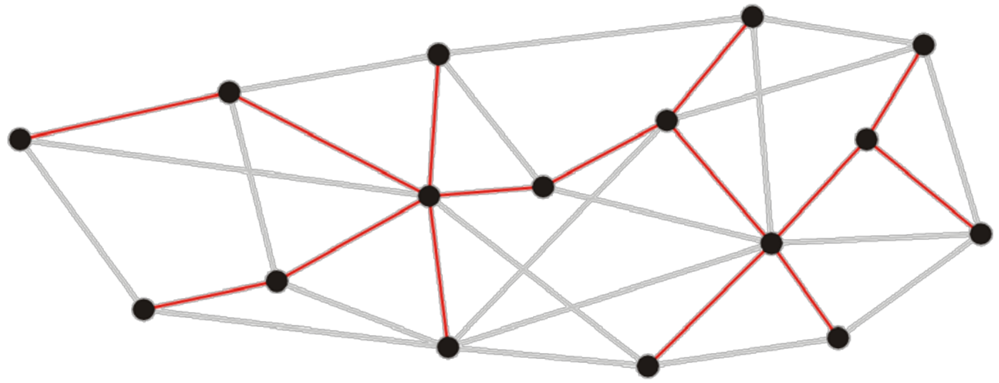
Discussion Week 9

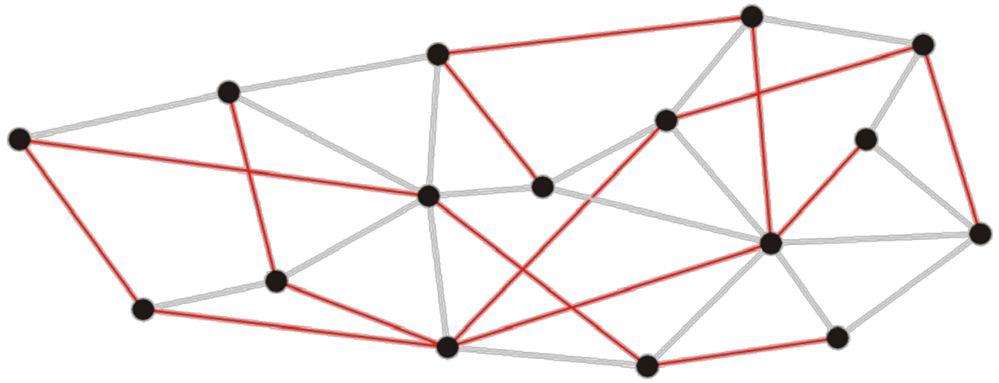
Minimum Spanning Tree

Spanning trees

Given a connected graph with n vertices, a spanning tree is defined as a subgraph that is a tree and includes all the n vertices

- It has $n - 1$ edges

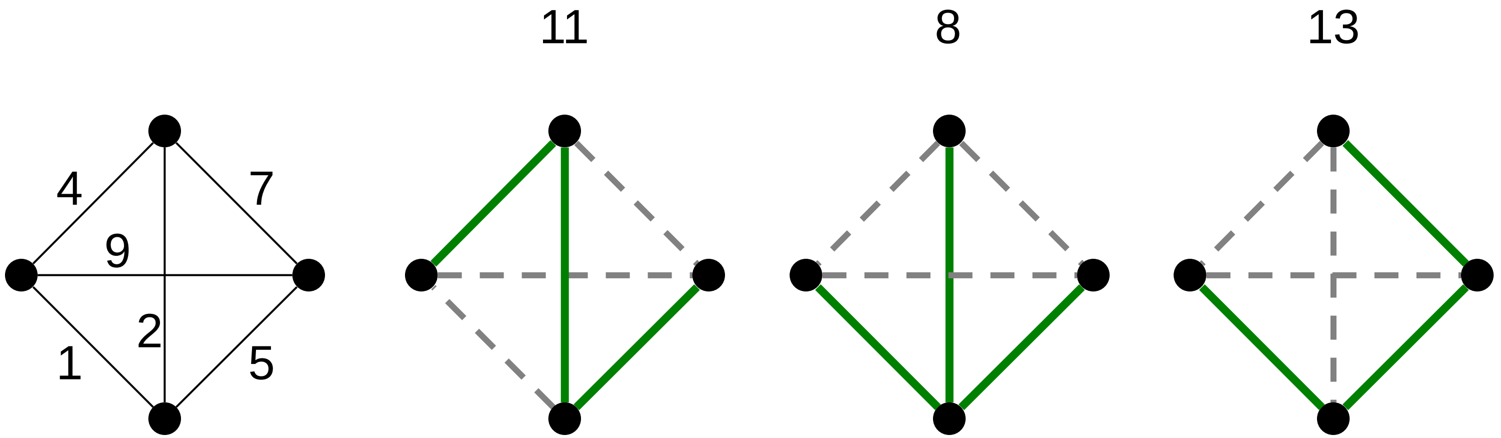




A spanning tree is not necessarily unique

Spanning trees on weighted graphs

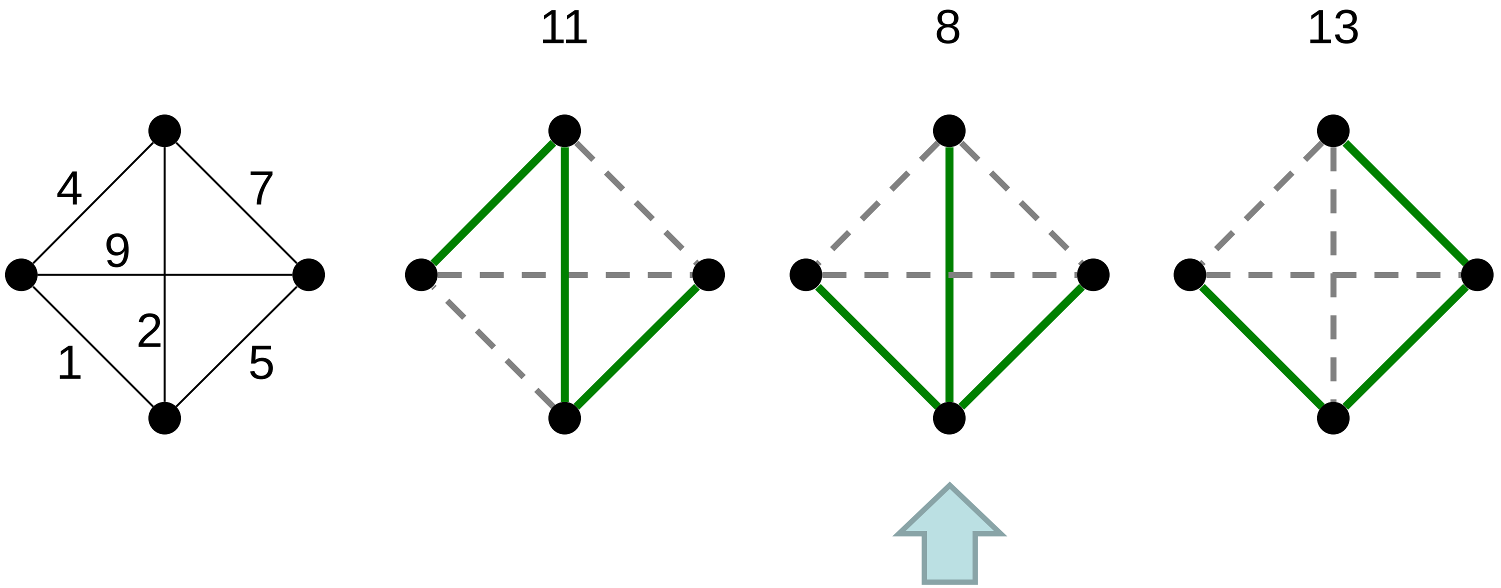
The weight of a spanning tree is the sum of the weights on all the edges which comprise the spanning tree



Minimum Spanning Trees

Which spanning tree minimizes the weight?

