COMP90038 Algorithms and Complexity

Lecture 20: Greedy Techniques – Prim and Dijkstra (with thanks to Harald Søndergaard & Michael Kirley)

Andres Munoz-Acosta

munoz.m@unimelb.edu.au

Peter Hall Building G.83

On the previous lecture

- We have talked a lot about dynamic programming:
 - DP is bottom-up problem solving technique
 - Similar to divide-and-conquer; however, problems are overlapping
 - Solutions often involve recursion
- We applied this idea to two graph problems:
 - Computing the **transitive closure** of a directed graph
 - Finding shortest distances in weighted directed graphs

Greedy algorithms

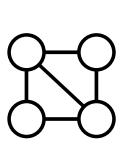
- A problem solving strategy is to take the locally best choice among all feasible ones
 - Once we do this, our decision is irrevocable
- We want to change 30 cents using the smallest number of coins
 - Assuming denominations of $\{25, 10, 5, 1\}$, we could use as many 25-cent pieces as we can, then do the same for 10-cent pieces, and so on, until we have reached 30 cents (25+5)
 - This **greedy** strategy would not work for denominations $\{25, 10, 1\}$ (25+1+1+1+1+1 compared to 10+10+10)

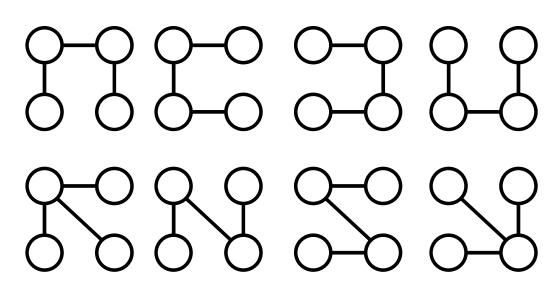
Greedy algorithms

- In general, it is unusual that locally best choices yield global best results
 - However, there are problems for which a greedy algorithm is correct and fast
 - In some other problems, a greedy algorithm is an acceptable **approximation** algorithm
- Here we shall look at:
 - Prim's algorithm for finding minimum spanning trees
 - Dijkstra's algorithm for single-source shortest paths

What is a spanning tree?

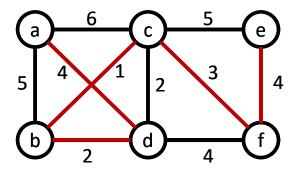
- Recall that a tree is a connected graph with no cycles
- A **spanning tree** of a graph $\langle V,E\rangle$ is a tree $\langle V,E'\rangle$ where E is a subset of E
- For example, the graph on the left has eight different spanning trees:





Minimum spanning trees of weighted graphs

- For a weighted graph, some spanning trees are more desirable than others
 - For example, suppose we have a set of "stations" to connect in a network, and also some possible connections, each with its own **cost**
- This is the problem of finding a spanning tree with the smallest possible cost
 - Such tree is a **minimum spanning tree** for the graph



- Prim's algorithm is an example of a greedy algorithm
 - It constructs a sequence of subtrees T, by adding to the latest tree the closest node not currently on it
- A simple version:

```
function PRIM(\langle V, E \rangle)
V_T \leftarrow \{v_0\}
E_T \leftarrow \emptyset
for i \leftarrow 1 to |V| - 1 do
find a minimum-weight edge (v, u) \in V_T \times (V \setminus V_T)
V_T \leftarrow V_T \cup \{u\}
E_T \leftarrow E_T \cup \{(v, u)\}
return E_T
```

- But how to find the **minimum-weight edge** (v,u)?
- An approach is to organise the nodes that are not yet included in the spanning tree T as a **priority queue**, using a **min-heap** by edge **cost**
- Which nodes are connected in T is stored by an array prev of nodes, indexed by V. Namely, when (v,u) is included, this is stored by setting prev[u] = v

