Recitation 8 – *Shortest Paths and DFS Numberings*

Parallel and Sequential Data Structures and Algorithms, 15-210 (Fall 2014)

October 14th, 2014

1 Announcements

- How's *ThesaurusLab*?
- *AbridgedLab* has been released!
- Mid-Semester Break Plans?
- More than half of the labs are done!

2 That's Some Deep Searching, Yo.

2.1 DFS Basics

When you apply DFS to a graph, it implicitly defines a spanning tree rooted at the start vertex. If more than one tree is needed to span all vertices in the graph (when will this happen?), then we call it a *DFS forest*.

Q: How could we run only a single DFS on a disconnected graph and still hit every vertex, with similar results to picking each component and running DFS in sequence?

A: Construct a dummy node and give it out edges to all of the nodes in the graph. Begin the DFS there.

Q: What are edges that correspond to edges in the DFS tree called?

A: Tree edges.

Q: What are edges that go from a vertex v to an ancestor u in the DFS tree called?

A: Back edges.

Q: What are edges that go from a vertex v to an descendant u in the DFS tree called?

A: Forward edges.

Q: What are the remaining edges called?

A: Cross edges, since they cross from one subtree of the DFS tree to another.

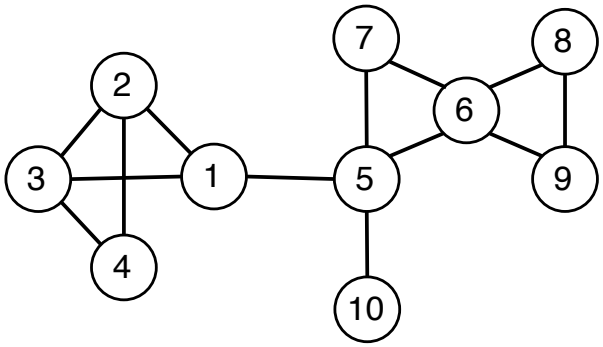
Q: In an *undirected* graph, if DFS finds a vertex that it has visited before, it has found a cycle. Is this sufficient?

A: There is a cycle in the undirected graph iff there is a back edge.

In an undirected graph, such as the one we will consider now, all edges are either tree edges or back edges (equivalently, we could say *forward edges* instead of *back edges*, since there's no notion of direction). Why can't there be any cross edges?

2.2 Drawing DFS Trees

What's the DFS tree of this graph? Use the vertices in increasing order.



Redraw the resulting DFS tree in a more familiar tree-like format.

2.3 Bridges

In *AbridgedLab*, your task is to find all bridges in an undirected graph.

Q: What is a bridge?

A: A bridge is an edge whose removal disconnects the graph. In other words, a bridge is not on a cycle.

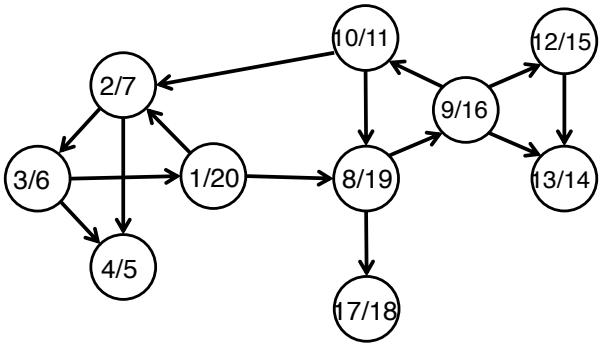
Q: Which edges are bridges in the above graph?

A: $(1, 5)$ and $(5, 10)$.

2.4 DFS Numberings

DFS can easily be modified to record the *discovery time* $d(u)$ and *finishing time* $f(u)$ for every vertex u. The discovery time corresponds to the time when a vertex is first visited and the finishing time corresponds to the time when the vertex is last visited.