- Such a tree is termed a *minimum spanning tree*



Minimum Spanning Trees

Simplifying assumption:

- All edge weights are distinct

This guarantees that given a graph, there is a unique minimum spanning tree.

Prim's Algorithm

Prim's algorithm for finding the minimum spanning tree states:

- Start with an arbitrary vertex to form a minimum spanning tree on one vertex
- At each step, add the edge with least weight that connects the current minimum spanning tree to a new vertex
- Continue until we have $n - 1$ edges and n vertices

Prim’s Algorithm

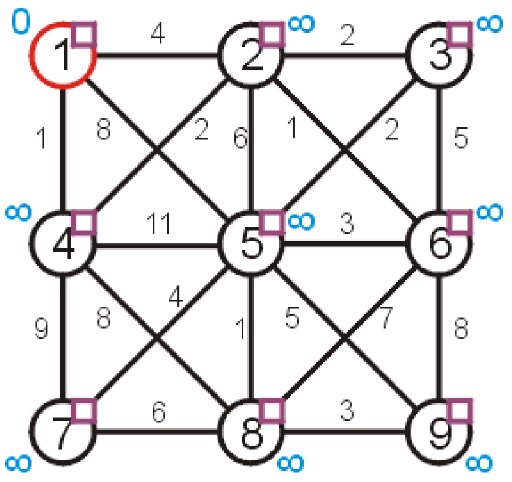
Associate with each vertex three items of data:

- A Boolean flag indicating if the vertex has been visited,
- The minimum distance (weight of a connecting edge) to the partially constructed tree, and
- A pointer to a vertex which will form the parent node in the resulting tree

Implementation:

- Add three member variables to the vertex class
- Or track three tables

Prim’s Algorithm

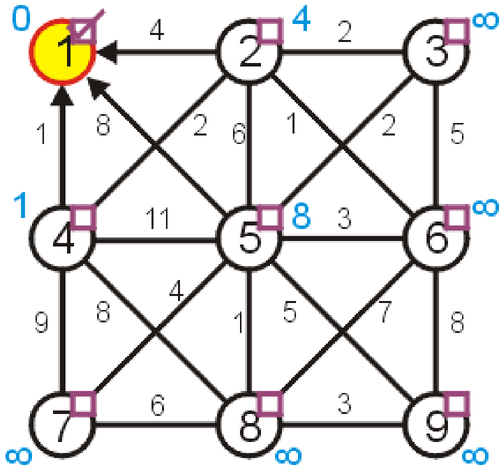


Visited or not

		Distance	Parent
1			
2			
3			
4			
5			
6			
7			
8			
9			

Prim’s Algorithm

Visiting vertex 1, we update vertices 2, 4, and 5

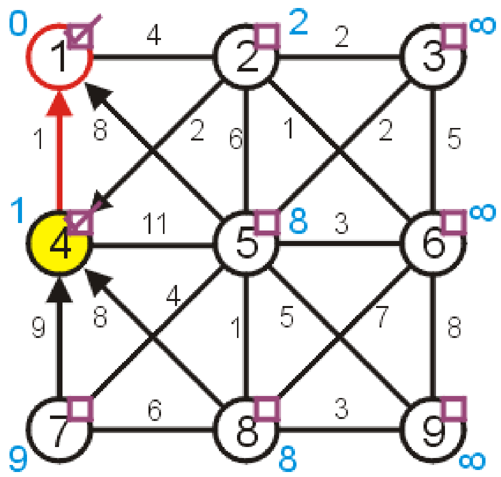


		Distance	Parent
1	T	0	0
2	F	4	1
3	F	∞	0
4	F	1	1
5	F	8	1
6	F	∞	0
7	F	∞	0
8	F	∞	0
9	F	∞	0

Prim's Algorithm

The next unvisited vertex with minimum distance is vertex 4

- Update vertices 2, 7, 8
- Don't update vertex 5

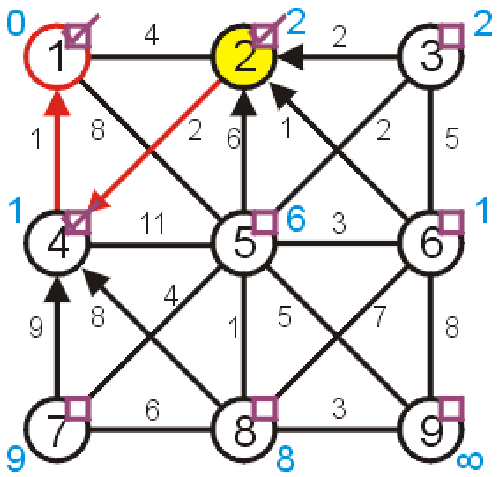


		Distance	Parent
1	T	0	0
2	F	2	4
3	F	∞	0
4	T	1	1
5	F	8	1
6	F	∞	0
7	F	9	4
8	F	8	4
9	F	∞	0

Prim’s Algorithm

Next visit vertex 2

- Update 3, 5, and 6

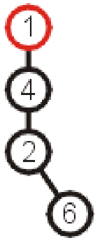


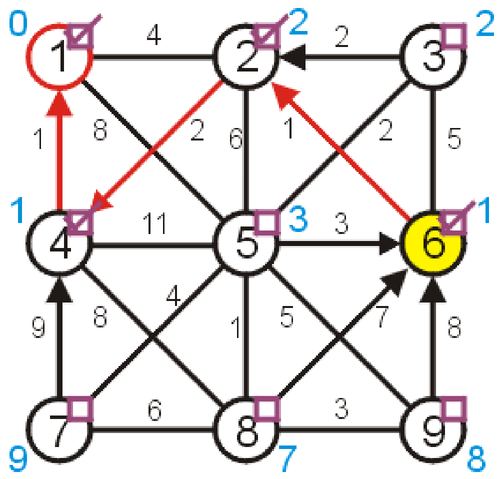
		Distance	Parent
1	T	0	0
2	T	2	4
3	F	2	2
4	T	1	1
5	F	6	2
6	F	1	2
7	F	9	4
8	F	8	4
9	F	∞	0