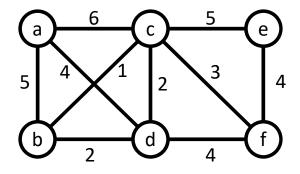
The complete algorithm is:

```
function PRIM(\langle V, E \rangle)
    for each v \in V do
        cost[v] \leftarrow \infty
        prev[v] \leftarrow nil
    pick initial node v_0
    cost[v_0] \leftarrow 0
    Q \leftarrow \texttt{InitPriorityQueue}(V)
                                                               > priorities are cost values
    while Q is non-empty do
        u \leftarrow \text{EjectMin}(Q)
        for each (u, w) \in E do
            if weight(u, w) < cost[w] then
                 cost[w] \leftarrow weight(u, w)
                 prev[w] \leftarrow u
                 UPDATE(Q, w, cost[w])
                                                              > rearranges priority queue
```

• On the first loop, we only create the table

```
function PRIM(\langle V, E \rangle)
for each v \in V do
cost[v] \leftarrow \infty
prev[v] \leftarrow nil

pick initial node v_0
cost[v_0] \leftarrow 0
Q \leftarrow INITPRIORITYQUEUE(V)
while Q is non-empty do
u \leftarrow EJECTMIN(Q)
for each (u, w) \in E do
if weight(u, w) < cost[w] then
cost[w] \leftarrow weight(u, w)
prev[w] \leftarrow u
UPDATE(Q, w, cost[w])
```



	a	b	С	d	е	f
cost	∞	∞	∞	∞	∞	∞
prev	nil	nil	nil	nil	nil	nil
		cost ∞	cost ∞ ∞	$\cos t \propto \infty$	$cost \infty \infty \infty \infty$	$cost \infty \infty \infty \infty \infty$

• Then we pick the first node as the initial one

```
function PRIM(\langle V, E \rangle)

for each v \in V do

cost[v] \leftarrow \infty

prev[v] \leftarrow nil

pick initial node v_0

cost[v_0] \leftarrow 0

Q \leftarrow INITPRIORITYQUEUE(V)

while Q is non-empty do

u \leftarrow EJECTMIN(Q)

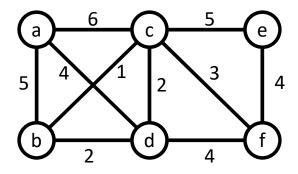
for each (u, w) \in E do

if weight(u, w) < cost[w] then

cost[w] \leftarrow weight(u, w)

prev[w] \leftarrow u

UPDATE(Q, w, cost[w])
```



Tree T		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil

 We take the first node out of the queue and update the costs

```
function PRIM(\langle V, E \rangle)

for each v \in V do

cost[v] \leftarrow \infty

prev[v] \leftarrow nil

pick initial node v_0

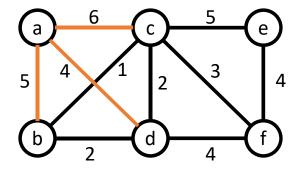
cost[v_0] \leftarrow 0

Q \leftarrow INITPRIORITYQUEUE(V)

while Q is non-empty do

u \leftarrow EJECTMIN(Q)
```

```
u \leftarrow \text{EJECTMIN}(Q)
\textbf{for } \text{each } (u, w) \in E \textbf{ do}
\textbf{if } weight(u, w) < cost[w] \textbf{ then}
cost[w] \leftarrow weight(u, w)
prev[w] \leftarrow u
UPDATE(Q, w, cost[w])
```



Tree T		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost		5	6	4	∞	∞
a	prev		a	a	a	nil	nil

 We eject the node with the lowest cost and update the queue

```
function PRIM(\langle V, E \rangle)

for each v \in V do

cost[v] \leftarrow \infty

prev[v] \leftarrow nil

pick initial node v_0

cost[v_0] \leftarrow 0

Q \leftarrow INITPRIORITYQUEUE(V)

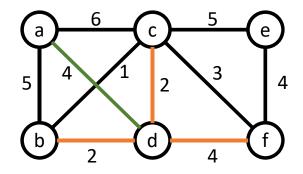
while Q is non-empty do

u \leftarrow EJECTMIN(Q)
for each (u, w) \in E do

if weight(u, w) < cost[w] then

cost[w] \leftarrow weight(u, w)
prev[w] \leftarrow u
```

UPDATE(Q, w, cost[w])



