Main topics for #lecture include #single_source_shortest_paths and #all-pairs_shortest_paths.

Directed Acyclic Graphs: Work in Topological Order

- We’re not quite done with single source shortest paths! But the algorithm I saved for last provides a nice transition to our next topic.
- Suppose we’re given a directed acyclic graph with arbitrary edge weights.
- We don’t have to worry about negative cycles at all with this case, because there aren’t any cycles at all in a DAG!
- Surprisingly, the key to solving single source shortest paths in this case is to consider dynamic programming. It’s back!
- Let dist(v) be the actual shortest path distance from s. If v = s, then dist(v) = 0. Otherwise, the shortest path contains some last edge u -> v. Therefore,
  - \( \text{dist}(v) = \begin{cases} 
  0 & \text{if } v = s \\
  \min_{u \rightarrow v} (\text{dist}(u) + w(u \rightarrow v)) & \text{otherwise} 
\end{cases} \)
- This identity is true for all directed graphs, but we can’t base an algorithm on it if there’s a directed cycle. We’d just keep recursing backwards forever.
- But since we’re working with a DAG, each recursive call visits an earlier vertex in topological order.
- So, we should just compute each dist value in topological order! Here’s the normal looking dynamic programming algorithm based directly off the recurrence.

```
DAGSSSP(s):
    for all vertices v in topological order
        if v = s
            dist(v) ← 0
        else
            dist(v) ← ∞
            for all edges u→v
                if dist(v) > dist(u) + w(u→v)  \(\{\text{if } u\rightarrow v \text{ is tense}\}\)
                    dist(v) ← dist(u) + w(u→v)  \(\{\text{relax } u\rightarrow v}\)  
```

- And here’s an algorithm that uses our relax procedure to actual compute the pred values instead.
Here’s an example run of the algorithm.

In both variants of this algorithm, we end up looking at every vertex and edge exactly once after computing the topological order in $O(V + E)$ time, so the algorithm runs in $O(V + E)$ time total.

Finally, you might not like the fact that we’re looking at incoming edges of a vertex when we’d normally consider outgoing edges in any of our algorithms. In particular, breadth-first search and Dijkstra’s algorithm consider outgoing edges of a vertex $u$ whenever we compute the true distance to $u$. We can do the same here:

By the time we process a vertex $v$, we have already checked every edge $u \rightarrow v$ and so $\text{dist}(v)$ will be correct.
All-Pairs Shortest Paths

- So that’s single source shortest paths. But what if you want to know the shortest paths from every vertex.
- This is the all-pairs shortest path problem. We want to compute dist(u, v), the length of the shortest path from u to v for all u and v.
- To keep things simple today, we’ll assume there are no negative cycles, but there could be negative weight edges.
- We can also compute values pred(u, v), the vertex immediately before v on the shortest path from u to v. But by now you hopefully see that actually computing these pred values requires only tiny changes over computing the dist values.
- So, there’s an obvious algorithm for this problem. Compute single source shortest paths from every vertex!

\[
\text{ObviousAPSP}(V, E, w):
\begin{align*}
\text{for every vertex } s \\
\text{dist}(s, \cdot) &\leftarrow \text{SSSP}(V, E, w, s)
\end{align*}
\]

- But depending on what kind of graph you have and what algorithm you call, this approach may lead to some pretty bad running times:
  - Unweighted: Use breadth-first search in \(V \times O(E) = O(VE) = O(V^3)\) time.
  - A DAG: Use the DAG algorithm we just discussed in \(V \times O(E) = O(VE) = O(V^3)\) time.
  - Non-negative edge weights: Use Dijkstra in \(O(V E \log V) = O(V^3 \log V)\) time. If we use Fibonacci heaps, that goes down to \(O((E + V \log V)) = O(V^3)\) time. We can even afford to manually scan the array of distance each time we want to extract from our queue for \(O((E + V^2)) = O(V^3)\) time total.
  - Otherwise: Use Bellman-Ford in \(V \times O(VE) = O(V^2 E) = O(V^4)\) time.
- The last case was much worse. Can we get that \(O(V^3)\) with negative length edges?
- I’ll show you one way to do it based on dynamic programming.
- One “obvious” recursive definition of dist(u, v) is the following. Just like before.

\[
\text{dist}(u, v) = \begin{cases} 
0 & \text{if } u = v \\
\min_{x \rightarrow v} \left( \text{dist}(u, x) + w(x \rightarrow v) \right) & \text{otherwise}
\end{cases}
\]

- But if we’re not in a DAG, the recursive calls suggested by this recursive definition may keep going back and back and back around a directed cycle. We’re stuck in an infinite loop!
- We need something that gets smaller in each recursive call so we know the recursion bottoms out.
- Earlier, we analyzed Bellman-Ford by considering shortest paths with at most i edges. Let’s take inspiration from this analysis by putting the number of edges into our recursively defined function.
• Let \( \text{dist}(u, v, \ell) \) denote the length of the shortest path from \( u \) to \( v \) that uses at most \( \ell \) edges.

• Now we have the following recursive function:

\[
\text{dist}(u, v, \ell) = \begin{cases} 
0 & \text{if } \ell = 0 \text{ and } u = v \\
\infty & \text{if } \ell = 0 \text{ and } u \neq v \\
\min \left\{ \text{dist}(u, v, \ell - 1), \min_{x \rightarrow v} \left( \text{dist}(u, x, \ell - 1) + w(x \rightarrow v) \right) \right\} & \text{otherwise}
\end{cases}
\]

• If there are no negative length cycles, then every shortest path uses at most \( V - 1 \) edges, so \( \text{dist}(u, v) = \text{dist}(u, v, V - 1) \).

