A proof by induction: edges vs. connected components

Instructor: László Babai

December 8, 2022

This note describes the solution of a graph theory exercise. The solution illustrates how to do inductive proofs for structures such as graphs. In comments we highlight typical errors.

Recall that the **connected components** of a graph $G = (V, E)$ are the equivalence classes of the accessibility relation on V. So each connected component is a subset of V.

Exercise 0.1. A subset $C \subseteq V$ is a connected component if and only if
(i) the induced subgraph $G[C]$ is connected, and
(ii) there is no edge from C to $V \setminus C$.

Theorem 0.2. *Let G be a graph with n vertices, m edges, and k connected components. Prove: $m \geq n - k$. Proceed by induction on $n - k$.*

Let $i = n - k$. We have been instructed to prove the result by induction on i. Let $P(i)$ denote the result for a given value i, so for all $i \geq 0$, $P(i)$ denotes the following statement:

> $P(i)$: Let G be a graph with n vertices, m edges, and k connected components.
> If $n - k = i$ then $m \geq n - k$.

The Theorem says $(\forall i \geq 0)(P(i))$.

First we prove a lemma.

Lemma 0.3. *Let H be a graph with ℓ connected components. If we add an edge to H then the resulting graph H^* will have either ℓ or $\ell - 1$ connected components.*

Proof. Let $e = \{u, v\}$ be the added edge. If v is accessible from u in H then the addition of e does not change the connected components. If v is not accessible from u in H then their connected components are merged in H^* and none of the other connected components are affected. □

Now we prove Theorem 0.2.

BASE CASE: $P(0)$ (case $i = 0$). We need to show that $m \geq 0$ which is always true.

Typical error: assumes $n = 1$. Trouble: we are not doing induction on n. The base is not $n = 1$; it is $i = 0$. In other words, the base case must cover all possible values of n and all graphs with $k = n$ components. So the base case covers infinitely many graphs.

Having settled the case $i = 0$, we now consider the cases where $i > 0$ (our current value of i is positive).

INDUCTIVE HYPOTHESIS (IH). $P(i')$ is true for all $i' < i$.

This means that for all $i' < i$ (including $i' = 0$) and for all graphs with n' vertices, m' edges, and k' connected components, we have $m' \geq n' - k'$ as long as $i' := n' - k' < i$.

Typical errors. (1) Some solutions define the IH this way: "The theorem is true for $m' \geq n' - k'$ where $n' - k' < n - k$." The trouble with this statement is that $m' \geq n' - k'$ should be the conclusion of the statement, whereas "for" should be followed by the list of inputs for which the statement is assumed by the IH to be true.
(2) Some solutions define the IH this way: "If $m' \geq n' - k'$ then $m \geq n - k$ assuming $n' - k' < n - k$." The trouble here is that this is not a statement of the IH but a statement of what the Inductive Step needs to prove.
(3) Notational confusion arises when the solution does not make a clear distinction between the current value of the parameters and the "past values" of the parameters (the values for which the validity of the Theorem is assumed by the IH). It is also important that the current values and the corresponding past values be notationally related. For example, we indicate this connection by an apostrophy: i, n, m, k: current values, i', n', m', k': past values.

INDUCTIVE STEP. In this step we need to prove that $P(i)$ follows from the Inductive Hypothesis.

Proof. We are given a graph G with n vertices, m edges, and k connected components, where $n - k = i$. Desired Conclusion: $m \geq n - k$.

We wish to use the IH. To this effect, we shall construct a graph G' with some n' vertices, m' edges, and k' connected components such that $n' - k' < i$ so we can apply the IH to G'.

We shall not change the set of vertices; in particular, $n' = n$. This means we want $k' > k$. We shall increase the number of connected components by deleting edges from G.

