A proof that BFS finds shortest paths

In class, I presented a proof that BFS actually computes shortest paths, a fact that is intuitively "obvious," and yet, a careful proof takes a bit of work. This handout gives the same proof, with some of the details expanded, and is provided for your reference.

First, some notation. There are n vertices in the graph, numbered $1 \ldots n$. The BFS starts at vertex r, which forms the root of the BFS tree, and a total of ℓ vertices are reachable from r (and hence visited during BFS). For a vertex v, we define

- $\mathit{level}[v]$ to be the level of v in the BFS tree,
- $\mathit{dist}[v]$ to be the actual distance from r to v in the graph, and
- $\mathit{pos}[v]$ to be the "position number" i of v, where $1 \le i \le \ell$, and v was the ith vertex to be inserted into the queue during BFS.

We want to show that $\mathit{level}[v] = \mathit{dist}[v]$ for all v. We do this by induction on $\mathit{pos}[v]$, and strengthen the induction hypothesis. Namely, we prove by induction on $i = 1 \ldots \ell$ that for v with $\mathit{pos}[v] = i$, we have

(I1) $\mathit{dist}[v] = \mathit{level}[v]$, and

(I2) for any vertex w, if $\mathit{dist}[w] < \mathit{dist}[v]$, then $\mathit{pos}[w] < \mathit{pos}[v]$.

The case $i = 1$ is trivially true, as the reader may verify. We now prove (I1) and (I2) for $1 < i \le \ell$, assuming (I1) and (I2) hold for all $i' < i$.

To prove (I1) for i, let $\mathit{pos}[v] = i$, let v' be the parent of v in the BFS tree. Suppose, by way of contradiction, that there is a path

$$r \rightsquigarrow w \to v$$

of length $h < \mathit{level}[v]$. First, we have

$$\mathit{pos}[v'] < \mathit{pos}[v], \tag{1}$$

since v is placed in the queue when v' is removed from the queue. Second, we have

$$\mathit{dist}[w] < \mathit{dist}[v'], \tag{2}$$

since by (1) we may apply the induction hypothesis (I1) at $\mathit{pos}[v']$, obtaining

$$\mathit{dist}[v'] = \mathit{level}[v'] = \mathit{level}[v] - 1 > h - 1 \ge \mathit{dist}[w'].$$

Third, we have

$$pos[w] < pos[v'], \tag{3}$$

since by (1) we may apply the induction hypothesis (I2) at $pos[v']$, together with (2), to obtain (3).

But now consider the point in time during the execution of BFS when w was removed from the queue. Since there is an edge $w \to v$, the BFS algorithm would visit v at this point in time, if it had not already at an earlier time. Thus, the parent of v has position number at most $pos[w]$, which by (3) is strictly less than $pos[v]$, and so v' cannot be the parent of v, as assumed — a contradiction.

To prove (I2), suppose $pos[v] = i$ and

$$dist[w] < dist[v]. \tag{4}$$

By way of contradiction, assume that $pos[w] \geq pos[v]$. Since by (4), $w \neq v$, it follows that $pos[w] \neq pos[v]$, and hence

$$pos[w] > pos[v]. \tag{5}$$

Let v' the parent of v in the BFS tree, and let

$$r \rightsquigarrow w' \to w$$

be a shortest path from r to w. This implies that $r \rightsquigarrow w'$ is a shortest path from r to w'. From this, we may conclude that

$$dist[w'] = dist[w] - 1. \tag{6}$$

First, just as in the proof of (I1), we have

$$pos[v'] < pos[v]. \tag{7}$$

Second, by hypothesis (I1) at $pos[v]$, which we just proved above, the tree path $r \rightsquigarrow v' \to v$ is a shortest path, and hence the tree path $r \rightsquigarrow v'$ is also a shortest path. From this, we may conclude that

$$dist[v'] = dist[v] - 1. \tag{8}$$

Now, (4), together with (6) and (8), imply that

$$dist[w'] < dist[v']. \tag{9}$$

Applying the induction hypothesis (I2) at $pos[v']$, which by (7) is less than $pos[v]$, along with (9), we obtain

$$pos[w'] < pos[v']. \tag{10}$$

Now, v was placed in the queue when v' was removed from the queue, and w was placed in the queue when w' was removed or at some earlier time. By (10), this means that w would be placed in the queue before v was placed in the queue, which contradicts (5).