Math 3322: Graph Theory[1] Mikhail Lavrov

Lecture 11: Trees and forests

September 17, 2024 *Kennesaw State University*

1 Counting edges in trees

A tree is an acyclic, connected graph. What can we say about the number of edges this requires?

The first thing we know is a fact we proved a few weeks ago, when we were discussing average and minimum degrees in graphs:

Theorem 1.1 (Corollary 3.2 from Lecture 6)**.** *If a graph has n vertices and at least n edges, then it contains a cycle.*

As a result, n-vertex trees can have at most $n-1$ edges, because we don't want then to have any cycles. Also, if a graph has no cycles and exactly $n-1$ edges, then it must be a tree: add any edge, and this theorem tells us that a cycle is created.

By itself, the argument above doesn't guarantee that all trees with n vertices do have exactly $n-1$ edges, but we're about to see that this is true. For lower bounds on the number of edges, we're going to have to look at connectedness.

Theorem 1.2. *A graph with n vertices and m edges has at least $n-m$ connected components.*

Proof. We'll prove this by induction on m.

When $m = 0$, if a graph has n vertices and 0 edges, then every vertex is an isolated vertex, so it is a connected component all by itself. There are always exactly $n = n - m$ connected components.

Assume that the theorem is true for graphs with $m-1$ edges, and let G be a graph with n vertices and m edges. We don't have to be super clever with the induction here; let uv be an arbitrary edge of G, and consider the $(m-1)$-edge graph $G - uv$.

By the induction hypothesis, $G - uv$ has at least $n - m + 1$ connected components. There are two cases for what edge uv can do to change this:

- If u and v are in the same connected component of $G - uv$, then adding edge uv does not do anything at all. There is already a $u - v$ path, so any walk that used edge uv could have used that $u - v$ path instead. Therefore if two vertices are in the same component of G, they're also in the same component of $G - uv$; G also has at least $n - m + 1$ connected components.
- If u and v are in different connected components of $G - uv$, then those two components become the same connected component of G. Any vertex in u's component can reach v (by going to u, then taking edge uv) and from there, it can reach any vertex in v's component.

[1] This document comes from the Math 3322 course webpage: http://facultyweb.kennesaw.edu/mlavrov/courses/3322-fall-2024.php

However, that's all that happens. If we look at a vertex w not in either of these connected components, then there's no walk starting at w that uses edge uv: it would first have to get to u or v without using that edge, which is impossible. So walks from w are the same in G and in $G - uv$. As a result, G has one connected component less than $G - uv$: at least $n - m$.

In both cases, G has at least $n - m$ connected components, so the induction is complete and the statement we want is true for all n and m. □

We can use these ideas to prove a few more characterizations of trees. Let's do that, and summarize all our results so far in a theorem:

Theorem 1.3. *The following conditions for a graph G with n vertices are all equivalent definitions of a tree:*

1. *G is minimally connected: connected, but deleting any edge will disconnect it.*
2. *G is maximally acyclic: acyclic, but adding any edge will create a cycle.*
3. *G is both acyclic and connected.*
4. *G is connected and has at most $n - 1$ edges.*
5. *G is acyclic and has at least $n - 1$ edges.*
6. *G is uniquely connected: there is exactly one path between any two vertices.*

Proof. We already know that conditions 1–3 are equivalent; we proved that last time. Assuming condition 3, we can:

- Use Theorem 1.1 to conclude that G can have at most $n - 1$ edges to be acyclic, proving condition 4;
- Use Theorem 1.2 to conclude that G must have at least $n - 1$ edges to be connected, proving condition 5.

So conditions 4–5 are true of all trees.

Suppose that condition 4 holds. Then G is not only connected but minimally connected: if we delete any edge, then the resulting $n - 2$ edges are not enough to connect G, by Theorem 1.2. Therefore condition 1 holds, and G is a tree.

Suppose that condition 5 holds. Then G is not only acyclic but maximally acyclic: if we add any edge, then the resulting n edges are guaranteed to contain a cycle, by Theorem 1.1. Therefore condition 2 holds, and G is a tree.

This shows that conditions 1–5 are equivalent. I've included condition 6 from practice problem 6 at the end of the previous lecture; it's also equivalent to the rest, but I won't prove that here. □

Conditions 4 and 5 of Theorem 1.3 are stated to be maximally easy to prove, but the words "at most" and "at least" are silly when using them. Together, those conditions tell us that every n-vertex tree has *exactly* $n - 1$ edges.