Prim's Algorithm

Next, we visit vertex 6:

- update vertices 5, 8, and 9

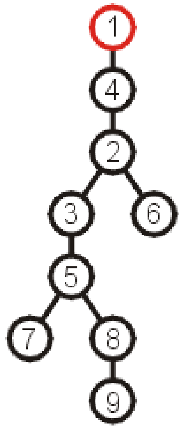


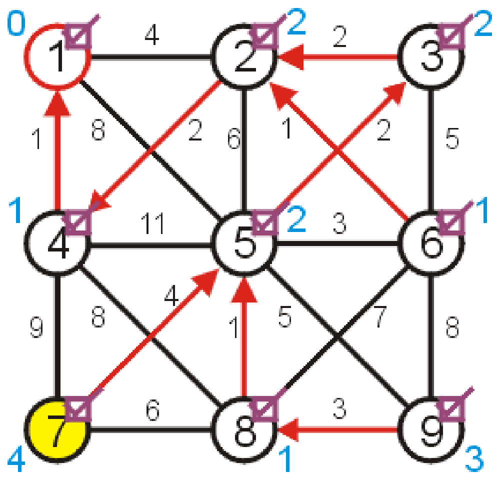


		Distance	Parent
1	T	0	0
2	T	2	4
3	F	2	2
4	T	1	1
5	F	3	6
6	T	1	2
7	F	9	4
8	F	7	6
9	F	8	6

Prim's Algorithm

And neither are there any vertices to update when visiting vertex 7

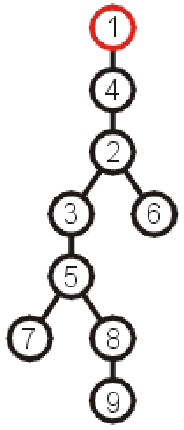


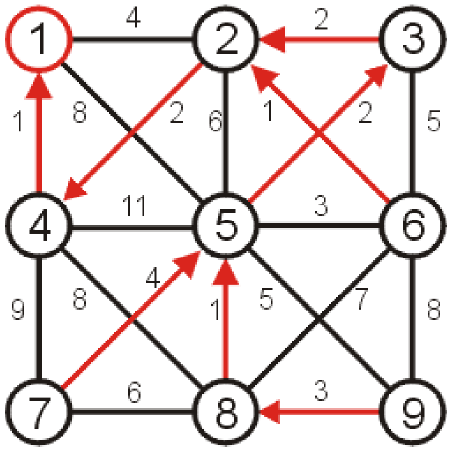


		Distance	Parent
1	T	0	0
2	T	2	4
3	T	2	2
4	T	1	1
5	T	2	3
6	T	1	2
7	T	4	5
8	T	1	5
9	T	3	8

Prim's Algorithm

Using the parent pointers, we can now construct the minimum spanning tree





		Distance	Parent
1	T	0	0
2	T	2	4
3	T	2	2
4	T	1	1
5	T	2	3
6	T	1	2
7	T	4	5
8	T	1	5
9	T	3	8

Implementation: Prim's Algorithm

Use a priority queue.

- Maintain set of explored nodes S.
- For each unexplored node v, maintain attachment cost a[v] = cost of cheapest edge from v to a node in S.
- O(n2) with an array; O(m log n) with a binary heap.

```
Prim(G, c) {
   foreach (v ∈ V) a[v] ← ∞
   Initialize an empty priority queue Q
   foreach (v ∈ V) insert v onto Q
   Initialize set of explored nodes S ← φ

   while (Q is not empty) {
      u ← delete min element from Q
      S ← S ∪ { u }
      foreach (edge e = (u, v) incident to u)
         if ((v ∉ S) and (c_e < a[v]))
            decrease priority a[v] to c_e
}
```

Kruskal's Algorithm

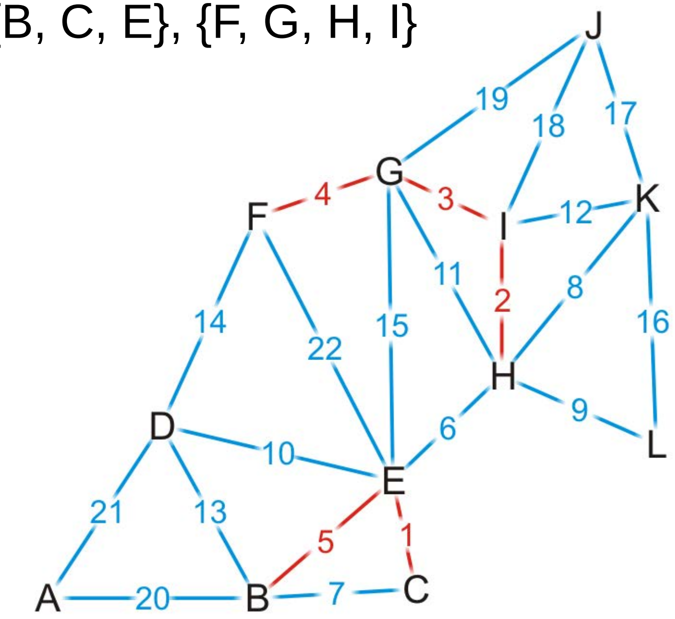
- Sort the edges by weight
- Go through the edges from least weight to greatest weight
  - add the edges to the spanning tree so long as the addition does not create a cycle
  - Does this edge belong to the minimum spanning tree?
    - Yes! The cut property (consider the subtree connected to one end of the edge as the set *S*).
- Repeatedly add more edges until:
  - $|V| - 1$ edges have been added, then we have a minimum spanning tree
  - Otherwise, if we have gone through all the edges, then we have a forest of minimum spanning trees on all connected sub-graphs

