CMSC 351: Shortest Path Algorithm

Justin Wyss-Gallifent

April 12, 2022

1 Introduction .................................................. 2
2 Intuition ....................................................... 2
3 Pseudocode .................................................... 5
4 Pseudocode Time Complexity ............................... 6
5 Thoughts, Problems, Ideas .................................. 7
6 Python Test and Output ..................................... 8
1 Introduction

Consider the following graph:

Suppose we wish to find the shortest path between two vertices \(s\) and \(t\). How can we go about this?

2 Intuition

Intuitively the idea is this: First label all vertices with a (tentative distance) value of \(\infty\) and with no predecessor.

Then we we label the starting vertex \(s\) as having distance 0 from itself and we put 0 on a queue of vertices which need to be dealt with.

Next we repeat the following until \(t\) gets a distance label:

- Pop a vertex \(v\) off the queue, go through all of the \(\infty\) vertices \(v\) is connected to, label each with \(v\)'s distance plus 1 and label its predecessor as \(v\) and put each on the queue.

Once we label \(t\) we have found its distance and we can use the predecessor labels to find the shortest path back to the origin.

In truth the distance labels are not necessary since we can use the length of the shortest path to calculate the distance.

Example 2.1. Let’s see how this works on a really easy graph. In the following suppose we wish to find the shortest path path from vertex \(s = 0\) to vertex \(t = 7\):

The first thing we do is label all the vertices with a value of \(\infty\) except for \(s = 0\) itself which we label with a value of 0. We also assign the queue.
We then pop 0 off the queue and we go over all the ∞ vertices connected to 0 by an edge (those are 1,3) and we label those with the value 0 + 1 = 1 and their predecessor as 0. We note that vertex 7 is not amongst them. We also push them onto the queue.

We then pop 1 off the queue and we go over all the ∞ vertices connected to 1 by an edge (those are 2,4) and we label those with the value 1 + 1 = 2 and their predecessor as 1. We note that 7 is not amongst them. We also push them onto the queue.

We then pop 3 off the queue and we go over all the ∞ vertices connected to 3 by an edge (that is 6) and we label that with the value 1 + 1 = 2 and its predecessor as 3. We note that 7 is not amongst them. We also push it onto the queue.
We then pop 2 off the queue and we go over all the \( \infty \) vertices connected to 2 by an edge (that is 5) and we label that with the value \( 2 + 1 = 2 \) and its predecessor as 2. We note that 7 is not amongst them. We also push it onto the queue.

We then pop 4 off the queue and we go over all the \( \infty \) vertices connected to 4 by an edge (that is 7) and we label that with the value \( 2 + 1 = 3 \) and its predecessor as 4. We note that 7 is amongst them so we return it and are done.

The pseudocode does not push it onto the queue here because the push happens after the \texttt{return} conditional.

Since we’ve located vertex 7 we can backtrack using the predecessors. The predecessors go 4, 1, 0 so the path is \( \langle 0, 1, 4, 7 \rangle \) and has length 3.

Note that the predecessor array in this case is:

\[
P = [\text{NULL}, 0, 1, 0, 1, 2, 3, \text{NULL}]
\]

We extract the path by observing that:

\[
\]
3 Pseudocode

Here is the pseudocode. The distance and predecessor data is stored in lists (if each vertex is an object these can instead just be properties of the object). The list of vertices to deal with next is stored in a queue.

\[ PRE: G \text{ is a graph with } n \text{ vertices.} \]
\[ PRE: s \text{ and } t \text{ are start and end vertices.} \]

\[ \text{function shortestpath}(G,s,t) \]
\[ \text{dist} = \text{distance array of size } n \text{ full of inf} \]
\[ \text{pred} = \text{predecessor array of size } n \text{ full of null} \]
\[ Q = \text{empty queue} \]
\[ \text{dist}[s] = 0 \]
\[ \text{\text{\textbackslash Push } s \text{ onto } Q \text{ so it’s the next (first) vertex we deal with.}} \]
\[ Q.\text{push}(s) \]
\[ \text{\text{\textbackslash While there are vertices to deal with,}} \]
\[ \text{\text{\textbackslash pop one off and analyze its adjacent nodes.}} \]
\[ \text{\text{\textbackslash If unlabeled, label and push onto the queue.}} \]
\[ \text{while } Q \text{ is nonempty} \]
\[ x = Q.\text{pop} \]
\[ \text{for each infinity vertex } y \text{ connected to } x \]
\[ \text{dist}[y] = \text{dist}[x] + 1 \]
\[ \text{pred}[y] = x \]
\[ \text{if } y == t \]
\[ \text{return}\text{pred} \]
\[ \text{end} \]
\[ Q.\text{push}(y) \]
\[ \text{end} \]
\[ \text{end} \]
\[ \text{\text\textbackslash POST: Returns pred, from which the path can be extracted.} \]

Note that some pseudocode you’ll see for this include an array which keeps track of whether each vertex has been visited or not. This isn’t necessary since this can be determined based upon whether or not a vertex has a finite distance assigned.
4 Pseudocode Time Complexity

The best-case is easy - this can be $\Theta(1)$ if $s = t$

The worse-case is not as obvious. It's tempting to look at the pseudocode for this but we need to be careful. Let's see why.