2 Forests

An acyclic graph, not necessarily connected, is called a **forest**. The reason for the name is that every connected component of a forest is connected (because it's a connected component) and acyclic (because the forest as a whole has no cycles), so it must be a tree. A forest is a graph whose components are all trees.

Suppose an n-vertex forest has k connected components: trees with $n_1, n_2, \dots, n_k$ vertices, where $n_1 + n_2 + \cdots + n_k = n$. The i^{th} tree has $n_i - 1$ edges, so the total number of edges in the forest is

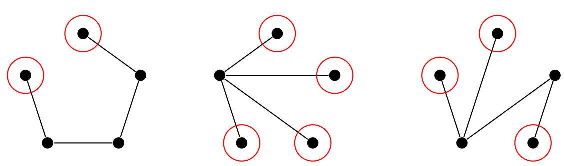
$$(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) = (n_1 + n_2 + \cdots + n_k) - k = n - k.$$

By Theorem 1.2, that's the minimum number of edges in a graph with n vertices and k connected components. This is not surprising, because each connected component in a forest has as few edges as it possibly can.

In fact, every graph G with n vertices, m edges, and exactly $n - m$ connected components must be a forest (that is, it cannot have any cycles). Otherwise, let uv be an edge of G that lies on one of it's cycles, so it's not a bridge. Then $G - uv$ has the same number of connected components as G: $n - m$. But $G - uv$ has n vertices and $m - 1$ edges, so it must have at least $n - m + 1$ connected components: contradiction!

3 Leaves in trees

A vertex of degree 1 in a graph is called a **leaf**. This terminology is mostly invented for use with trees, because leaves are very important here. Here are all the possible 5-vertex trees; all of them have leaves:



At least one leaf must exist (with one exception) by another theorem we have already proved about cycles:

Theorem 3.1 (Theorem 2.1 from Lecture 6)**.** *If a graph has minimum degree* 2*, then it contains a cycle.*

Trees have no cycles, so they cannot have minimum degree 2. In other words, every tree has a vertex with degree 1 (a leaf) or degree 0 (an isolated vertex).

An isolated vertex is its own connected component, so it can't exist in a tree that has any other vertices. (However, a 1-vertex graph with 0 edges is a perfectly good tree.) In a tree with at least 2 vertices, we must get at least 1 leaf.

Slightly more is true:

Theorem 3.2. *Any tree with at least* 2 *vertices has at least* 2 *leaves.*

Proof. Suppose the tree has $n \geq 2$ vertices, but only one leaf (we know it has at least one).

Then we have a single vertex of degree 1, and the other $n-1$ vertices have degree at least 2. The sum of degrees is at least

$$\underbrace{2+2+\cdots+2}_{n-1 \text{ times}}+1 = 2n-1.$$

However, a tree must always have $n-1$ edges, so the sum of degrees is $2(n-1) = 2n-2$. It cannot also be at least $2n-1$ at the same time, so we arrive at a contradiction: there cannot be only one leaf. □

4 Induction on trees

The existence of leaves in trees is especially useful because it gives us a handy template for induction on trees. If G is a tree and v is a leaf, then $G-v$ is also a tree! The easiest way to check this is to check that $G-v$ has $n-1$ vertices (if G had n vertices), $n-2$ edges (still one less than the number of vertices), and is acyclic (because deleting a vertex can't create a cycle). Then, use condition 5 of Theorem 1.3.

So if we're proving a theorem about all trees, then we can induct on the number of vertices. To apply the induction hypothesis, pick a leaf v, and apply it to $G-v$. Then verify that the statement remains true if we add v back in.

Here is a silly example:

Theorem 4.1. *All trees are bipartite.*

Proof. We induct on n, the number of vertices in the tree. When $n = 1$, there is only one vertex u, and no edges; so the partition (A, B) with $A = \{u\}$ and $B = \varnothing$ is a bipartition. (All edges go between A and B, because there are no edges!)

Assume that all $(n-1)$-vertex trees are bipartite, and let G be an n-vertex tree. Let v be a leaf of G, and let w be its only neighbor. By the induction hypothesis, $G-v$ has a bipartition (A, B).

Without loss of generality, w is on side A of the bipartition. Then consider the partition $(A, B \cup \{v\})$. Every edge of G except for edge vw has one endpoint in A and one in B. As for vw, one of its endpoints (w) is in A, and the other (v) is in $B \cup \{v\}$. So this is a bipartition of G, and G must also be bipartite.