Tree T		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
a	cost		5	6	4	∞	∞
a	prev		a	a	a	nil	nil
a,d	cost		2	2		∞	4
a, a	prev		d	d		nil	d

We eject the next node based on alphabetical order.
 Why is (f) not updated?

```
function PRIM(\langle V, E \rangle)

for each v \in V do

cost[v] \leftarrow \infty

prev[v] \leftarrow nil

pick initial node v_0

cost[v_0] \leftarrow 0

Q \leftarrow INITPRIORITYQUEUE(V)

while Q is non-empty do

u \leftarrow EJECTMIN(Q)
```

```
u \leftarrow \text{EJECTMIN}(Q)

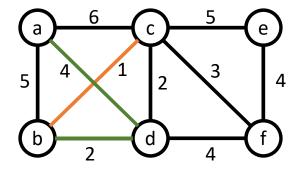
for each (u, w) \in E do

if weight(u, w) < cost[w] then

cost[w] \leftarrow weight(u, w)

prev[w] \leftarrow u

UPDATE(Q, w, cost[w])
```



Tree T		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	\propto
	prev	nil	nil	nil	nil	nil	ni
	cost	0	∞	∞	∞	∞	α
	prev	nil	nil	nil	nil	nil	ni
2	cost		5	6	4	∞	α
а	prev		a	a	a	nil	ni
a,d	cost		2	2		∞	4
a, a	prev		d	d		nil	C
a,d,b	cost			1		∞	2
α, α, υ	prev			b		nil	C

• We now update (f)

```
function PRIM(\langle V, E \rangle)

for each v \in V do

cost[v] \leftarrow \infty

prev[v] \leftarrow nil

pick initial node v_0

cost[v_0] \leftarrow 0

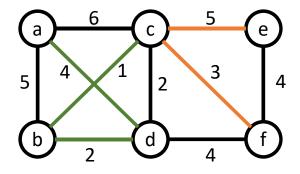
Q \leftarrow INITPRIORITYQUEUE(V)

while Q is non-empty do

u \leftarrow EJECTMIN(Q)
for each (u, w) \in E do

if weight(u, w) < cost[w] then

cost[w] \leftarrow weight(u, w)
prev[w] \leftarrow u
UPDATE(Q, w, cost[w])
```



Tree T		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
2	cost		5	6	4	∞	∞
a	prev		a	a	a	nil	nil
a d	cost		2	2		∞	4
a,d	prev		d	d		nil	d
a d h	cost			1		∞	4
a,d,b	prev			b		nil	d
adh a	cost					5	3
a,d,b,c	prev					C	C

We reach the last choice

```
function PRIM(\langle V, E \rangle)

for each v \in V do

cost[v] \leftarrow \infty

prev[v] \leftarrow nil

pick initial node v_0

cost[v_0] \leftarrow 0

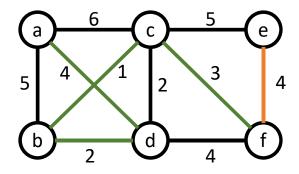
Q \leftarrow INITPRIORITYQUEUE(V)

while Q is non-empty do

u \leftarrow EJECTMIN(Q)
for each (u, w) \in E do

if weight(u, w) < cost[w] then

cost[w] \leftarrow weight(u, w)
prev[w] \leftarrow u
UPDATE(Q, w, cost[w])
```



Tree T		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
2	cost		5	6	4	∞	∞
a	prev		a	a	a	nil	nil
a,d	cost		2	2		∞	4
a, u	prev		d	d		nil	d
a d h	cost			1		∞	4
a,d,b	prev			b		nil	d
a,d,b,c	cost					5	3
a, u, D, C	prev					С	C
a,d,b,c,f	cost					4	
a, u, D, C, I	prev					f	

• The resulting tree is {a,d,b,c,f,e}

```
function PRIM(\langle V, E \rangle)

for each v \in V do

cost[v] \leftarrow \infty

prev[v] \leftarrow nil

pick initial node v_0

cost[v_0] \leftarrow 0

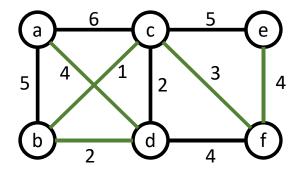
Q \leftarrow INITPRIORITYQUEUE(V)

while Q is non-empty do

u \leftarrow EJECTMIN(Q)
for each (u, w) \in E do

if weight(u, w) < cost[w] then

cost[w] \leftarrow weight(u, w)
prev[w] \leftarrow u
UPDATE(Q, w, cost[w])
```