Let's assume for now that the initialization step before the while loop is $\Theta(1)$.

Since no vertex ever gets put on the queue more than once the while loop iterates at most $n$ times.

It's tempting to think that if each vertex is connected to every other vertex that the inner for each edge loop might also run $n$ times, giving a time complexity of $O(n^2)$. However while not wrong, this goes a little overboard because we can see that if each vertex were connected to every other vertex that $t$ would be found immediately.

Instead, first note that the while loop will never iterate more than $n - 1$ times since each time corresponds to a pop and we will never pop more than $n - 1$ nodes off the queue. We would not, for example, pop all $n$ off because we would have found the target before then.

Then note that the for loop is technically following edges and the same edge will never be followed twice.

Thus in a worst case we have:

- Some constant-time operations before the while loop.
- The body of the while loop i not counting the for loop, iterates at most $n - 1$ times, taking constant time for each.
- The body of the for loop iterates at most $E$ times, where $E$ is the number of edges, taking constant time for each.

Thus we can say for certain that in the worst-case:

$$T(n, E) = \Theta(c + n - 1 + E) = \Theta(n + E)$$
5 Thoughts, Problems, Ideas

1. Is it more advantageous to have a graph $G$ represented by an adjacency matrix or an adjacency list in order to implement this pseudocode algorithm in actual code? Explain.

2. Modify the pseudocode so that it returns the length of the shortest path from $s$ to $t$.

3. This algorithm may be modified to find a shortest path tree by not targeting a specific vertex but rather proceeding until all vertices have been accounted for. In this case the predecessor list which is returned can be used to reconstruct the entire tree. Give the pseudocode for this algorithm.

4. Suppose the graph were weighted instead of unweighted. If the line

   \[ \text{dist}[y] = \text{dist}[x] + 1 \]

were replaced by

   \[ \text{dist}[y] = \text{dist}[x] + \text{edgeweight}(x \text{ to } y) \]

would this effectively find the shortest total weighted distance from $s$ to $t$? If so, explain why. If not, explain why with a graph.

5. Modify the shortest path algorithm so that it finds the shortest path from $s$ to itself (a cycle) and returns FALSE if no such cycle exists.
6 Python Test and Output

The following code is applied to the graph above. This follows the model of the pseudocode and in addition creates and returns a list of the vertices in the order in which they were visited.
def shortestpath(AL, n, u, up):
    dist = [float('inf')] * n
    pred = [None] * n
    Q = []
    dist[u] = 0
    print('Dist: ' + str(dist))
    Q.append(u)
    print('Queue: ' + str(Q))
    while len(Q) != 0:
        x = Q.pop(0)
        print('Pop ' + str(x) + ' and process vertices ' + str(AL[x]))
        for y in AL[x]:
            if dist[y] == float('inf'):
                print('Process: ' + str(y))
                dist[y] = dist[x] + 1
                print('Dist: ' + str(dist))
                pred[y] = x
                if y == up:
                    return (dist, pred)
                Q.append(y)
                print('Queue: ' + str(Q))
            else:
                print('Already done: ' + str(y))

AL = [
    [1, 3],
    [0, 2, 4],
    [1, 5],
    [0, 4, 6],
    [1, 3, 7],
    [2, 8],
    [3, 8],
    [4],
    [5, 6]
]

n = 9

u = 0
up = 7
[dist, pred] = shortestpath(AL, n, u, up)

path = []
x = up
while x != None:
    path.append(x)
    x = pred[x]
path.reverse()
print('Path: ' + str(path))
print('Length: ' + str(len(path) - 1))
Output:

```
Dist: [0, inf, inf, inf, inf, inf, inf, inf, inf]
Queue: [0]
Pop 0 and process vertices [1, 3]
__Process: 1
__Dist: [0, 1, inf, inf, inf, inf, inf, inf, inf]
__Queue: [1]
__Process: 3
__Dist: [0, 1, inf, 1, inf, inf, inf, inf, inf]
__Queue: [1, 3]
Pop 1 and process vertices [0, 2, 4]
__Already done: 0
__Process: 2
__Dist: [0, 1, 2, 1, inf, inf, inf, inf, inf]
__Queue: [3, 2]
__Process: 4
__Dist: [0, 1, 2, 1, 2, inf, inf, inf, inf]
__Queue: [3, 2, 4]
Pop 3 and process vertices [0, 4, 6]
__Already done: 0
__Already done: 4
__Process: 6
__Dist: [0, 1, 2, 1, 2, inf, 2, inf, inf]
__Queue: [2, 4, 6]
Pop 2 and process vertices [1, 5]
__Already done: 1
__Process: 5
__Dist: [0, 1, 2, 1, 2, 3, 2, inf, inf]
__Queue: [4, 6, 5]
Pop 4 and process vertices [1, 3, 7]
__Already done: 1
__Already done: 3
__Process: 7
__Dist: [0, 1, 2, 1, 2, 3, 2, 3, inf]
Path: [0, 1, 4, 7]
Length: 3
```