Analysis

We could use disjoint sets

- Consider edges in the same connected sub-graph as forming a set

{B, C, E}, {F, G, H, I}



Analysis

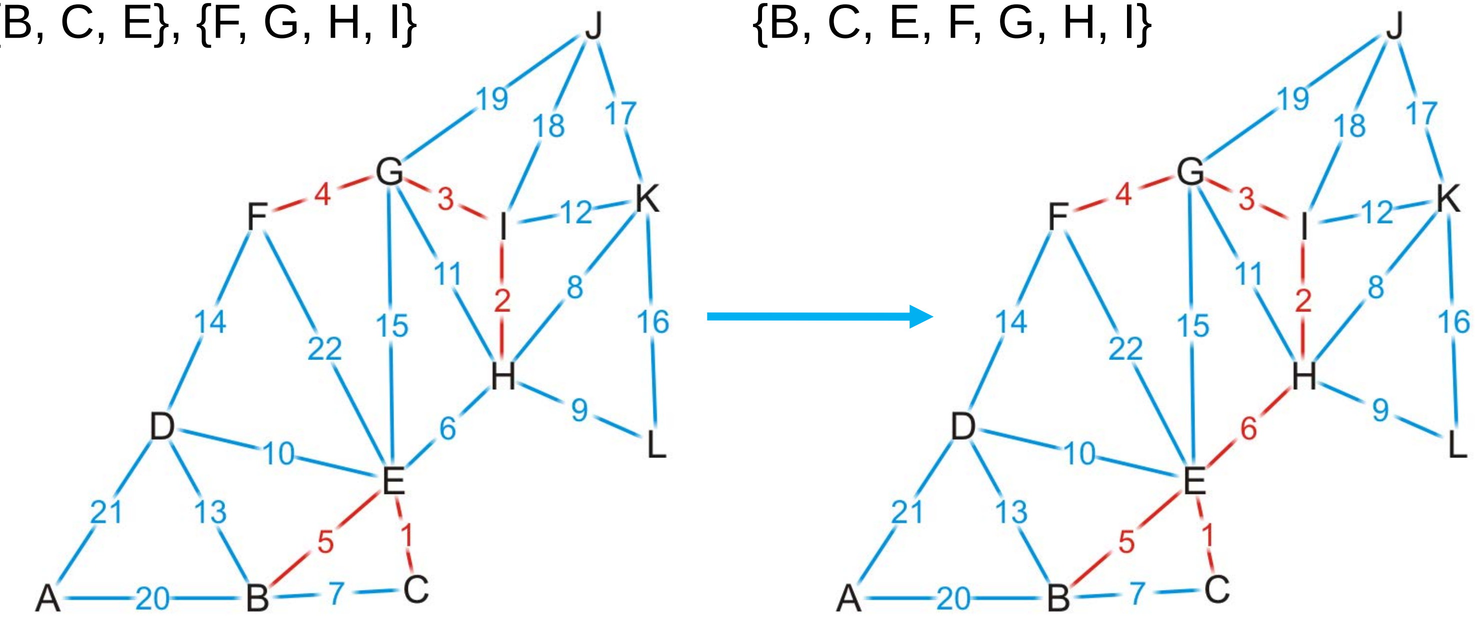
Instead, we could use disjoint sets

- Consider edges in the same connected sub-graph as forming a set
- If the vertices of the next edge are in different sets, take the union of the two sets

Add edge (E, H)?

{B, C, E}, {F, G, H, I}

{B, C, E, F, G, H, I}



Analysis

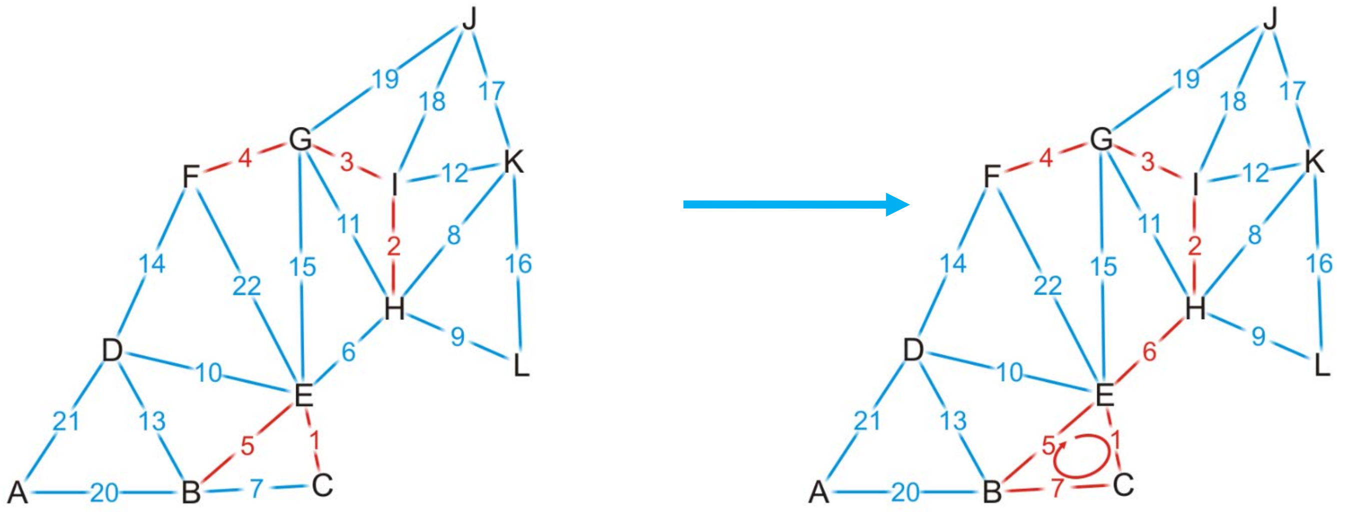
Instead, we could use disjoint sets

- Consider edges in the same connected sub-graph as forming a set
- If the vertices of the next edge are in different sets, take the union of the two sets

Add edge (B, C)?

{B, C, E}, {F, G, H, I}

- Do not add an edge if both vertices are in the same set



Analysis

The disjoint set data structure has run-time $\mathbf{O}(\alpha(n))$, which is effectively a constant

Thus, checking and building the minimum spanning tree is now $\mathrm{O}(|E|)$

The dominant time is now the time required to sort the edges, which is $\mathrm{O}(|E| \ln(|E|)) = \mathrm{O}(|E| \ln(|V|))$

- If there is an efficient $\Theta(|E|)$ sorting algorithm, the run-time is then $\Theta(|E|)$