Tree T		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
a	cost		5	6	4	∞	∞
a	prev		a	a	a	nil	nil
a,d	cost		2	2		∞	4
a, u	prev		d	d		nil	d
a,d,b	cost			1		∞	4
a, u, D	prev			b		nil	d
adh a	cost					5	3
a,d,b,c	prev					С	С
adhaf	cost					4	
a,d,b,c,f	prev					f	
adhaf o	cost						
a,d,b,c,f,e	prev						

Analysis of Prim's algorithm

- First, a crude analysis: For each node, we look through the edges to find those incident to the node, and pick the one with smallest cost. Thus we get $O(|V| \times |E|)$. However, we are using cleverer data structures.
- Using adjacency lists for the graph and a min-heap for the priority queue, we perform |V| 1 heap deletions (each at cost $O(\log |V|)$) and |E| updates of priorities (each at cost $O(\log |V|)$).
- Altogether $(|V|-1+|E|) O(\log |V|)$.
- Since, in a connected graph, $|V|-1 \le |E|$, this is $O(|E| \log |V|)$.

- Another classical greedy weighted-graph algorithm is Dijkstra's algorithm, whose overall structure is the same as Prim's
- On **Lecture 19** we talked about Floyd's algorithm:
 - It gave us the shortest paths, **for every pair of nodes**, in a (directed or undirected) weighted graph.
 - Assumes an adjacency matrix representation and had complexity $O(|V|^3)$
- **Dijkstra's algorithm** is also a shortest-path algorithm for (directed or undirected) weighted graphs
 - It finds all shortest paths from a fixed start node
 - Its complexity is the same as that of Prim's algorithm

The complete algorithm is:

```
function Dijkstra(\langle V, E \rangle, v_0)

for each v \in V do

dist[v] \leftarrow \infty
prev[v] \leftarrow nil
dist[v_0] \leftarrow 0
Q \leftarrow \text{InitPriorityQueue}(V)
vhile Q \text{ is non-empty do}
u \leftarrow \text{EjectMin}(Q)
for each (u, w) \in E do

if dist[u] + weight(u, w) < dist[w] then
dist[w] \leftarrow dist[u] + weight(u, w)
prev[w] \leftarrow u
UPDATE(Q, w, dist[w])
▷ rearranges priority queue
```

• On the first loop, we only create the table

```
function DIJKSTRA(\langle V, E \rangle, v_0)

for each v \in V do

dist[v] \leftarrow \infty

prev[v] \leftarrow nil

dist[v_0] \leftarrow 0

Q \leftarrow \text{INITPRIORITYQUEUE}(V)

while Q is non-empty do

u \leftarrow \text{EJECTMIN}(Q)

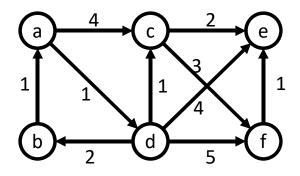
for each (u, w) \in E do

if dist[u] + weight(u, w) < dist[w] then

dist[w] \leftarrow dist[u] + weight(u, w)

prev[w] \leftarrow u

Q \leftarrow \text{UPDATE}(Q, w, dist[w])
```



Covered		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil

Then we pick the first node as the initial one

```
function Dijkstra(\langle V, E \rangle, v_0)

for each v \in V do

dist[v] \leftarrow \infty

prev[v] \leftarrow nil

dist[v_0] \leftarrow 0

Q \leftarrow \text{InitPriorityQueue}(V)

while Q is non-empty do

u \leftarrow \text{EjectMin}(Q)

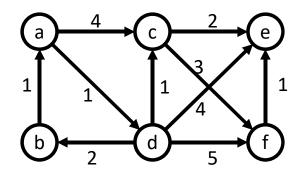
for each (u, w) \in E do

if dist[u] + weight(u, w) < dist[w] then

dist[w] \leftarrow dist[u] + weight(u, w)

prev[w] \leftarrow u

u \leftarrow \text{Update}(Q, w, dist[w])
```



