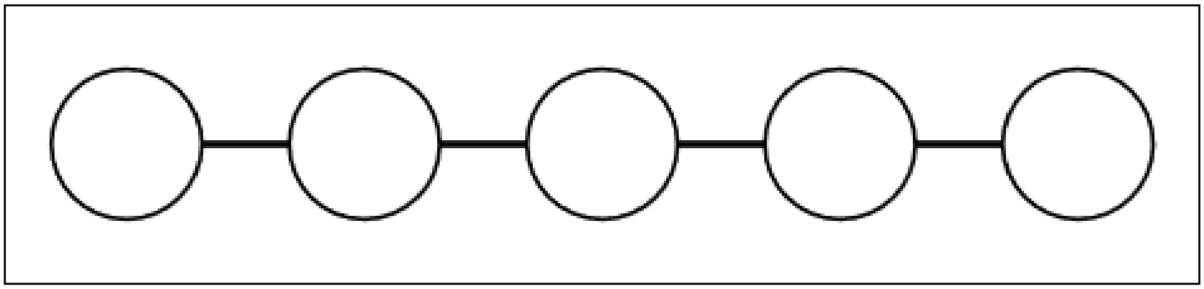
15.5 Cycles and Trees

Graph connectivity is one of the most fundamental properties that a graph can have, and so it is useful in practice to be able to efficiently determine whether a graph is connected or not. In the previous section, we developed an algorithm for checking whether two vertices in a graph are connected. It is possible to extend this to an algorithm that checks whether an entire graph is connected, and a good exercise to do so!

However, over the next few sections we're going to look at graph connectivity through a slightly different lens: determining *theoretical properties* that can help us determine whether a graph is connected or not. As we've seen a few times in these course notes, understanding such properties can help us develop new algorithms for solving problems, and graph connectivity is no exception.

A "lower bound" for connectivity

Suppose we have a graph with n vertices. We'll start by asking the following question: how many *edges* (in terms of n) must this graph have in order to be connected? Or put another way, what is the *minimum* number of edges required in a connected graph?[1]