• We can memoize this recurrence by looping over all \( \ell \) from 0 to \( V - 1 \) and then over all \( u \). We’d consider the different \( v \) in topological order.

• This strategy was described by Alfonso Shimbel in 1943, so Erickson calls it ShimbelAPSP.

<table>
<thead>
<tr>
<th>ShimbelAPSP((V, E, w)):</th>
</tr>
</thead>
<tbody>
<tr>
<td>for all vertices ( u )</td>
</tr>
<tr>
<td>for all vertices ( v )</td>
</tr>
<tr>
<td>if ( u = v )</td>
</tr>
<tr>
<td>( \text{dist}[u, v, 0] \leftarrow 0 )</td>
</tr>
<tr>
<td>else</td>
</tr>
<tr>
<td>( \text{dist}[u, v, 0] \leftarrow \infty )</td>
</tr>
<tr>
<td>for ( \ell \leftarrow 1 ) to ( V - 1 )</td>
</tr>
<tr>
<td>for all vertices ( u )</td>
</tr>
<tr>
<td>for all vertices ( v \neq u )</td>
</tr>
<tr>
<td>( \text{dist}[u, v, \ell] \leftarrow \text{dist}[u, v, \ell - 1] )</td>
</tr>
<tr>
<td>for all edges ( x \rightarrow v )</td>
</tr>
<tr>
<td>if ( \text{dist}[u, v, \ell] &gt; \text{dist}[u, x, \ell - 1] + w(x \rightarrow v) )</td>
</tr>
<tr>
<td>( \text{dist}[u, v, \ell] \leftarrow \text{dist}[u, x, \ell - 1] + w(x \rightarrow v) )</td>
</tr>
</tbody>
</table>

• Each edge is considered once per value of \( \ell \) and \( u \), so the whole thing takes \( O(V^2 E) = O(V^4) \) time. Oh, it didn’t get any better.

• If you look at the recursive definition, the variable \( u \) doesn’t change. We’re really just computing shortest paths from each vertex \( u \) to all vertices \( v \) separately. In fact, computing the shortest paths using at most \( \ell \) edges from \( u \) is essentially another way to describe Bellman-Ford with \( u \) as the source. Can we do better?

**Floyd-Warshall**

• It turns out we need a better third variable for our recursive definition.

• There’s a different dynamic programming formulation discovered (in some form) by many people independently as usual.

• Instead of tracking how many edges appear in a path, we’ll instead track which vertices are allowed to appear in the path.
• Number the vertices arbitrarily from 1 to V.

• \( \pi(u, v, r) := \text{the shortest path from } u \text{ to } v \text{ where every intermediate vertex (that is, every vertex except } u \text{ and } v \text{) is numbered at most } r. \)

• In particular, \( \pi(u, v, V) \) is the true shortest path from \( u \) to \( v \) since it can use any of the vertices.

• If \( r = 0 \), we can’t have any intermediate vertices. So either \( \pi(u, v, 0) = u \rightarrow v \) or it’s not defined.

• Now, either \( \pi(u, v, r) \) uses intermediate vertex \( r \) or it doesn’t.

  ![Diagram](image)

• If it doesn’t, then \( \pi(u, v, r) = \pi(u, v, r - 1). \)

• If it does, then it contains a subpath from \( u \) to \( r \) and a subpath from \( r \) to \( v \) so \( \pi(u, v, r) = \pi(u, r, r - 1) \) followed by \( \pi(r, v, r - 1). \)

• So now, let \( \text{dist}(u, v, r) \) be the length of \( \pi(u, v, r). \)

\[
\text{dist}(u, v, r) = \begin{cases} 
  w(u \rightarrow v) & \text{if } r = 0 \\
  \min \left\{ \begin{array}{c}
  \text{dist}(u, v, r - 1) \\
  \text{dist}(u, r, r - 1) + \text{dist}(r, v, r - 1)
  \end{array} \right\} & \text{otherwise}
\end{cases}
\]

• Again, the shortest path from \( u \) to \( v \) may use any vertex, so \( \text{dist}(u, v) = \text{dist}(u, v, V). \)

• We need to compute \( V \times V \times V = \Theta(V^3) \) values, but it takes only constant time for each. So the algorithm will take \( \Theta(V^3) \) time.

• Jeff calls this algorithm KleeneAPSP (clay knee), because Kleene discovered this recursive pattern while studying finite automata.

```plaintext
KLEENEAPSP(V, E, w):
    for all vertices u
        for all vertices v
            \( \text{dist}[u, v, 0] \leftarrow w(u \rightarrow v) \)
    for \( r \leftarrow 1 \) to \( V \)
        for all vertices u
            for all vertices v
                if \( \text{dist}[u, v, r - 1] < \text{dist}[u, r, r - 1] + \text{dist}[r, v, r - 1] \)
                    \( \text{dist}[u, v, r] \leftarrow \text{dist}[u, v, r - 1] \)
                else
                    \( \text{dist}[u, v, r] \leftarrow \text{dist}[u, r, r - 1] + \text{dist}[r, v, r - 1] \)
```

• We can clean this algorithm up a bit. We only need to maintain the shortest paths from each \( u \) to \( v \) we’ve found so far, not which specific vertices they were allowed to go through.

• We also don’t need to keep track of the specific vertex numbers as long as we loop through all the vertices.
This version still runs in $O(V^3)$ time.

This is the cleaned up version of that dynamic programming algorithm. It’s usually referred to as Floyd-Warshall. A formal proof of correctness involves a similar induction proof to the one we used for Bellman-Ford. Namely, we can prove by induction that after the $r$th iteration of the outer loop, $\text{dist}(u, v)$ is at most $\text{dist}(u, v, r)$. I’ll spare you the details.