Covered		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil

Then we pick the first node as the initial one

```
function DIJKSTRA(\langle V, E \rangle, v_0)

for each v \in V do

dist[v] \leftarrow \infty

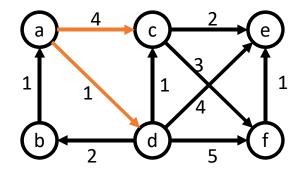
prev[v] \leftarrow nil

dist[v_0] \leftarrow 0

Q \leftarrow \text{INITPRIORITYQUEUE}(V)

while Q is non-empty do
```

```
while Q is non-empty do u \leftarrow \operatorname{EJECTMIN}(Q) for each (u,w) \in E do if \operatorname{dist}[u] + \operatorname{weight}(u,w) < \operatorname{dist}[w] then \operatorname{dist}[w] \leftarrow \operatorname{dist}[u] + \operatorname{weight}(u,w) \operatorname{prev}[w] \leftarrow u UPDATE(Q,w,\operatorname{dist}[w])
```



Covered		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
a	cost		∞	4	1	∞	∞
а	prev		nil	a	a	nil	nil

• Then eject the node with the shortest distance from the queue. Then, we update all the paths by adding 1.

```
function DIJKSTRA(\langle V, E \rangle, v_0)

for each v \in V do

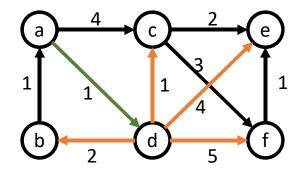
dist[v] \leftarrow \infty

prev[v] \leftarrow nil

dist[v_0] \leftarrow 0

Q \leftarrow \text{INITPRIORITYQUEUE}(V)
```

```
while Q is non-empty do u \leftarrow \operatorname{EJECTMIN}(Q) for each (u,w) \in E do if \operatorname{dist}[u] + \operatorname{weight}(u,w) < \operatorname{dist}[w] then \operatorname{dist}[w] \leftarrow \operatorname{dist}[u] + \operatorname{weight}(u,w) \operatorname{prev}[w] \leftarrow u UPDATE(Q,w,\operatorname{dist}[w])
```



Covered		a	b	С	d	e	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
2	cost		∞	4	1	∞	∞
а	prev		nil	a	а	nil	nil
a d	cost		3	2		5	6
a,d	prev		d	d		d	d

 Our next node will be the one with the shortest path in overall (b)

```
function DIJKSTRA(\langle V, E \rangle, v_0)

for each v \in V do

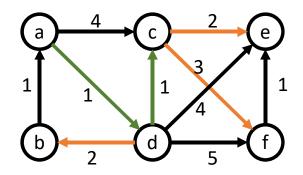
dist[v] \leftarrow \infty

prev[v] \leftarrow nil

dist[v_0] \leftarrow 0

Q \leftarrow \text{InitPriorityQueue}(V)
```

```
while Q is non-empty do u \leftarrow \operatorname{EJECTMIN}(Q) for each (u,w) \in E do if \operatorname{dist}[u] + \operatorname{weight}(u,w) < \operatorname{dist}[w] then \operatorname{dist}[w] \leftarrow \operatorname{dist}[u] + \operatorname{weight}(u,w) \operatorname{prev}[w] \leftarrow u UPDATE(Q,w,\operatorname{dist}[w])
```



Covered		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
a	cost		∞	4	1	∞	∞
	prev		nil	a	а	nil	nil
ا ما	cost		3	2		5	6
a,d	prev		d	d		d	d
a,d,c	cost		3			4	5
	prev		d			С	С