Here are the DFS numbers for an example directed graph:



The following is a theorem that relates DFS numberings to the structure of the DFS tree. Fill in the following 3 cases for the relationships between d and f numbers, for any two vertices u and v:

1. If the intervals $d(u) \rightarrow f(u)$ and $d(v) \rightarrow f(v)$ are entirely disjoint, ...

2. If the interval $d(u) \rightarrow f(u)$ is contained entirely within the interval $d(v) \rightarrow f(v)$, ...

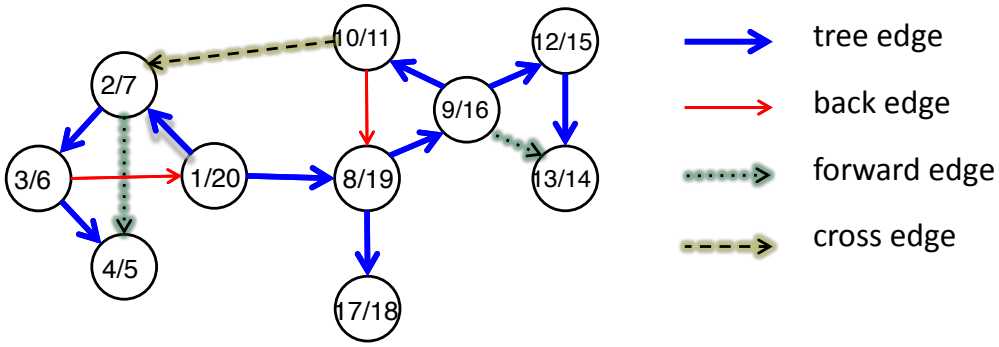
3. If the interval $d(v) \rightarrow f(v)$ is contained entirely within the interval $d(u) \rightarrow f(u)$, ...

Q: Based on this theorem, how would you use the discovery times and finishing times to classify edges into tree, forward, back or cross edges?

A: An edge (u, v) is:

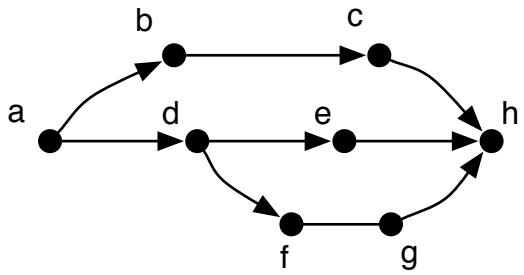
1. A tree/forward edge iff $d(u) < d(v) < f(v) < f(u)$
2. A back edge iff $d(v) < d(u) < f(u) < f(v)$
3. A cross edge iff $d(v) < f(v) < d(u) < f(u)$

Now classify the edges of the above graph into tree, forward, back or cross edges!



3 Topologicalizationalistically Sort This Graph

Let's do a topological sort on the following DAG:



4 So You Think You Know Your Way Around CMU?

4.1 Dijkstra's Basics

It's just after 15–210 lecture on Wednesday, and you need to get from Rashid to your next class in Baker. There are many ways to get there! You could:

- Take the Pausch Bridge to the Cut, walk straight across to the Doherty entrance, then diagonally across to Baker.
- Take the indoor bridge to Newell-Simon then to Wean, then walk across the Mall to Baker.
- Take the exit to Forbes, walk to Craig, get a delicious serving of *froyo* from Razzy, back to Gates, then option 1 to Baker!

Of course, you want the fastest route!

For Dijkstra's algorithm, we need to maintain a table $D(v)$ which contains the shortest path from the start node s (GHC in this example) to v *following only nodes that we've already explored*. We start out with $D(s) = 0$ and $D(v) = \infty$ for all $v \neq s$. Repeat the following steps until all nodes are explored:

1. Find a vertex u such that $D(u) = \min_{v \in N \setminus X} D(v)$, where X is the set of visited vertices.

2. For all $v \in N_G(u)$, set $D(v) = \min\{D(v),\ D(u) + w(u,v)\}$, where $w(u,v)$ is the weight of the edge from u to v.

Q: What data structure do we use to find the u with minimum $D(u)$?