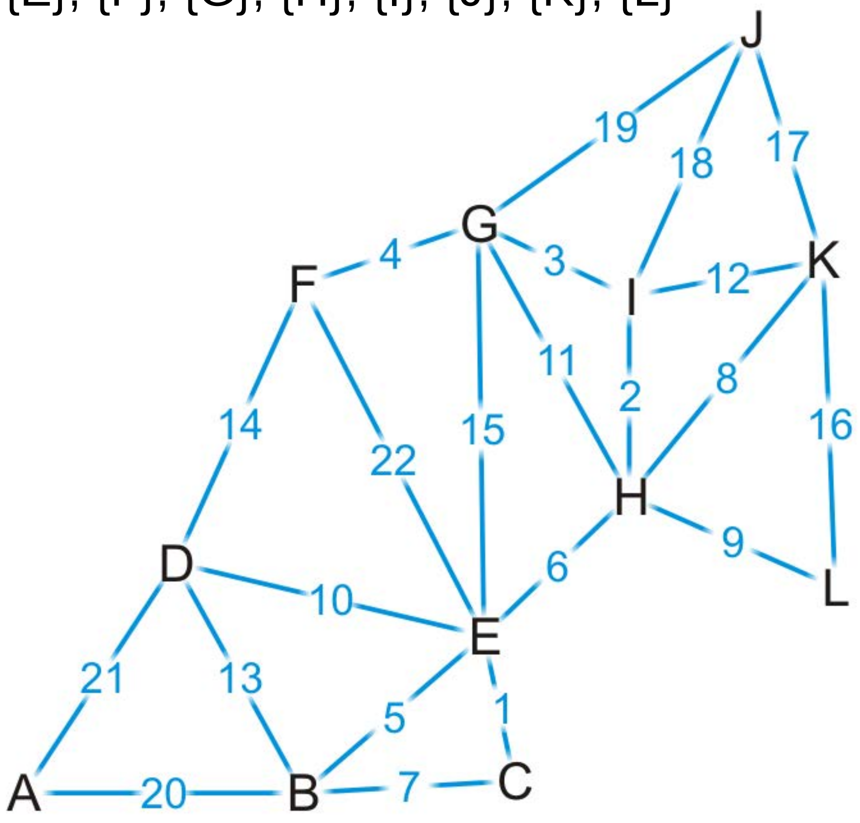
Example

Going through the example again with disjoint sets

Example

We start with twelve singletons

{A}, {B}, {C}, {D}, {E}, {F}, {G}, {H}, {I}, {J}, {K}, {L}

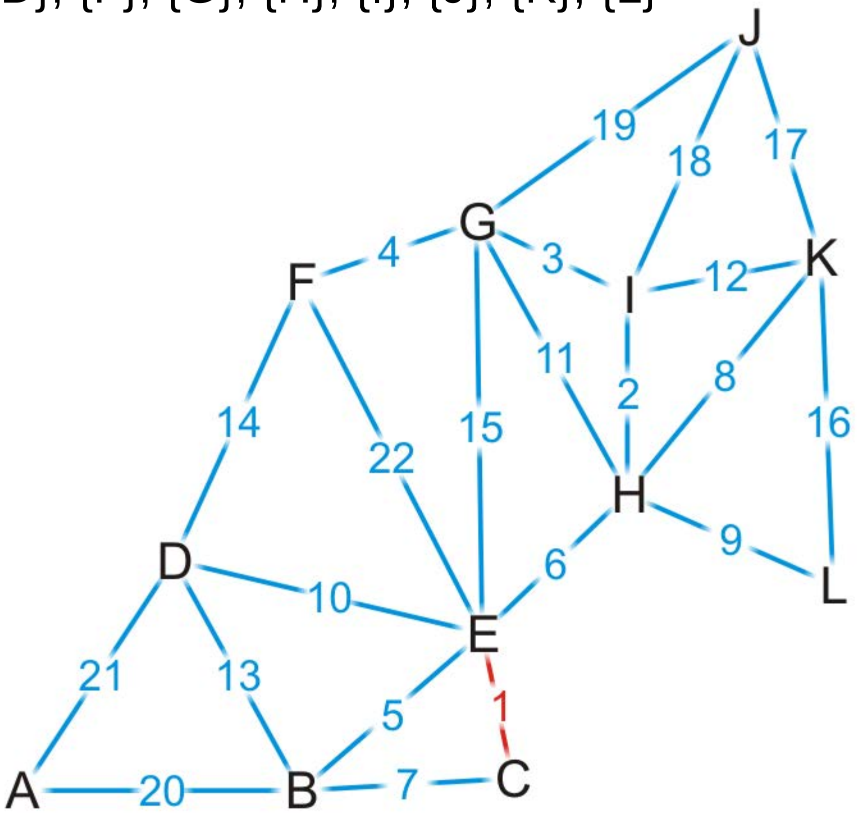


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

We start by adding edge {C, E}

{A}, {B}, {C, E}, {D}, {F}, {G}, {H}, {I}, {J}, {K}, {L}

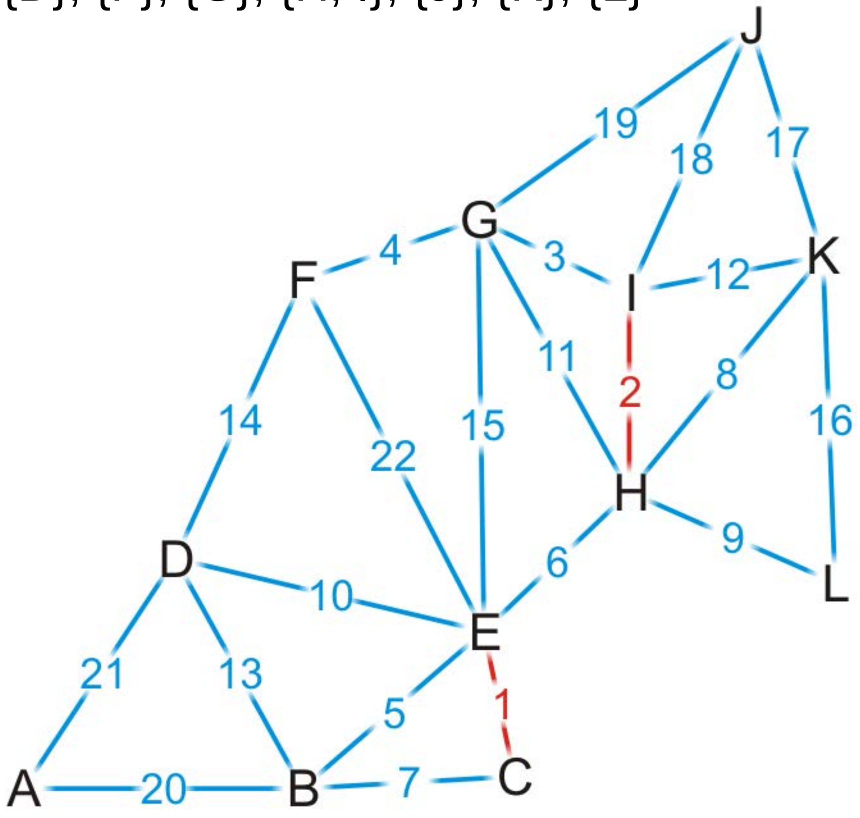


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