Now, we continue evaluating from (c)

```
function DIJKSTRA(\langle V, E \rangle, v_0)

for each v \in V do

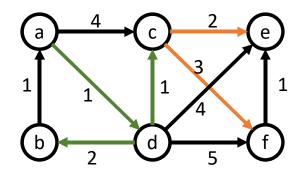
dist[v] \leftarrow \infty

prev[v] \leftarrow nil

dist[v_0] \leftarrow 0

Q \leftarrow \text{INITPRIORITYQUEUE}(V)
```

```
 \begin{aligned} \textbf{while } Q \text{ is non-empty } \textbf{do} \\ u &\leftarrow \text{EJECTMIN}(Q) \\ \textbf{for } \text{ each } (u,w) \in E \textbf{ do} \\ \textbf{if } dist[u] + weight(u,w) < dist[w] \textbf{ then} \\ dist[w] &\leftarrow dist[u] + weight(u,w) \\ prev[w] &\leftarrow u \\ \text{UPDATE}(Q,w,dist[w]) \end{aligned}
```



Covered		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
a	cost		∞	4	1	∞	∞
	prev		nil	a	а	nil	nil
1	cost		3	2		5	6
a,d	prev		d	d		d	d
2 4 6	cost		3			4	5
a,d,c	prev		d			С	С
a,d,c,b	cost					4	5
	prev					С	C

We arrive at our last decision.

```
function DIJKSTRA(\langle V, E \rangle, v_0)

for each v \in V do

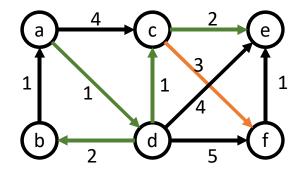
dist[v] \leftarrow \infty

prev[v] \leftarrow nil

dist[v_0] \leftarrow 0

Q \leftarrow \text{INITPRIORITYQUEUE}(V)

while Q is non-empty do
```



Covered		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
a	cost		∞	4	1	∞	∞
	prev		nil	a	а	nil	nil
2 4	cost		3	2		5	6
a,d	prev		d	d		d	d
a,d,c	cost		3			4	5
	prev		d			С	С
a,d,c,b	cost					4	5
	prev					С	С
a,d,c,b	cost						5
, e	prev						С

• Our complete tree is {a,d,c,b,e,f}

```
function DIJKSTRA(\langle V, E \rangle, v_0)

for each v \in V do

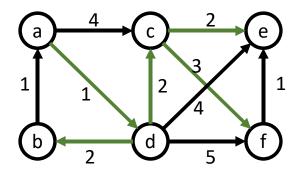
dist[v] \leftarrow \infty

prev[v] \leftarrow nil

dist[v_0] \leftarrow 0

Q \leftarrow \text{InitPriorityQueue}(V)
```

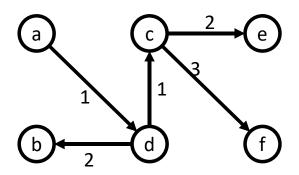
```
while Q is non-empty do u \leftarrow \text{EJECTMIN}(Q) for each (u,w) \in E do if dist[u] + weight(u,w) < dist[w] then dist[w] \leftarrow dist[u] + weight(u,w) prev[w] \leftarrow u UPDATE(Q, w, dist[w])
```



Covered		a	b	С	d	е	f
	cost	∞	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
	cost	0	∞	∞	∞	∞	∞
	prev	nil	nil	nil	nil	nil	nil
a	cost		∞	4	1	∞	∞
	prev		nil	a	а	nil	nil
a,d	cost		3	2		5	6
	prev		d	d		d	d
2 4 6	cost		3			4	5
a,d,c	prev		d			С	С
a,d,c,b	cost					4	5
	prev					С	С
a,d,c,b	cost						5
, e	prev						С
a,d,c,b	cost						
,e,f	prev						

Tracing paths

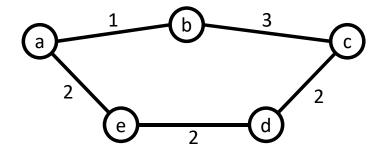
 The array prev is not really needed, unless we want to retrace the shortest paths from node a



• This tree is referred as the **shortest-path tree**

Spanning trees and shortest-path trees

• The shortest-path tree that results from Dijkstra's algorithm is very similar to the minimum spaning tree.



- Exercise:
 - Assume that you started from node a.
 - Which edge is missing in the minimal spanning tree?
 - Which edge is missing from the shortest-path tree?

Next lecture

• Huffman trees and codes (Levitin Section 9.4)