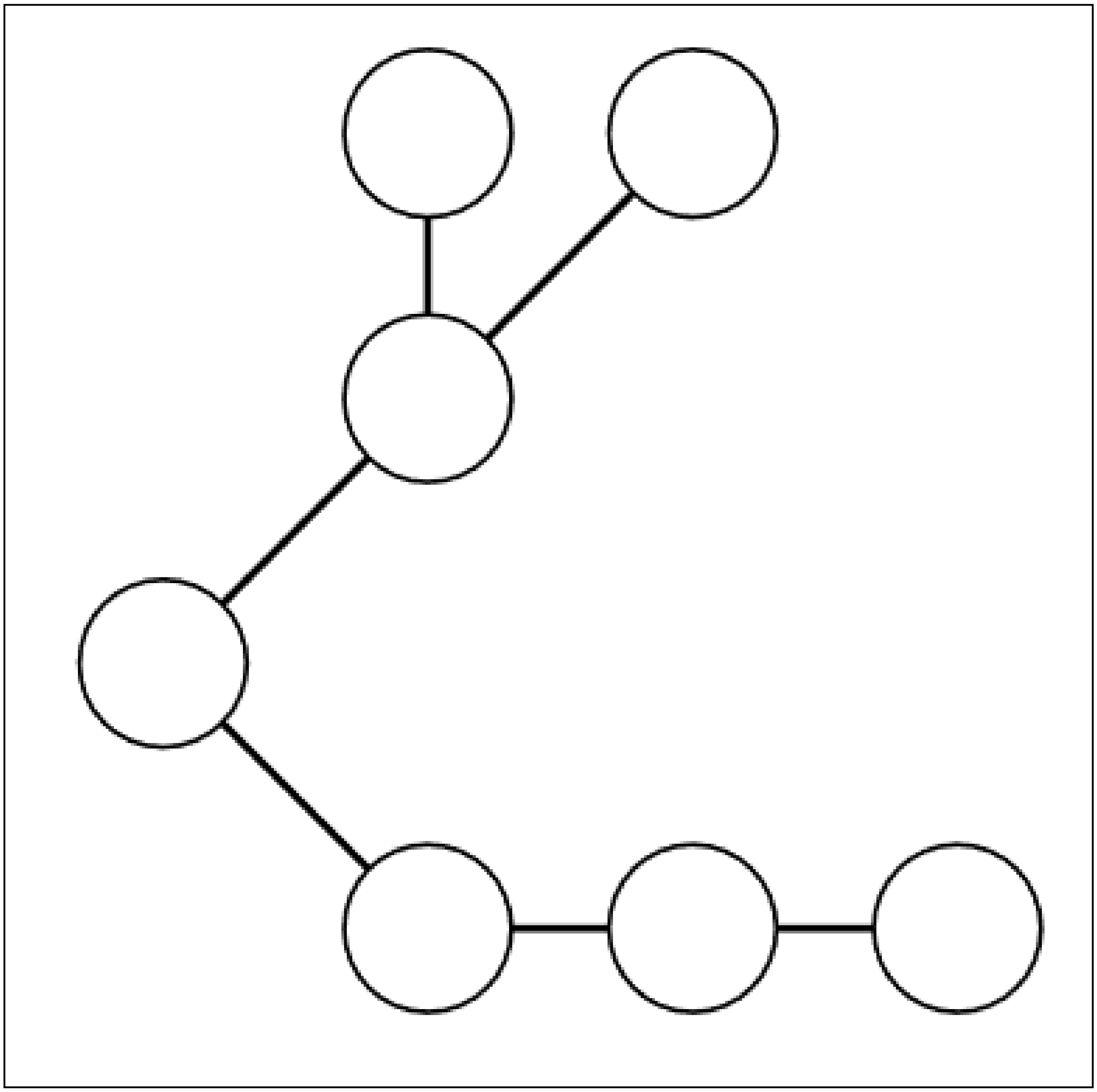


We might consider some simple examples to gain some intuition here. For example, suppose we have a graph with n vertices which is just a path, as shown on the right. This has $n - 1$ edges, and if you remove any edge from it, the resulting graph will be disconnected (we leave a proof of this as an exercise).

But this isn't the only possible configuration for such a graph. The graph on the right certainly isn't a path, but removing any edge from this graph disconnects it as well. And you might notice that the number of edges is again one fewer than the number of vertices.

It turns out that these examples do give us the right intuition: *any* connected graph $G = (V, E)$ must have $|E| \geq |V| - 1$.[2] The tricky part is proving this. Once again, we must struggle with the fact that even though the previous examples gave us some intuition, it is a challenge to generalize these examples to obtain an argument that works on all graphs

[1] Yet another way: how many edges are *necessary* for a graph to be connected?



satisfying these vertex and edge counts.

Cycles

To get a formal proof, we'll need some way of characterizing exactly when we can remove an edge from a graph without disconnecting it. The following definition is an excellent start.

Definition. Let $G = (V, E)$ be a graph. A **cycle** in G is a sequence of vertices $v_0, \dots, v_k$ satisfying the following conditions:

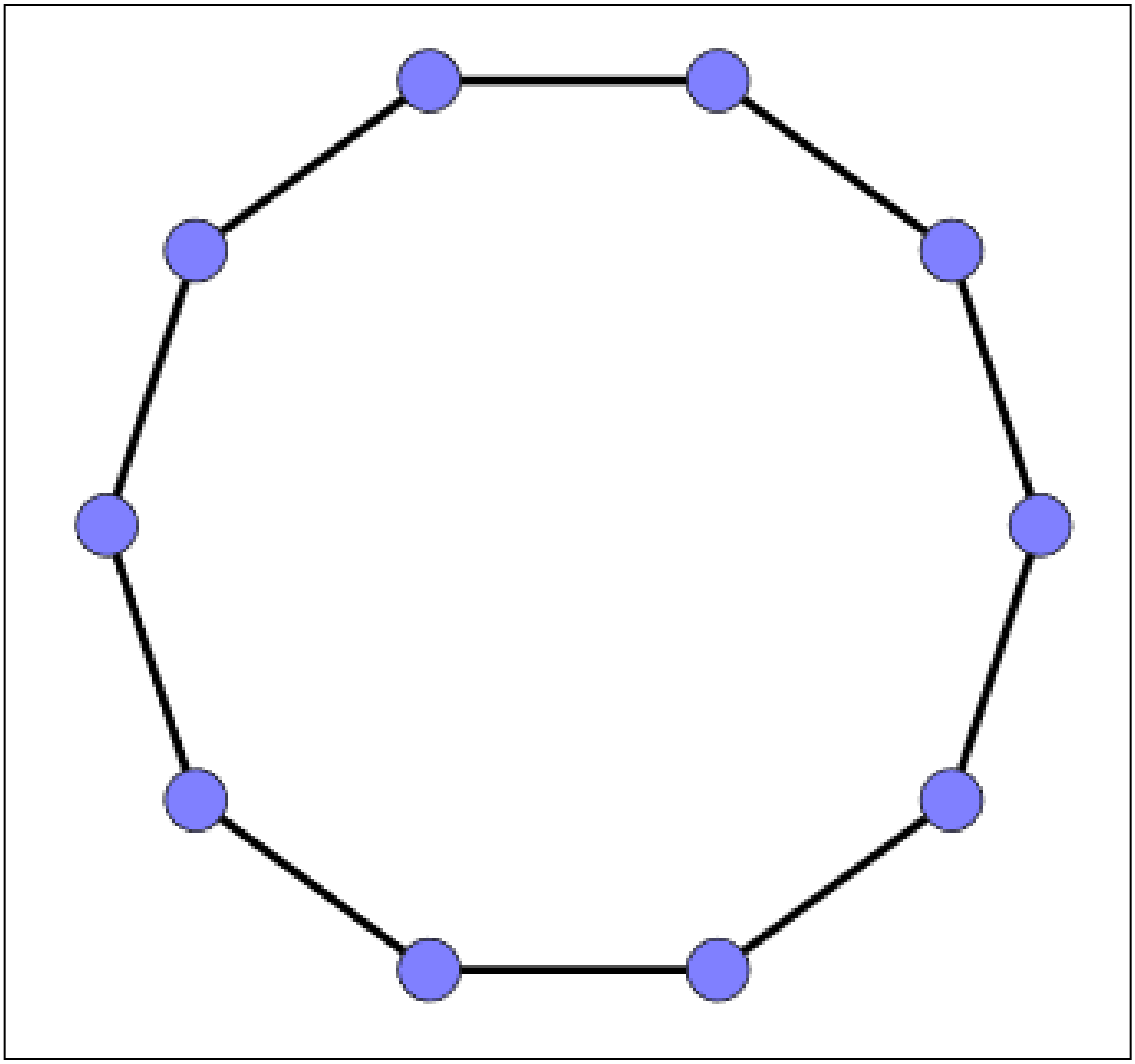
- $k \geq 3$

[2] The contrapositive is also an interesting statement: if a graph has fewer than $|V| - 1$ edges, it cannot be connected.

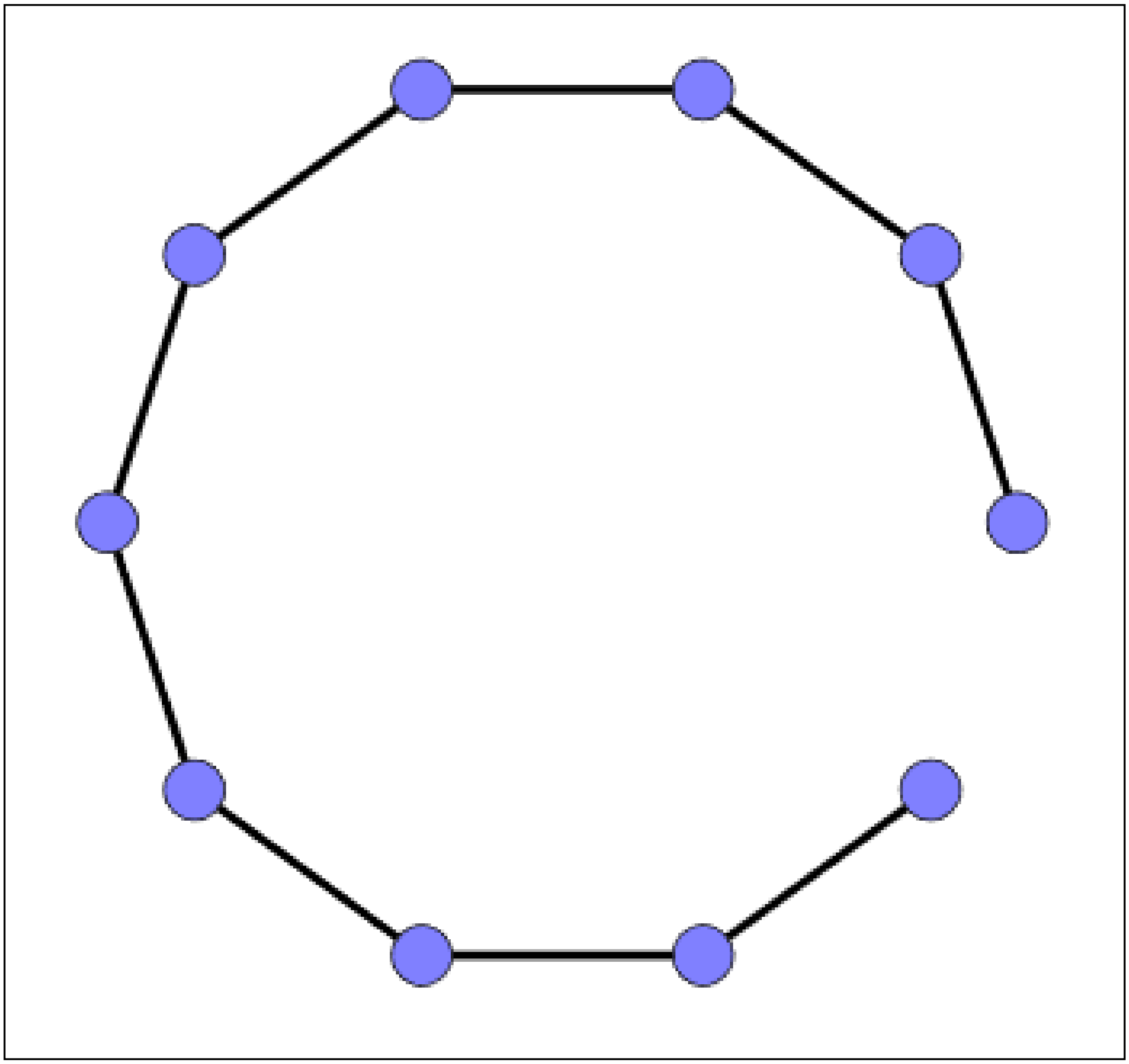
- $v_0 = v_k$, and all other vertices are distinct from each other and v_0
- each consecutive pair of vertices is adjacent

