Math 3322: Graph Theory[1] Mikhail Lavrov

Lecture 11: Trees

September 21, 2021 *Kennesaw State University*

1 Counting edges in trees and forests

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. *If a graph has n vertices and at least n edges, then it contains a cycle.*

As a result, n-vertex trees and forests can have at most $n-1$ edges, because we don't want then to have any cycles. We know that a tree is a maximally acyclic graph: it has no cycles, but if you add any edge, then a cycle is created. This proves a fact from the previous lecture: if a graph has at least $n-1$ edges and no cycles, then in fact it has exactly $n-1$ edges, and is maximally acyclic (one more edge, and the theorem guarantees a cycle), so it's a tree.

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.

[1]This document comes from the Math 3322 course webpage: http://facultyweb.kennesaw.edu/mlavrov/courses/3322-fall-2021.php

- 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.

  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. □

In particular, a connected graph has only one connected component, so it must have at least $n - 1$ edges. If a connected graph has *at most* $n - 1$ edges, then it must have exactly $n - 1$. Such a graph must be minimally connected: delete any edge, and you will have only $n - 2$ edges, which is not enough to be connected. Therefore a connected graph with $n - 1$ edges must be a tree.

We conclude that all trees have $n - 1$ edges exactly. Any fewer, and it wouldn't be connected; any more, and it wouldn't be acyclic.

What about forests? Suppose an n-vertex forest has k connected components. Each component is a tree, so if the i^{th} component has n_i vertices, then it has $n_i - 1$ edges. The total number of edges in the forest is

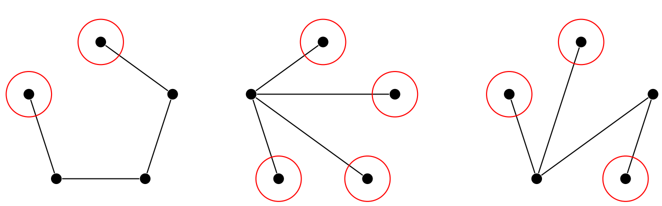
$$(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) = (n_1 + n_2 + \cdots + n_k) - k = n - k.$$

By the theorem we just proved, 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!

2 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. All the trees we've seen have multiple leaves (circled in red below):



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

Theorem 2.1. *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 2.2. *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. $\square$

3 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).

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 3.1. *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.

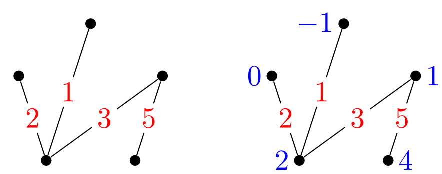
By induction, trees with any number of vertices are bipartite. $\square$

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 3.2. *Suppose every edge of a tree is labeled with a number. Then it's possible to also label each* vertex *with a number in such a way that the label on every edge is the sum of the labels on its endpoints.*

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 put any number we like on u.

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 label the vertices of $G - v$ in such a way that the label on every edge is the sum of the labels on its endpoints.

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: the numbers on v and on w have to add up to the number on edge vw. So just label v with the value (# on edge vw) − (# on vertex w).

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. Suppose we change the rules slightly: each edge must be labeled by the absolute value of the difference between the numbers on the endpoints. The **graceful tree conjecture** says that we can pick labels for the vertices and edges in such a way that the vertices are labeled 0 through $n-1$ (in some order) and the edges are labeled 1 through $n-1$ (in some order).

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.