We focus on the BFS algorithm as depicted in Figure 1. It performs a breadth-first search of $G$ starting at a node $s$.

\[
\begin{align*}
\text{BFS}(G, s): & \\
1 & \text{Unmark all vertices} \\
2 & \text{visit and mark } (s) \\
3 & \text{Enqueue}(Q, s) \\
4 & \text{BFS} \_\text{tree} \leftarrow \emptyset \\
5 & \text{while } Q \neq \emptyset \text{ do} \\
6 & \quad v \leftarrow \text{Dequeue}(Q) \\
7 & \quad \text{for every } w \in \text{Adj}[v] \text{ do} \\
8 & \quad \quad \text{if } w \text{ not yet visited then} \\
9 & \quad \quad \quad \text{visit and mark } (w) \\
10 & \quad \quad \text{Enqueue}(Q, w) \\
11 & \quad \text{BFS} \_\text{tree} \leftarrow \text{BFS} \_\text{tree} \cup \{(v, w)\}
\end{align*}
\]

Figure 1: The BFS Algorithm

We say that node $v$ is \textit{reachable} from $s$, denoted $s \leadsto v$, in $G$ if there is a path from $s$ to $v$ in $G$. By \textit{time} $r$ we shall refer to the $r^{th}$ time at which Line 5 (the start of the while loop) is reached.

\textbf{Definition 1} Define $\delta(v)$, node $v$'s distance (from $s$), to be number of edges on the shortest path from $s$ to $v$ if $v$ is reachable from $s$, and $\delta(v) = \infty$ otherwise.

We start with the following lemma that captures some of the basic properties of the BFS algorithm.
Lemma 1 Every time Line 5 is reached, as well as when the algorithm completes, the following hold:

1. All marked nodes are reachable from s in G;
2. If w /∈ Q is a marked node and (w, w') ∈ E, then w' is marked;
3. If Q ≠ Ø and the node v at the front of Q is of distance δ(v) = k then
   (a) The set of marked nodes consists exactly of all nodes of distance ≤ k, as well as all nodes of distance k + 1 that are in Q; and
   (b) The queue Q consists of a sequence of nodes of distance k followed by a (possibly empty) sequence of nodes of distance k + 1.

Proof: We prove the claim by induction on the number r of times Line 5 has been reached in the algorithm.

Base: For r = 1 we have that Q = ⟨s⟩, node s is marked, its distance is δ(s) = 0 and all nodes of distance > 0 are not in Q and are unmarked. Finally, there are no marked nodes outside of Q. All parts immediately follow.

Inductive Step: Assume that the claim holds for r. If Q = Ø there, then Lines 6-11 are skipped and the algorithm completes with the properties holding. Otherwise, Q ≠ Ø, Lines 6-11 are performed and Line 5 is reached once more at time r + 1. In the sequel, we shall denote the node at the front of Q at r by v. We shall show that the claims hold at r + 1.

1. Part 1 at time r + 1 follows from the fact that it holds at time r and that only neighbors w of v are added to Q between time r and r + 1. Since v is in Q at r, it is marked by part 3(a). Thus, by the inductive hypothesis for part 1, v is reachable from s and so w is also reachable, since there is a path s ↦ v → w in G.

2. The node v at the front of Q at time r is marked by part 3(a). It is removed from Q on Line 6, while all of its unmarked neighbors are marked by Lines 8 and 9. Thus, the claim is true for v. By the inductive hypothesis, part 2 holds at time r. It follows that part 2 holds at r + 1 since v is the only node that is both marked and not in Q at time r + 1 that was not so already at time r.

3. For part 3, observe by part 1 that the node v at the front of Q at r is reachable from s. Thus, δ(v) < ∞. Denote δ(v) = m. Lines 6–11 will remove v from Q and
add those of its neighbors that are unmarked at \( r \) to the end of \( Q \). By part 3(a) for time \( r \) we have that all added neighbors are unmarked and hence of distance \( > m \). But since \( \delta(v) = m \) we have that \( \delta(w) \leq m + 1 \) for every neighbor \( w \) of \( v \), since there is a path \( s \leadsto v \leadsto w \) of length \( m + 1 \). Thus, the added nodes are of distance \( m + 1 \). We consider two possibilities:

- **The node \( v' \) following \( v \) in \( Q \) at \( r \) is of distance \( \delta(v') = m \):** In this case part 3(a) holds at \( r + 1 \) by the inductive assumption together with the fact that by Lines 9 and 10 every node that is added to \( Q \) is marked. For part 3(b), the fact shown above that the nodes that are marked (on Line 9) and added to \( Q \) (on Line 10) are of distance \( m + 1 \), combined with the inductive assumption of 3(b) for time \( r \), imply the claim for time \( r + 1 \).