In other words, a cycle is like a path, except it starts and ends at the same vertex. The length of a cycle is the number of edges used by the sequence, which is also the number of distinct vertices in the sequence (the above notation describes a cycle of length k). Cycles must have length at least three; two adjacent vertices are not considered to form a cycle.

To use our example of cities and roads, if there is a cycle in the graph, it is possible to make a trip which starts and ends at the same city, and travels no road or city more than once.



Getting back to our motivation, cycles are a form of "connectedness redundancy" in a graph. Vertices in a cycle are all obviously connected to each other, but even if one edge is removed, the result is a path. In this case, the cycle's vertices are still connected to each other—albeit with possibly a much longer path to travel. Even though the diagrams on the right illustrate this property for a cycle itself, we will now show that this property holds even when this cycle is part of a larger graph.



Lemma. Let $G = (V, E)$ be a graph and $e \in E$. If G is connected and e is in a cycle of G, then the graph obtained by removing e from G is still connected.

Translation. There are a lot of quantified variables here, and some assumptions which are perhaps not obvious from the English. It is certainly a worthwhile exercise to translate this statement explicitly. The trickiest part is the condition on e (that it is part of a cycle of G); remember that we generally represent such conditions as assumptions in a logical implication.

For brevity, we will use the notation $G - e$ to represent the graph obtained by removing edge e from G.

$$\forall G = (V, E),\ \forall e \in E,\ \ (G \text{ is connected} \land e \text{ is in a cycle of } G) \Rightarrow G - e \text{ is connected.}$$

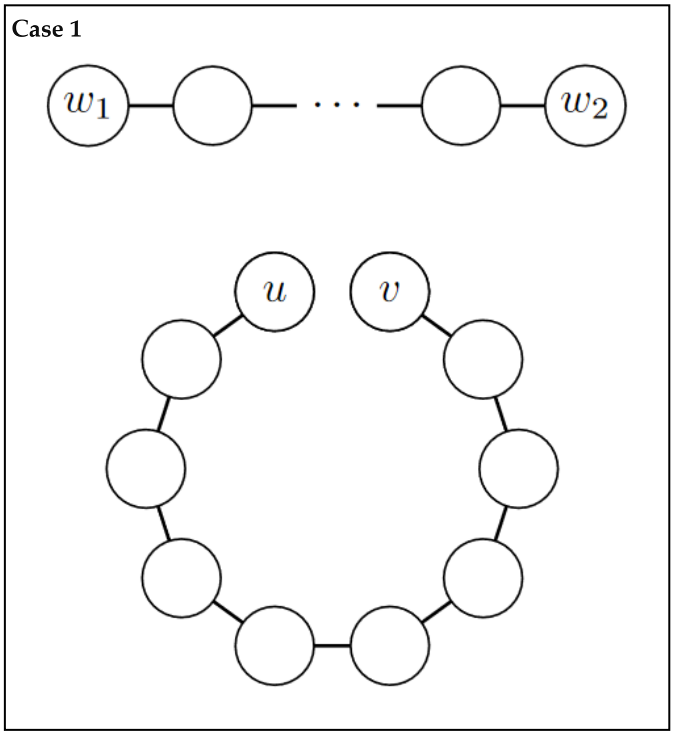
Discussion. This is a statement about a particular transformation: if we start with a connected graph and remove an edge in a cycle, then the resulting graph is still connected.