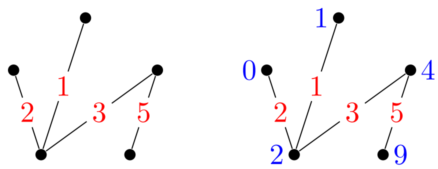
By induction, trees with any number of vertices are bipartite. □

It's a silly example, because there is a shorter, non-inductive proof. A graph is bipartite if and only if it has no odd cycles. Trees have no odd cycles: they have no cycles at all!

Here's another, fancier example.

Theorem 4.2. *Suppose every edge vw of a tree is labeled with a number y_{vw}. Then it's possible to also label each* vertex v *with a number x_v in such a way that for every edge vw, $y_{vw} = |x_v - x_w|$.*

Here is a diagram of the theorem statement. On the left are the edge labels, which we're given. On the right, we've also added vertex labels, which the theorem promises that we can always find. Note that there's no requirement that the labels be positive, or nonzero, or different.



Proof. We induct on n, the number of vertices in the tree. When $n = 1$, there is only one vertex u, and no edges; so there are no constraints, and we can make the label x_u be any number.

Assume that every $(n-1)$-vertex tree-labeling puzzle like this one has a solution, and let G be an n-vertex tree with a number on every edge. Let v be a leaf of G, and let w be its only neighbor. By the induction hypothesis, we can labe every vertex of $G - v$ to satisfy the theorem.

To extend this to a labeling of G, all we have to do is label v. There is only one thing that label has to do: we must have $|x_v - x_w| = y_{vw}$. So just set $x_v = x_w + y_{vw}$.

By induction, the theorem holds for trees on any number of vertices. $\square$

This is not a super fancy theorem, but it is adjacent to a big open problem in graph theory. The **graceful tree conjecture** says that there's some solution to this labeling problem in which:

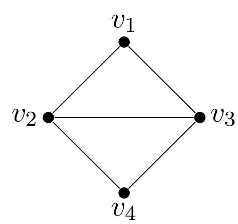
- The edge labels y_{vw} are a permutation of $1, 2, \ldots, n-1$.
- The vertex labels x_v are a permutation of $1, 2, \ldots, n$.

Though the question was asked in 1966, we still don't know the answer today. As of 2010, computer searches had verified the graceful tree conjecture for all trees with 35 vertices or fewer, according to Wikipedia.

5 Practice problems

1. Let T be tree whose degree sequence has the form $4, 3, 2, 1, 1, 1, \ldots$ (that is, $4, 3, 2$ followed by some number of 1's).
   (a) Determine the number of 1's in the degree sequence of T.
   (b) There is more than one possibility for a tree T with this degree sequence. Give two non-isomorphic trees with this degree sequence, and explain why they are not isomorphic.
2. Find all 6-vertex trees up to isomorphism. (There are six of them.)
3. Let G be a graph with 10 vertices and 10 edges.
   (a) If G contains exactly one cycle, how many connected components must it have? Give an example of such a graph.
   (b) If G contains exactly two cycles, how many connected components must it have? Give an example of such a graph.
   (c) If G contains exactly three cycles, there's two possible values for the number of connected components. Why is that? Give examples for each possibility.

   (Note: we consider two cycles to be the same if they are the same cycle subgraph *of G. For example, the cycle (x, y, z, x) is the same cycle as (z, y, x, z).)*
4. Find all 8 spanning trees of the graph below. (I do not care about isomorphism in this problem; if several spanning trees are isomorphic but different graphs, list them all.)



5. Give an example to demonstrate that Theorem 4.2 does not necessarily hold for a graph that is not a tree.
6. (a) Put together results we already know to show that if v, w are vertices of a graph G in the same connected component, then adding edge vw to G creates at least one new cycle.
   (b) Imitate the proof of Theorem 1.2 to prove that a graph with n vertices and m edges has at least $m - n + 1$ cycles. *(Hint: use part (a).)*
7. For graphs in general, we need a complicated procedure (the Havel–Hakimi algorithm) to determine whether a degree sequence is realizable. For trees, however, the rule is very simple: given any sequence of $n \geq 2$ positive integers whose sum is $2(n - 1)$, there is a tree with that degree sequence!

   The proof *almost*, but not entirely, follows the strategy in Section 4. The only difference is that we're trying to prove the existence of a tree, so we can't start by taking a tree and deleting a leaf. Instead, we do a corresponding operation on our degree sequence...