- **The node \( v' \) following \( v \) in \( Q \) at \( r \) is of distance \( \delta(v') = m + 1 \):** In this case, part 3(b) at time \( r \) implies that \( v \) is the only node of distance \( m \) in \( Q \) at \( r \). Parts 3(a) and 3(b) should hold at time \( r + 1 \) with respect to \( k = m + 1 \).

(a) Since the only nodes that were added to \( Q \) between \( r \) and \( r + 1 \) are of distance \( m + 1 \), we have from 3(b) at time \( r \) that there are no nodes of distance \( k + 1 = m + 2 \) in \( Q \). Thus, Moreover, by 3(a) no node of distance greater than \( m + 1 \) was marked at \( r \) and since only nodes of distance \( m + 1 \) were marked on lines 8 and 9, no nodes of distance \( > m + 1 \) are marked at \( r + 1 \). To complete 3(a) we need to show that all nodes of distance \( \leq m + 1 \) are marked at time \( r + 1 \). For all nodes of distance \( \leq m \) this is already true at time \( r \) and so also at time \( r + 1 \). For every node \( u' \) of distance \( m + 1 \) there is a node \( u \) of distance \( m \) such that \((u, u') \in E\). If \( u \neq v \) (where \( v \) is the node removed from \( Q \) on line 6), then at time \( r \) we had that \( u \) was marked and \( u \notin Q \). By the inductive assumption part 3 holds at \( r \), and it implies that \( u' \) is marked at \( r \), and remains marked at \( r + 1 \), as claimed. Alternatively, \( u = v \), in which case it is marked on line 9 if it has not been marked beforehand.

(b) Finally, for part 3(b) observe that \( k = m + 1 \) at \( r + 1 \). The inductive assumption guarantees that at time \( r \) the queue \( Q \) consists of a sequence of nodes of distance \( m \), followed by a sequence of nodes of distance \( m + 1 \). By assumption, the only node of distance \( m \) in \( Q \) at \( r \) is \( v \). Since \( v \) is removed on line 6, there are no nodes of distance \( m \) in \( Q \) at time \( r + 1 \). Finally, as argued above, the only nodes added to \( Q \) are of distance \( k = m + 1 \). It follows that all nodes in \( Q \) at \( r + 1 \) are of distance \( k \), which implies that the claim in part 3(b) holds at \( r + 1 \).
**Lemma 2**  Every node in $Q$ is eventually removed from $Q$.

**Proof:** Suppose that $u$ is in $Q$ at time $r$. Let $f$ be the number of nodes between $u$ and the front of $Q$. A straightforward induction on $f$ shows that $u$ will be removed by Line 6 immediately following time $r + f$. □

We can now prove the basic theorem that captures our intuition about the behavior of the BFS algorithm:

**Theorem 1**  The BFS algorithm marks all and only the nodes that are reachable from the source $s$ in $G$. Moreover, the order in which the nodes are visited respects their distance from $s$: $0, 1, \ldots, k, k + 1, \ldots$.

**Proof:** By part 1 of Lemma 1 we have that only nodes reachable from $s$ are ever visited and marked. For reachable nodes $v$, we prove by induction on $\delta(v)$ that $v$ will be marked before any nodes of distance $> \delta(v)$ are marked.

**Base:** If $\delta(v) = 0$ then $v = s$ is the source node. On Line 1 all nodes are unmarked. The source $s$ is marked on Line 2, before any of the other nodes are marked. The claim holds in this case.

**Inductive step:** Assume that the claim holds for all nodes of depth $m$, and let $\delta(v) = m + 1$. Since $\delta(v) = m + 1$ there is a node $w$ such that $\delta(w) = m$ and $(w, v) \in E$. By assumption, all such nodes $w$ are marked by the algorithm. Each of these nodes is enqueued on Line 10 immediately after being marked on Line 9. By Lemma 2 all these nodes are eventually removed from $Q$. When the first of these, say $w'$, is removed, the node $v$ is marked. Since $\delta(w') = m$, we have by Lemma 1 part 3(a) that all nodes of distance $m + 2$ or greater are unmarked, completing the inductive step. □