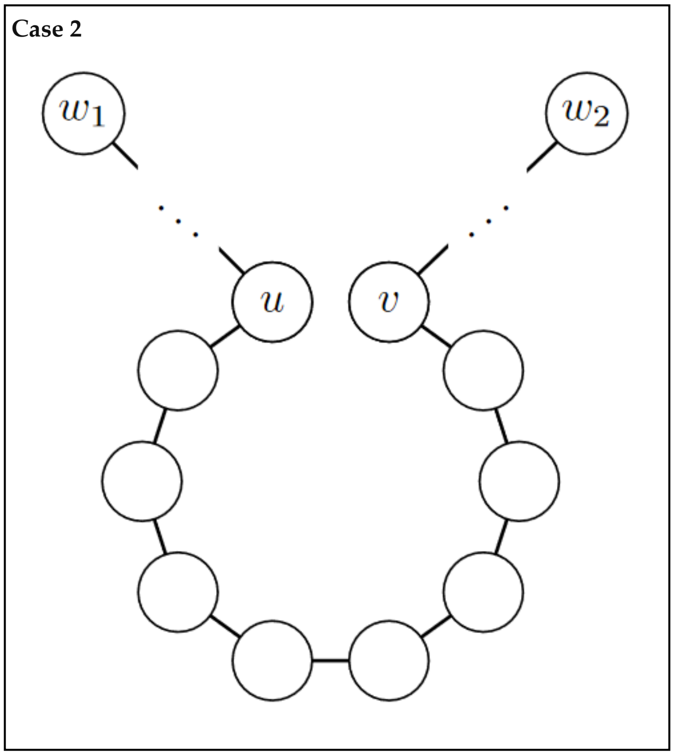
We get to assume that the original graph is connected and has a cycle, but that's it. We don't know anything else about the graph's structure, nor even which edge in the cycle e is.

That said, it seems like we should be able to simply make an argument based on the transitivity of connectedness: if we remove the edge $\{u, v\}$ from the cycle, then we already know that u and v are still connected, so all the other vertices should still be connected too).

Proof. Let $G = (V, E)$ be a graph, and $e \in E$ be an edge in the graph. Assume that G is connected and that e is in a cycle. Let $G' = (V, E \setminus \{e\})$ be the graph formed from G by removing edge e. We want to prove that G' is also connected, i.e., that any two vertices in V are connected in G'.

Let $w_1, w_2 \in V$. By our assumption, we know that w_1 and w_2 are connected in G. We want to show that they are also connected in G', i.e., there is a path in G' between w_1 and w_2.





Let P be a path between w_1 and w_2 in G (such a path exists by the definition of connectedness). We divide our proof into two cases: one where P uses the edge e, and another where it does not.

Case 1: P does not contain the edge e. Then P is a path in G' as well (since the only edge that was removed is e).