We add edge {H, I}

{A}, {B}, {C, E}, {D}, {F}, {G}, {H, I}, {J}, {K}, {L}

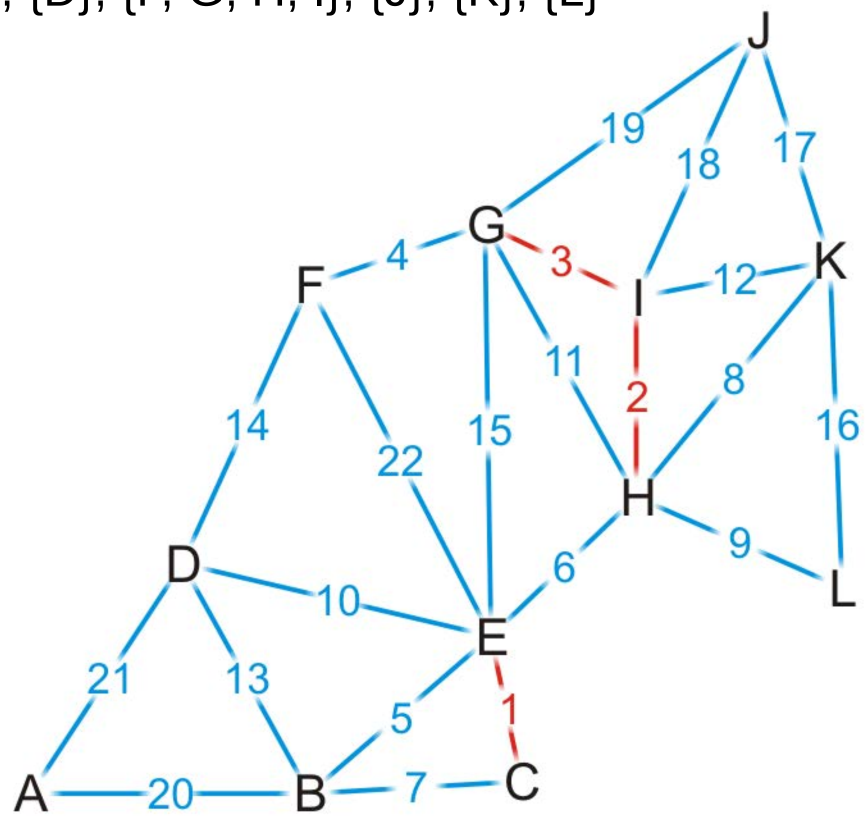


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

Similarly, we add {G, I}, {F, G}, {B, E}

{A}, {B, C, E}, {D}, {F, G, H, I}, {J}, {K}, {L}

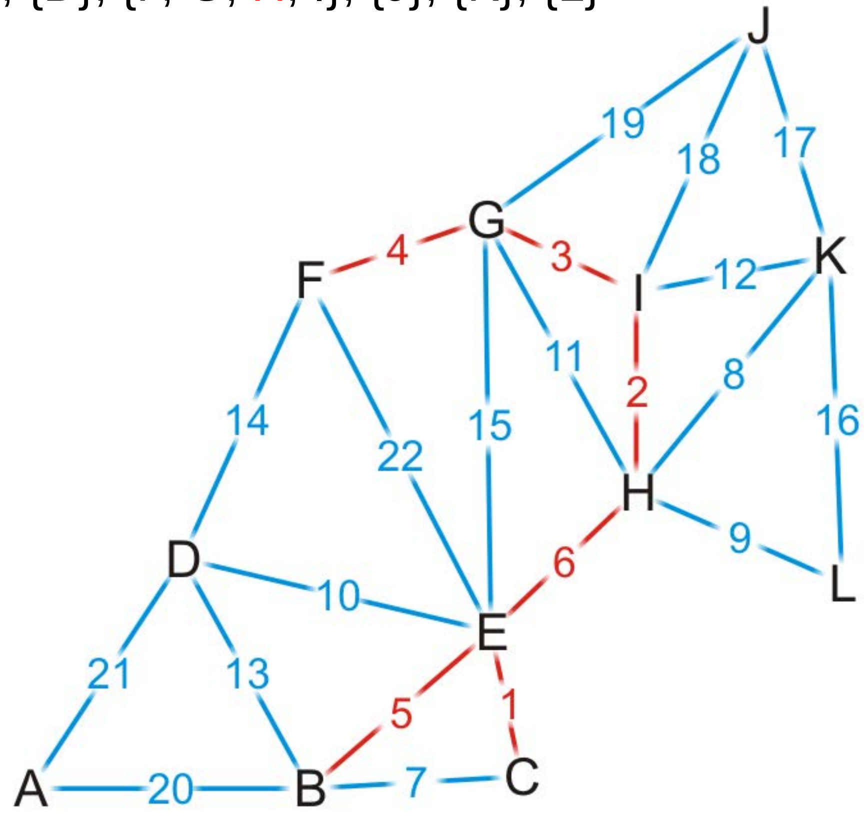


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

The vertices of {E, H} are in different sets

{A}, {B, C, E}, {D}, {F, G, H, I}, {J}, {K}, {L}

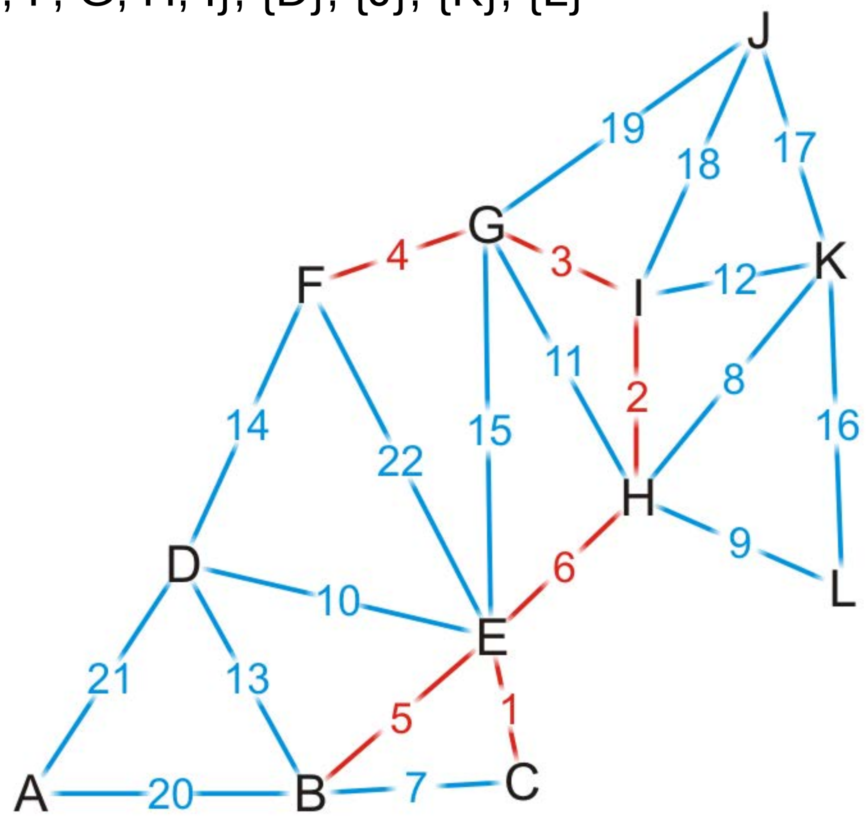