Since $n - k = i \geq 1$, our input graph G must have at least one edge, say $e = \{u, v\}$. Let $C \subseteq V(G)$ be the connected component containing e. Let us now remove as few edges from the subgraph $G[C]$ induced by C as possible to make $G'[C]$ disconnected, where G' is the resulting graph after the edge-removal. Let F denote the set of removed edges; so $|F| \geq 1$. Now if we add back any of the edges from F to G', then C will again induce a connected subgraph (because $|F|$ is minimal), so, by the Lemma, $k = k' - 1$, where k' is the number of connected components of G'. Let m' be the number of edges of G'; so $m' \leq m - 1$. Now $i' := n' - k' < n - k = i$, so we can apply the IH to G'. We obtain that $m' \geq n' - k' = n - (k + 1) = n - k - 1$. Therefore $m \geq m' + 1 \geq (n - k - 1) + 1 = n - k$. □

Typical error. The biggest conceptual error in many proofs is an attempt to construct G from G'. This is not how induction works. G is given to us (it is our current input), and we need to prove the desired conclusion for G. To use the IH, we need to construct another graph, G', from G, to which we can apply the IH. So we get a conclusion about G'. Finally, we need to use this result to prove our desired conclusion about G.

Again, the lack of notational consistency and clear distinction between the parameters of G and G' may add to the confusion.

A simpler proof.
Based on the Lemma, it is easier to prove the Theorem by induction on m.

The base case is when $m = 0$ (but n is any positive integer).

The IH is that the Theorem holds for all graphs G' with $m' < m$ edges.

The inductive step takes our input graph G with n vertices, m edges, and k connected components; we assume $m \geq 1$ (since the case $m = 0$ was the base case). Now delete an edge to get a graph G' with $n' = n$ vertices, $m' = m - 1$ edges, and k' connected components. By the IH we have $m' \geq n' - k'$. By the Lemma, $k \geq k' - 1$. Therefore $m = m' + 1 \geq n' - k' + 1 = n - k' + 1 \geq n - (k + 1) + 1 = n - k$.

Third solution. We again obey the instruction to proceed by induction on $n - k$, with an additional trick.

Recall that a **forest** is a cycle-free graph. A connected forest is a tree, so each connected component of a forest induces a tree.

First we prove the Theorem for forests, in the following stronger form.

Theorem 0.4. *If a forest has n vertices, m edges, and k connected components, then $m = n - k$.*

Lemma 0.5. *If we delete an edge from a forest, the number of connected components increases by exactly 1.*

We leave the proof as an exercise to the reader.

Now we prove Theorem 0.4] by induction on $n - k$.

BASE CASE: $n - k = 0$. This means each vertex by itself is a connected component, so there are no edges, and therefore $m = 0 = n - k$.

INDUCTIVE HYPOTHESIS: Theorem true for forests with $n' - k' < n - k$.

INDUCTIVE STEP. Let G be a forest with n vertices, m edges, and k connected components. We assume $m \geq 1$ (since the case $m = 0$ was the base case).

Let e be an edge of G and let $G' = G - e$ (we delete the edge e but do not delete any vertices), so $n' = n$ and $m' = m - 1$. Then, by Lemma 0.5, G' has $k' = k + 1$ connected components. Therefore $m = m' + 1 = n' - k' + 1 = n - (k + 1) + 1 = n - k$. □

A **spanning forest** of a graph G is a forest H that is a spanning subgraph of G such that the connected components of H are the same as the connected components of G.

Lemma 0.6. *Every graph has a spanning forest.*

Proof. By induction on s, the number of cycles in G.
BASE CASE: $s = 0$. This means G is a forest already, we are done.

IH: Lemma 0.6 holds for graphs with $s' < s$ cycles.

Inductive step: We have $s \geq 1$. Let e be an edge that belongs to a cycle. Apply the IH to the graph $G' = G - e$ (which has fewer cycles than G). The resulting spanning forest of G' is also a spanning forest of G because the removal of e from G does not change the accessibility relation in G. □

Proof of Theorem 0.2. Let G^* be a spanning forest of G, with $n^* = n$ vertices, m^* edges, and $k^* = k$ connected components. Since G^* is a subgraph of G, it has $m^* \leq m$ edges. Now, by Theorem 0.4, $m \geq m^* = n^* - k^* = n - k$. □