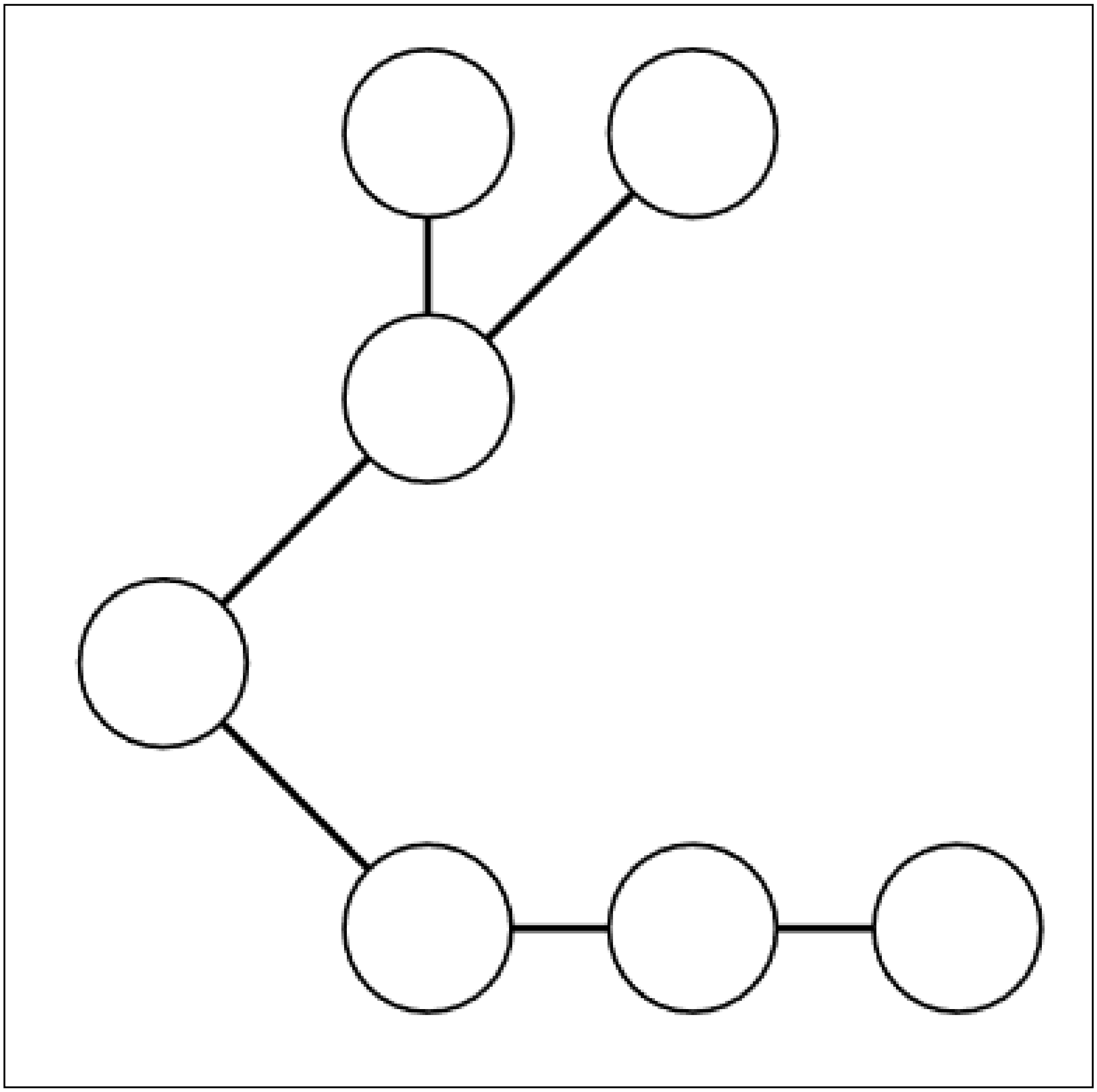
Case 2: P does contain the edge e. Let u be the endpoint of e which is closer to w_1 on the path P, and let v be the other endpoint.

This means that we can divide the path P into three parts: P_1, the part from w_1 to u, the edge $\{u, v\}$, and then P_2, the part from v to w_2. Since P_1 and P_2 cannot use the edge $\{u, v\}$—no duplicates—they must be paths in G' as well. So then w_1 is connected to u in G', and w_2 is connected to v in G'. But we know that u and v are also connected in G' (since they were part of the cycle), and so by the transitivity of connectedness, w_1 and w_2 are connected in G'.

■

Graphs with no cycles

The previous lemma told us that if we have a connected graph with a cycle, it is always possible to remove an edge from the cycle and still keep the graph connected. Since we are interested in talking about the minimum number of edges necessary for connecting a graph, we'll now think about graphs which don't have any cycles.



Definition. Let $G = (V, E)$ be a graph. We say that G is a **tree** when it is connected and has no cycles.

Wait a second, we already defined trees back in Chapter 13! It turns out that we can view trees as just a special type of graph, one that has no cycles. The one subtle difference is that the trees we studied earlier had a root, representing the "start" or "top" of a tree. With this graph-based definition of trees, we don't have a designated root element, and won't necessarily draw trees top-down, as illustrated by the diagram on the right.

Now, let's return to why we're studying trees in this chapter. We would like to say that trees are the "minimally-connected" graphs: that is, the graphs which have the fewest number of edges possible but are still connected. It may be tempting to simply assert this based on the definition and what we have already proven, but let G be a connected graph, and consider the following statements carefully:

1. *If* G has a cycle, *then* there exists an edge e in G such that $G - e$ is connected.
2. *If* G does not have a cycle, *then* there does not exist an edge e in G such that $G - e$ is connected.

We know that (1) is true by the previous example. Does it follow that (2) is also true?

No! The two statements may look very similar, but they are *not* logically equivalent. In fact, (2) is logically equivalent to the *converse* of (1): if we let p be the statement "G has a cycle" and q be the statement "there exists an edge e in G such that $G - e$ is connected," then (1) is simply $p \Rightarrow q$, while (2) is $\neg p \Rightarrow \neg q$.

So we actually need to prove this second statement directly, which is what we'll do next.

Lemma. Let G be a graph. If G does not have a cycle, then there does not exist an edge e in G such that $G - e$ is connected.