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

Adding edge {E, H} creates a larger union

{A}, {B, C, E, F, G, H, I}, {D}, {J}, {K}, {L}

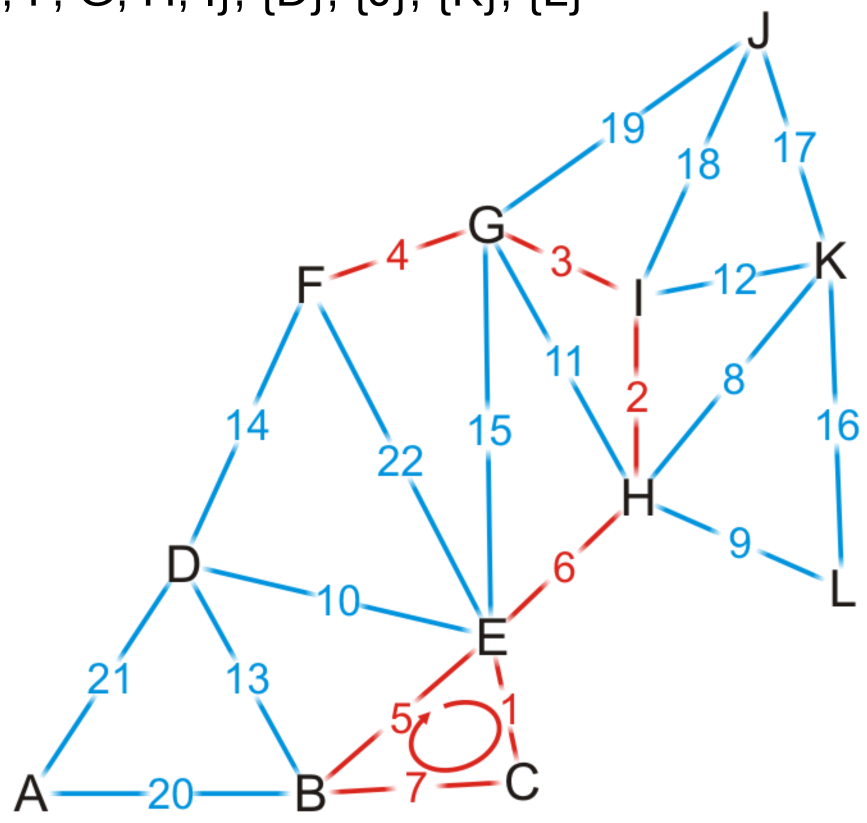


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

We try adding {B, C}, but it creates a cycle

{A}, {B, C, E, F, G, H, I}, {D}, {J}, {K}, {L}

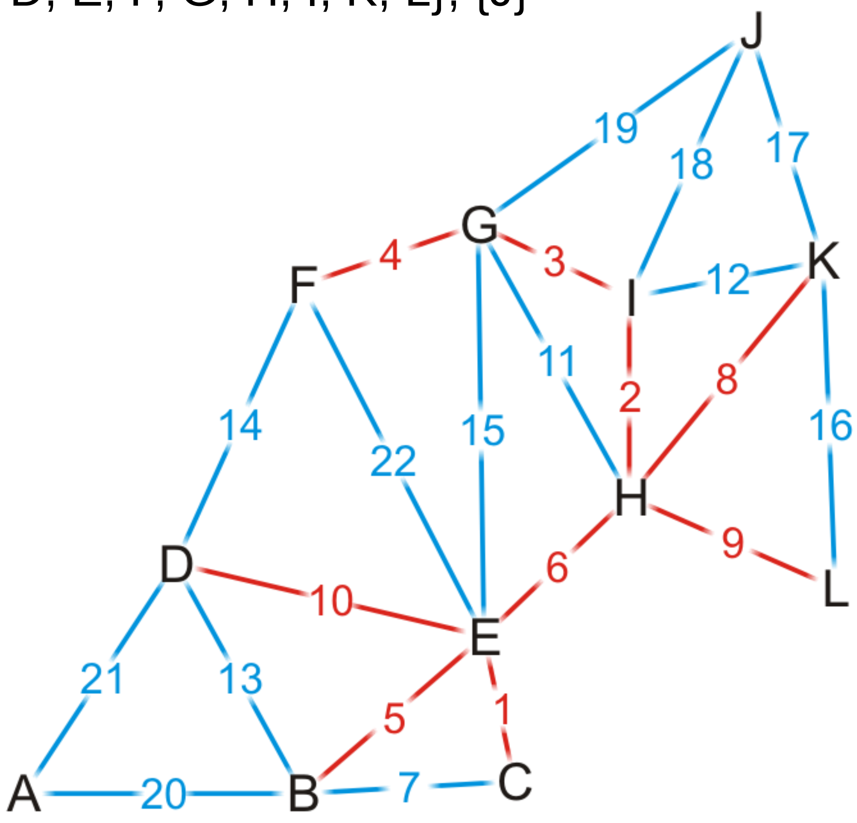


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
→ {B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