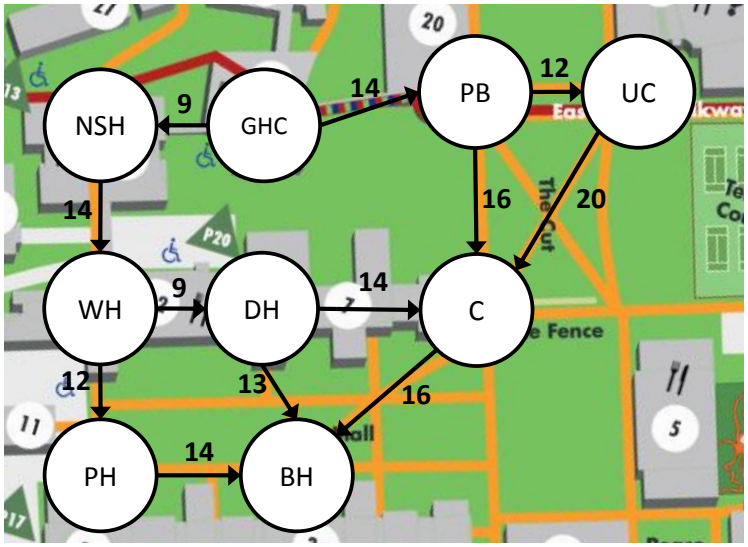
A: A priority queue! Using the `PQ.deleteMin` operation:

```
val deleteMin : 'a pq -> (key * 'a) option * 'a pq
```

Q: To update $D(v)$ for neighbors v of u, we just insert $(v,\ d + w(u,v))$ into the priority queue. Why do we not need to check if this is less than the existing value for v?

A: Because the first time we visit v, if there are multiple pairs $(v,\ d_1), \ldots, (v,\ d_n)$ in the priority queue, `deleteMin` will always give us the one with the lowest d_i. Thus, if we insert an existing node with a larger distance, it'll get ignored, and if we insert it with a smaller distance, it will just overwrite the existing value!

4.2 Finding our way to Baker!



Let's work through what Dijkstra's algorithm does on the graph of CMU above, with edge weights given by arbitrarily-scaled straight-line distances. For simplicity, we don't visit already-visited nodes. The first row has been done for you.

	Node	PQ Operations (after `deleteMin`)	PQ State (*lowest values only*)
1	GHC	`insert(PB, 14)`, `insert(NSH, 9)`	$\{NSH \mapsto 9,\ PB \mapsto 14\}$
2	NSH		
3	PB		
4	WH		
5	UC		
6	C		
7	DH		
8	PH		
9	BH		

In steps 5, 7 and 8 we inserted values for C, C, and BH, respectively, which were greater than the existing value, so these inserts are ignored. In step 7, we inserted a value for BH which was smaller than the existing value, thus updating the shortest path.

Hey guess what? This is indeed the shortest path to Baker: the path of distance 45 consisting of $\langle \text{GHC}, \text{NSH}, \text{WH}, \text{DH}, \text{BH} \rangle$. *DISCLAIMER: Walking directions are in BETA. Use caution: this route may contain stairs. The 15–210 course staff assumes no responsibility if following these directions makes you late to class!*

4.3 A*

You may now be left with a rather sour view of Dijkstra's Algorithm. You would never go to the UC trying to find a shortest route from GHC to Baker, and yet the algorithm did just that. But why? Think about what intuition you're using that Dijkstra doesn't have.

Here's where A* saves the day.

Q: How is A* different from Dijkstra's?

A: Instead of examining just the distance $d(v)$ of a vertex v, it considers $d(v) + h(v)$, where h is a *heuristic function*. More about this in the *AbridgedLab* handout.

Q: What are the practical applications of A*?

A: Glad you asked! You're welcome :)

http://www.youtube.com/watch?v=DlkMs4ZHHr8

4.4 Bonus - Strongly Connected Components

Recall the definition of strongly connected components from lecture: for a directed graph $G = (V, E)$, a set of vertices $V' \subset V$ is a *strongly connected component* iff $\forall u, v \in V'$ there exists a path from u to v, and a path from v to u, and $\forall w \in V \backslash V', \forall u \in V'$, either there is no path from u to w, or there is no path from w to u.

Trivially, every node of G is in some strongly connected component, because a single node can be a strongly connected component itself.

We'd like to identify the strongly connected components of G quickly. Determine an $O(m \lg n)$ algorithm (assuming an adjacency list representation of the graph) to do this.

Hint: Consider what information you can get from DFS, and try a topological sort.