Translation.

$$\forall G = (V, E),\ G \text{ does not have a cycle} \Rightarrow \neg\big(\exists e \in E,\ G - e \text{ is connected}\big).$$

In general, having to prove that there does *not* exist some object satisfying some given conditions is challenging; it is often easier to assume such an object exists, and then prove that its existence violates one or more of the given assumptions. This can be formalized by writing the contrapositive form of our original statement.

$$\forall G = (V, E),\ \big(\exists e \in E,\ G - e \text{ is connected}\big) \Rightarrow G \text{ has a cycle}.$$

Discussion. So we can *assume* that there exists an edge e with this nice property that removing it keeps the graph connected. From this, we need to prove that G has a cycle. Note that we only need to show that a cycle exists—it may or may not have anything to do with e, but it is probably a good bet that it does.

The key insight is that if we remove e, we remove one possible path between its endpoints. But since the graph must still be connected after removing e, there must be another path between its endpoints.

Proof. Let $G = (V, E)$ be a graph. Assume that there exists an edge $e \in E$ such that $G - e$ is still connected.

Let $G' = (V, E \setminus \{e\})$ be the graph obtained by removing e from G. Our assumption is that G' is connected.

Let u and v be the endpoints of e. By the definition of connectedness, there exists a path P in G' between u and v; this path does not use e, since e isn't in G'. Then taking the path P and adding the edge e to it is a cycle in G. ■

Thus we now can state and prove the following fact about trees.

Lemma. Let G be a tree. Then removing any edge from G results in a graph that is not connected.

Proof. This follows directly from the previous lemma. By definition, G does not have any cycles, and so there does not exist an edge that can be removed from G without disconnecting it. ■

We can say that a tree is the "backbone" of a connected graph. While a connected graph may have many edges and many cycles, it is possible to identify an underlying tree structure in the graph that, if it remains unchanged, ensures the graph remains connected, regardless of any other edges removed.[3]

[3] In fact, many such trees may exist. This insight is the basis of *minimum spanning trees*, a well-studied problem in computer science that you will learn about in future courses.

Counting tree edges

Now, let us return to our original motivation of counting edges to prove the following remarkable result, which says that the number of edges in a tree depends *only* on the number of vertices.

Theorem. (**Number of edges in a tree**) Let $G = (V, E)$ be a tree. Then $|E| = |V| - 1$.

Translation.

$$\forall G = (V, E),\ G \text{ is a tree} \Rightarrow |E| = |V| - 1.$$

Discussion. We have previously observed that this property seems to hold on trees that we drew ourselves. But of course this is not a formal proof, since we cannot assume anything about the particular structure of a tree.

A natural alternate strategy is to take a tree, remove a vertex from it, and use induction to show that the resulting tree satisfies this relationship between its numbers of vertices and edges.

This only works, though, if we can pick a vertex whose removal from G results in a tree—and in particular, results in a connected graph. To do this, we need to pick a vertex that is a *leaf* of the tree, i.e., a vertex that has degree one.

Rather than proceeding with the proof directly, we recognize that a likely claim we'll need to use in our proof is that picking such a leaf is always possible. Rather than embedding a subproof within the main proof, we will do it separately first.

Lemma. Let $G = (V, E)$ be a tree. If $|V| \geq 2$, then G has a vertex with degree one.

Translation.

$$\forall G = (V, E),\ \big(G \text{ is a tree} \land |V| \geq 2\big) \Rightarrow \big(\exists v \in V,\ d(v) = 1\big).$$

Discussion. What does it mean for a vertex to have degree one? Intuitively, it means that we're at the "end" of the tree, and can't go any further. This makes sense visually on a diagram, but how can we formalize this? Suppose we start at an arbitrary vertex, and traverse edges to try to get as far away from it as possible. Because there are no cycles, we cannot revisit a vertex. But the path has to end somewhere, so it seems like its endpoint must have just one neighbour.

Proof. Let $G = (V, E)$ be a tree. Assume that $|V| \geq 2$. We want to prove that there exists a vertex $v \in V$ which has exactly one neighbour.

Let u be an arbitrary vertex in V. Let v be a vertex in G that is at the maximum possible distance from u, i.e., the path between v and u has maximum possible length (compared to paths between u and any other vertex). We will prove that v has exactly one neighbour.

Let P be the shortest path between v between u. We know that v has *at least* one neighbour: the vertex immediately before it on P. v cannot be adjacent to any other vertex on P, as otherwise G would have a cycle. Also, v cannot be adjacent to any other vertex w *not* on P, as otherwise we could extend P to include w, and this would create a longer path.

