, d_i - 2, \ldots, d_{n-1} - 1} = \binom{n-2}{d_1 - 1, \ldots, d_{n-1} - 1, 0}.$$

This completes the proof of the theorem.

A second proof of Cayley's formula

We next give a second proof, due to Prüffer, for Cayley's formula. While a bit more involved, this has the advantage of being more constructive.

Key construction: given a tree T with vertex set $V = \{1, \ldots, n\}$, we associate to T an element $\mathcal{P}(T) \in V^{n-2}$, as follows. Let us consider the usual order on the above set V. We construct a sequence of trees $T_1, \ldots, T_{n-1}$ and two sequences of elements of V: $x_1, \ldots, x_{n-1}$ and $y_1, \ldots, y_{n-1}$ by recursion, as folllows.

Let $T_1 = T$, x_1 is the smallest element of T_1 of degree 1, and y_1 the unique element of T_1 joined to x_1 by an edge in T_1.

Let T_2 be the tree obtained from T_1 by removing the vertex x_1 and the edge joining x_1 and y_1 (we use here the corollary to Lemma 2). Let x_2 be the smallest element of T_2 of degree 1, and y_2 the unique element of T_2 joined to x_2 by an edge in T_2.

A second proof of Cayley's formula, cont'd

We repeat the construction, ending with T_{n-1}, x_{n-1}, and y_{n-1}. Note: $V = \{x_1, \ldots, x_{n-1}, y_{n-1}\}$. More generally: $V(T_i) = \{x_i, \ldots, x_{n-1}, y_{n-1}\}$.

We put $\mathcal{P}(T) = (y_1, \ldots, y_{n-2}) \in V^{n-2}$. Cayley's formula follows if we show that the map

$$\{\text{trees with vertex set } V\} \to V^{n-2}, \quad \text{mapping } T \text{ to } \mathcal{P}(T)$$

is bijective.
We begin with a few remarks regarding this construction:

- Each T_i has ≥ 2 vertices of degree 1, hence $x_i \leq n-1$ for $1 \leq i \leq n-1$. This implies that $y_{n-1} = n$.
- For $i \leq n-1$, the vertices of T_i are $x_i, x_{i+1}, \ldots, x_{n-1}, y_{n-1} = n$.
- The edges of T_i are those joining the following pairs: $(x_i, y_i), \ldots, (x_{n-1}, y_{n-1})$.

A second proof of Cayley's formula, cont'd

• For every $v \in V$, v appears among $y_1, \ldots, y_{n-2}$ exactly $\deg_T(v) - 1$ times: it appears $\deg_T(v)$ times among $x_1, y_1, \ldots, x_{n-1}, y_{n-1}$ (because of the previous remark) and once among $x_1, \ldots, x_{n-1}, y_{n-1}$.

• Applying the same observation with T replaced by T_i, we see that if v is a vertex of T_i, then v appears among $y_i, y_{i+1}, \ldots, y_{n-2}$ precisely $\deg_{T_i}(v) - 1$ times.

In particular, the vertices of T_i of degree 1 are precisely those elements of V that don't lie in $\{x_1, \ldots, x_{i-1}\} \cup \{y_i, \ldots, y_{n-2}\}$.

By construction, we thus see that x_i is the smallest element of V that does not lie in $\{x_1, \ldots, x_{i-1}\} \cup \{y_i, \ldots, y_{n-2}\}$. It follows that we can recursively recover $(x_1, \ldots, x_{n-1})$ (and thus the tree T) from $(y_1, \ldots, y_{n-2})$. This implies that the map we defined is injective.

A second proof of Cayley's formula, cont'd

For surjectivity, given $(y_1, \ldots, y_{n-2}) \in V^{n-2}$ we let x_1 be the smallest element of V that does not lie in $\{y_1, \ldots, y_{n-2}\}$. For $2 \leq i \leq n-1$, assuming that $x_1, \ldots, x_{i-1}$ have already been defined, we let x_i be the smallest element of V that is not in $\{x_1, \ldots, x_{i-1}\} \cup \{y_i, \ldots, y_{n-2}\}$. Once we have $x_1, \ldots, x_{n-1}$, we let G be the graph with edges joining the pairs of vertices $(x_1, y_1), \ldots, (x_{n-1}, y_{n-1})$.

I leave as an exercise checking that G is a tree that is mapped via our map to $(y_1, \ldots, y_{n-2})$. This completes the proof of the theorem.

Remark. Note that since we already know the number of trees with vertex set V (by the first proof we gave), the fact that the map we defined in the second proof is surjective also follows from the fact we have shown it is injective.

An example

Let's consider the trees on $\{1, \ldots, 7\}$ in which 2 and 3 have degree 3 and 5 has degree 2 (thus the other vertices must have degree 1). It follows from the first proof of Cayley's formula that the number of such trees is $\frac{5!}{2! \cdot 2!} = 30$.

In order to list all such trees, we can try Prüffer's approach: note that we clearly have $x_1 = 1$ and $y_1 \in \{2, 3, 5\}$. We next must have $x_2 = 4$ and $y_2 \in \{2, 3, 5\}$ (moreover, we can have $y_2 = 5$ if and only if $y_1 \neq 5$). However, it seems tricky to analyze in this way all possibilities.

An example, cont'd

It seems easier to analyze the situation directly. We know that each of the vertices 1, 4, 6, and 7 must be joined by one edge with one of 2, 3, and 5. This gives 4 edges. There will be 2 other edges joining vertices out of 2, 3, and 5. If these edges are joining 2 with 5 and 3 with 5, the other 4 vertices are only joined with 2 or 3. There are $\binom{4}{2} = 6$ such possibilities.

On the other hand, if those two edges are joining 2 with 3 and 2 with 5, then out of 1, 4, 6, and 7, we have to join two to 3, one with 2, and one with 5. We have $\binom{4}{2} = 12$ such possibilities.

The case when we have 3 joined with 2 and 3 joined with 5 is symmetric (switch 2 and 3). We thus get 12 more trees.

In total, we have $6 + 12 + 12 = 30$ trees.