CSCB63 Tutorial 6 — BFS produces shortest paths

An important result about Breadth-First Search is that it finds the shortest paths from the start vertex. In this tutorial, we will prove that BFS correctly computes distances from the start vertex to each other vertex.

```
0. BFS(s):
1.   for all v:
2.     v.seen := false
3.     v.d = inf
4.   queue := new Queue()
5.   queue.enqueue(s)
6.   s.seen := true
7.   s.d := 0                          // distance from s to s

8.   while not queue.is_empty():
9.     u := queue.dequeue()
10.      for each v in u's adjacency list:
11.       if not v.seen:
12.         v.seen := true
13.         v.d = u.d + 1              // distance from s to v
14.         queue.enqueue(v)
```

Definition 1. Let $G = (V, E)$ be a directed or undirected graph, with $v, u \in V$. Define the *shortest-path distance* $\delta(u, v)$ from u to v as the minimum number of edges on any path from u to v in G. If there is no path from u to v, then define $\delta(u, v) = \infty$.

Lemma 1. Let $G = (V, E)$ be a directed or undirected graph, let $s \in V$. Then for any edge $(u, v) \in E$,

$$\delta(s, v) \leq \delta(s, u) + 1$$

Proof. Suppose u is reachable from s. Then there is path p from s to v which is 1 edge longer than the shortest path from s to u. Thus, the shortest path from s to v is no longer than p. Thus, $\delta(s, v) \leq \delta(s, u) + 1$.

Suppose u is not reachable from s. Then $\delta(s, u) = \infty$ and so $\delta(s, v) \leq \delta(s, u) + 1 = \infty$. □

Lemma 2. Let $G = (V, E)$ be a directed or undirected graph, let $s \in V$, and suppose we run `BFS(s)`. Then for any vertex v, at each point during the execution of the algorithm (including at termination),

$$\delta(s, v) \leq v.d$$

Proof. By induction on the number of `enqueue` calls.

<u>Base case</u>: After the first call to `enqueue`, when s is enqueued, we have

$$\delta(s, s) = 0 \overset{line7}{=} s.d$$

$$\delta(s, v) \leq \infty \overset{line3}{=} v.d, \text{ for any } v \neq s$$

<u>Inductive case</u>: Assume $\forall v \in V, \delta(s, v) \leq v.d$ after k calls to `enqueue`. On line 9, we dequeue a vertex u. By induction hypothesis, we have $\delta(s, u) \leq u.d$. Before the next call to `enqueue`, we encounter node v via edge (u, v) (line 10). Then after the $k + 1^{st}$ call on line 14, we have:

$$\begin{aligned}
\delta(s,v) &\le \delta(s,u)+1 && \text{Lemma 1}\\
&\le u.d+1 && \text{IH}\\
&= v.d && \text{line 13}
\end{aligned}$$

The value $v.d$ is never changed again, since we mark the vertex seen (line 12) and only enqueue unseen vertices (line 11). □

Next we prove that at every point during BFS, the values $v.d$ of all nodes v in the queue are either all the same or look like this: ..., $k, k, k+1, k+1$, ... Formally:

Lemma 3. If during execution of BFS the queue contains vertices $(v_1, v_2, \ldots, v_n)$ where v_1 is at the head of the queue, then

$$v_n.d \le v_1.d + 1 \text{ and } v_i.d \le v_{i+1}.d \text{ for all } 1 \le i < n$$

Proof. By induction on the number of queue operations (both enqueue and dequeue).

Base case: queue contains 1 element, so $v_1 = v_n$.

Inductive case: Assume that after k queue operations, we have $v_n.d \le v_1.d + 1$ and $v_i.d \le v_{i+1}.d$ for all $1 \le i < n$.

Suppose we dequeue (line 9). Then the queue becomes either empty or $(v_2, \ldots, v_n)$. We have

$$\begin{aligned}
v_n.d &\le v_1.d+1 && \text{IH}\\
&\le v_2.d+1 && v_1.d \le v_2.d \text{ by IH}
\end{aligned}$$

Suppose we enqueue (line 14). Then the queue becomes $(v_1, \ldots, v_n, v_{n+1})$. Then by line 13, we have $v_{n+1}.d = u.d + 1$ for a node u which has just been dequeued on line 9. Before we dequeued, the queue was $(u, v_1, \ldots, v_n)$. Then

$$\begin{aligned}
v_{n+1}.d &= u.d+1 && \text{line 9}\\
&\le v_1.d+1 && \text{IH}
\end{aligned}$$

□

Lemma 4. Suppose vertex u is enqueued before vertex v during BFS. Then $u.d \le v.d$ at the time v is enqueued.

Proof. Follows from the previous Lemma. □

Theorem 1. Upon termination of BFS on a graph $G = (V, E)$ from a start vertex $s \in V$, for every node $v \in V$, $v.d = \delta(s, v)$.

Proof. By contradiction. Suppose there are vertices $v \in V$, for which upon termination of BFS, we have $v.d \ne \delta(s, v)$. Since by Lemma 2 we have $v.d \ge \delta(s, v)$, then $v.d > \delta(s, v)$. Of these vertices take the vertex v with the *smallest* $\delta(s, v)$. Then $v \ne s$, since $\delta(s, s) = 0 = s.d$, and v is reachable from s, since otherwise $\delta(s, v) = \infty = v.d$.

There exists some shortest path from s to v, in which some vertex u immediately precedes v. Because we chose v to be the vertex with smallest $\delta(s, v)$ that differs from $v.d$, it must be that $u.d = \delta(s, u)$. Then $\delta(s, v) = \delta(s, u) + 1 = u.d + 1 < v.d$.

Consider the moment u gets dequeued during BSF. At this point v is either seen or unseen. If it is seen, it was already enqueued, and so by Lemma 4 $v.d \le u.d$. Contradiction. If it is not seen, then on line 13 we set $v.d = u.d + 1$. Contradiction.

Note that to properly conclude the proof, we must show that every vertex v gets visited by BFS. Left as an exercise. □