And so v has exactly one neighbour (the one on P immediately before v). ■

With this lemma in hand, we can now give a complete proof of the number of edges in a tree theorem. The key will be to use induction, removing from the original graph a vertex with just one neighbour, so that the number of edges also only changes by one. Our proof will use induction on the number of vertices, just like we saw back in Section 15.2. Here is the statement we'll prove:

$$\forall n \in \mathbb{Z}^+,\ \forall G = (V, E),\ (G \text{ is a tree} \land |V| = n) \Rightarrow |E| = n - 1.$$

Proof. We will proceed by induction on n, the number of vertices in the tree. Let $P(n)$ be the following predicate (over positive integers):

$P(n) : \forall G = (V, E),\ (G \text{ is a tree} \land |V| = n) \Rightarrow |E| = n - 1.$

We want to prove that $\forall n \in \mathbb{Z}^+,\ P(n)$.

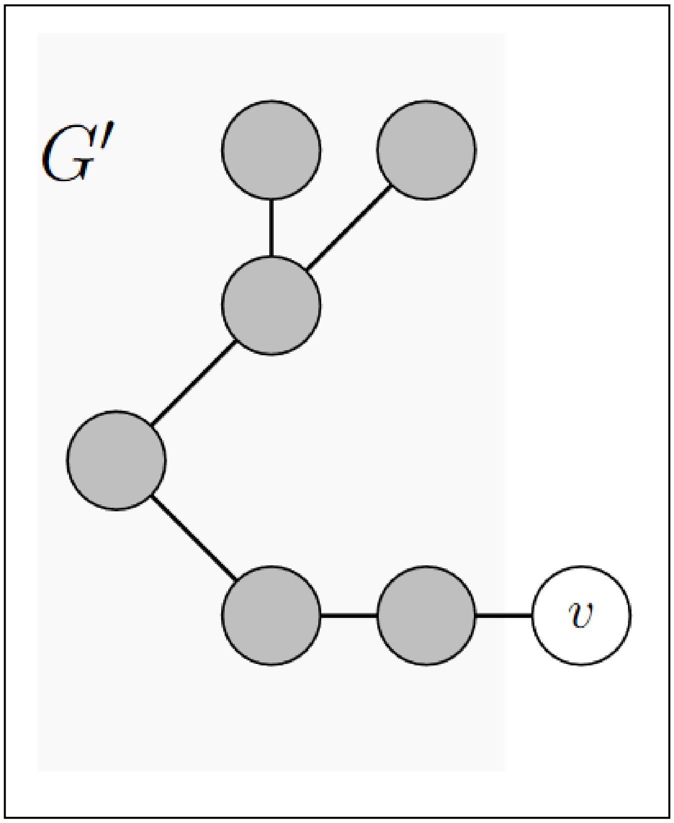
Case 1: Let $n = 1$. Let $G = (V, E)$ be an arbitrary graph, and assume that G is a tree with one vertex.

In this case, G cannot have any edges. Then $|E| = 0 = n - 1$.

Case 2: Let $k \in \mathbb{Z}^+$, and assume that $P(k)$ is true, i.e., for all graphs $G = (V, E)$, if G is a tree and $|V| = k$, then $|E| = k - 1$. We want to prove that $P(k + 1)$ is also true. Unpacking $P(k + 1)$, we get:

$\forall G = (V, E),\ (G \text{ is a tree} \land |V| = k + 1) \Rightarrow |E| = k.$

So let $G = (V, E)$ be a tree, and assume $|V| = k + 1$. We want to prove that $|E| = k$.



By the previous tree Lemma, since $k + 1 \geq 2$, there exists a vertex $v \in V$ that has exactly one neighbour. Let $G' = (V', E')$ be the graph obtained by removing v and the one edge on v from G. Then $|V'| = |V| - 1 = k$ and $|E'| = |E| - 1$.

We know that G' is also a tree. Then the induction hypothesis applies, and we can conclude that $|E'| = |V'| - 1 = k - 1$.

This means that $|E| = |E'| + 1 = k$, as required. ■

Putting it all together

Combining everything together, we can conclude the following required number of edges for any connected graph.

Since every connected graph contains at least one tree (just keep removing edges in cycles until you cannot remove any more), this constraint on the numbers of edges in a tree translates immediately into a lower bound on the number of edges in any connected graph (in terms of the number of vertices of that graph).

Theorem. Let $G = (V, E)$ be a graph. If G is connected, then $|E| \geq |V| - 1$.

So for example, suppose we have 100 cities that we’re connecting by airplane routes, and we want to be able to travel from one city to any other city (possibly by taking multiple flights). In this case, our previous theorem tells us that we’ll need *at least 99* routes between cities to make this possible. We might, of course, decide to open more routes, but we cannot open fewer without disconnecting cities from each other!

CSC110/111 Course Notes Home