We add edge {H, K}, {H, L} and {D, E}

{A}, {B, C, D, E, F, G, H, I, K, L}, {J}

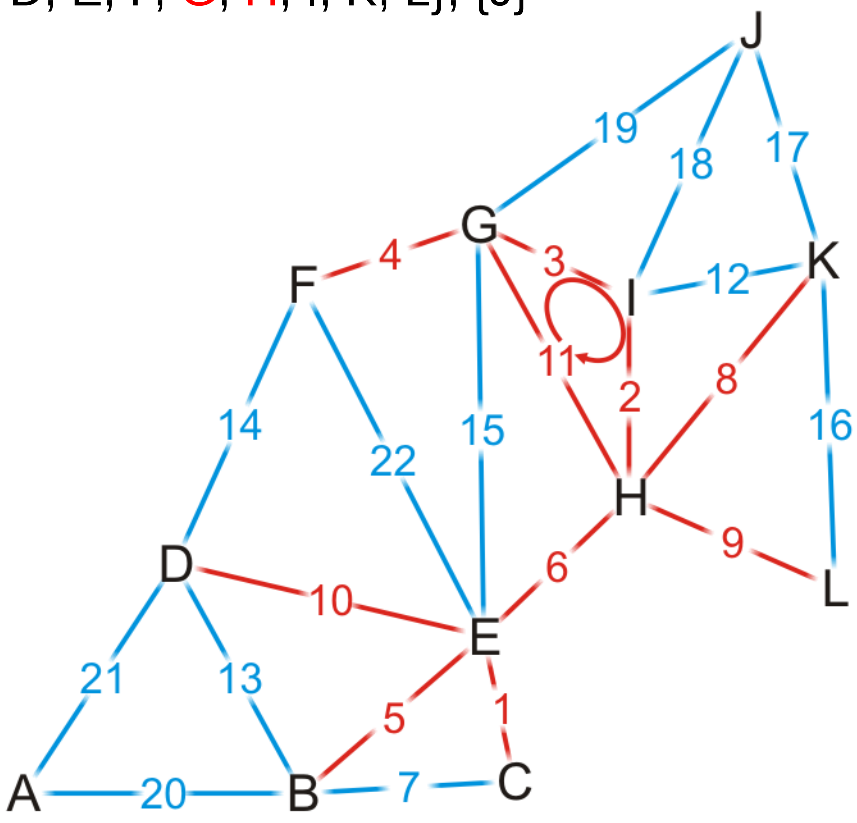


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

Both G and H are in the same set

{A}, {B, C, D, E, F, G, H, I, K, L}, {J}

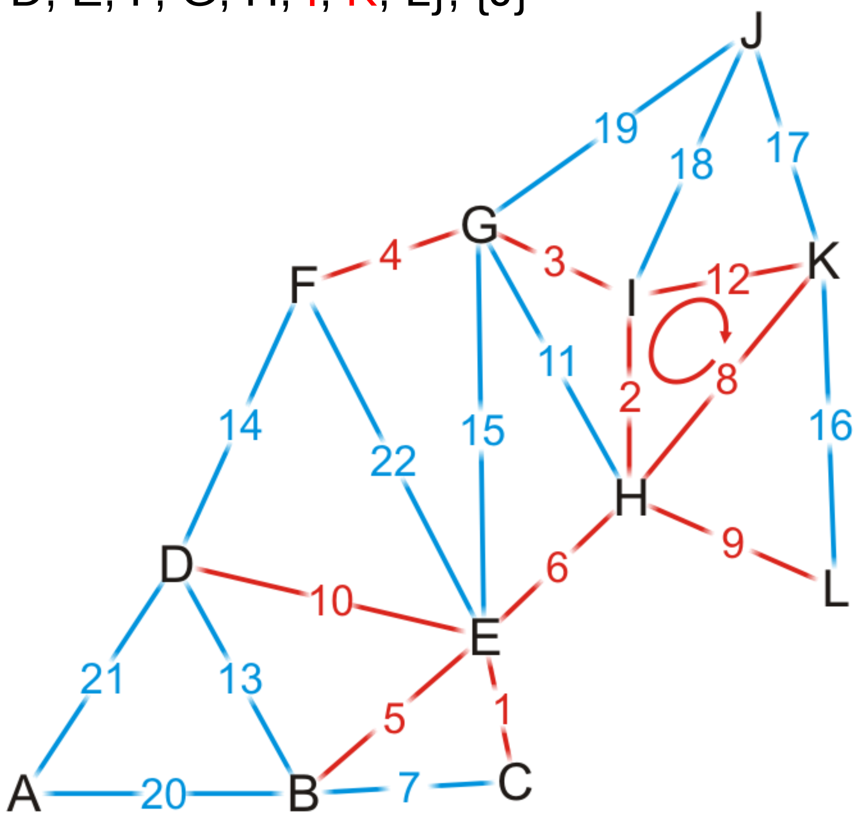


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

Both {I, K} are in the same set

{A}, {B, C, D, E, F, G, H, I, K, L}, {J}

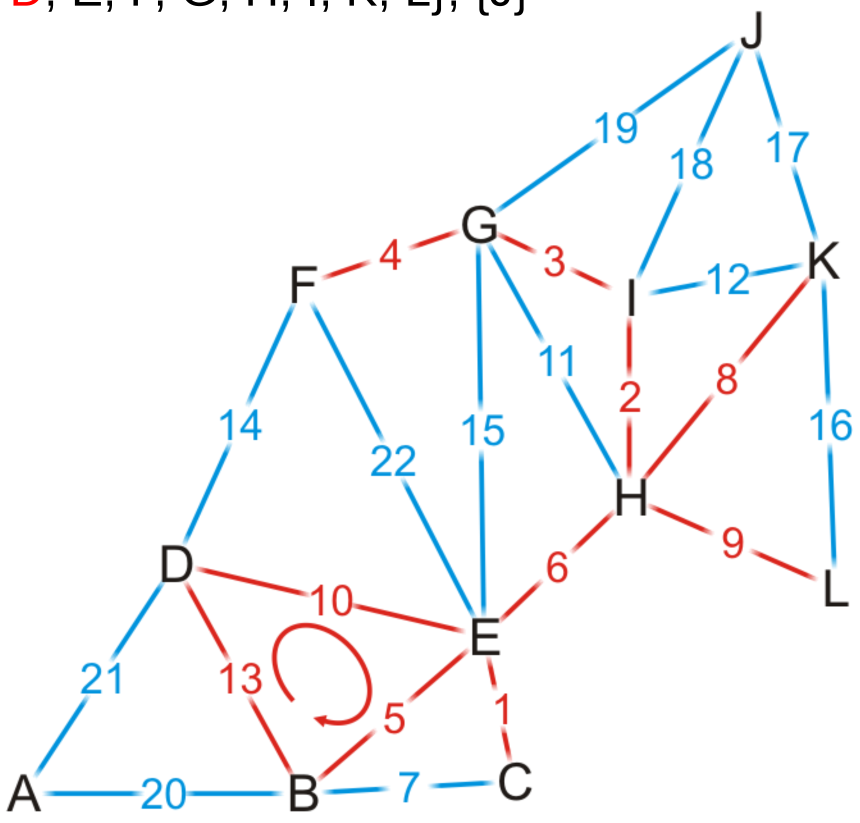


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

Both {B, D} are in the same set

{A}, {B, C, D, E, F, G, H, I, K, L}, {J}

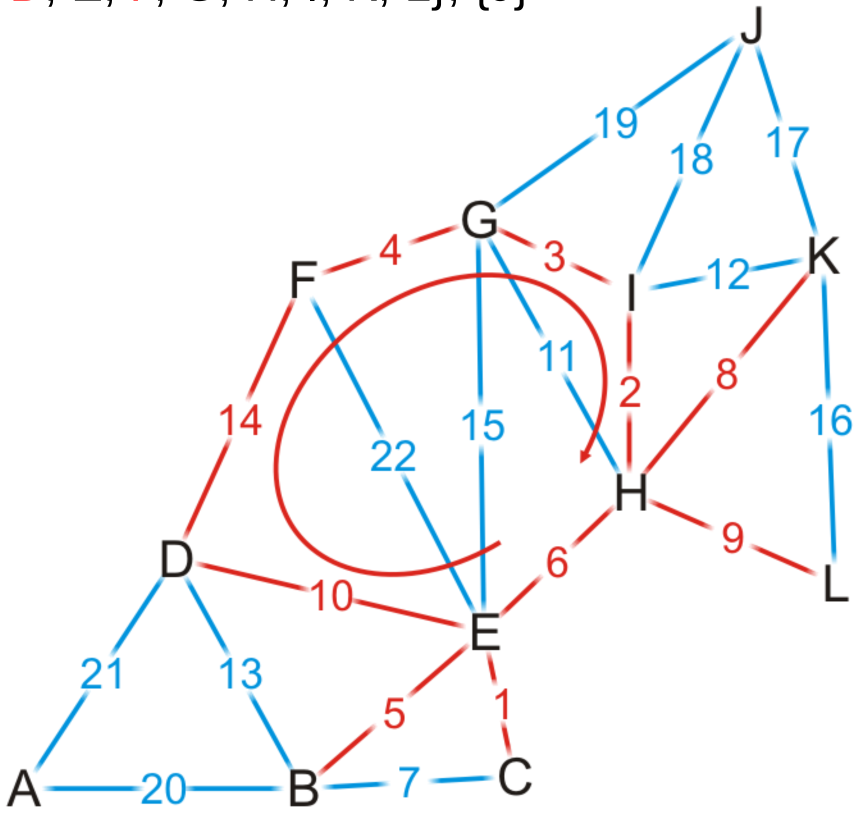


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

Both {D, F} are in the same set

{A}, {B, C, D, E, F, G, H, I, K, L}, {J}

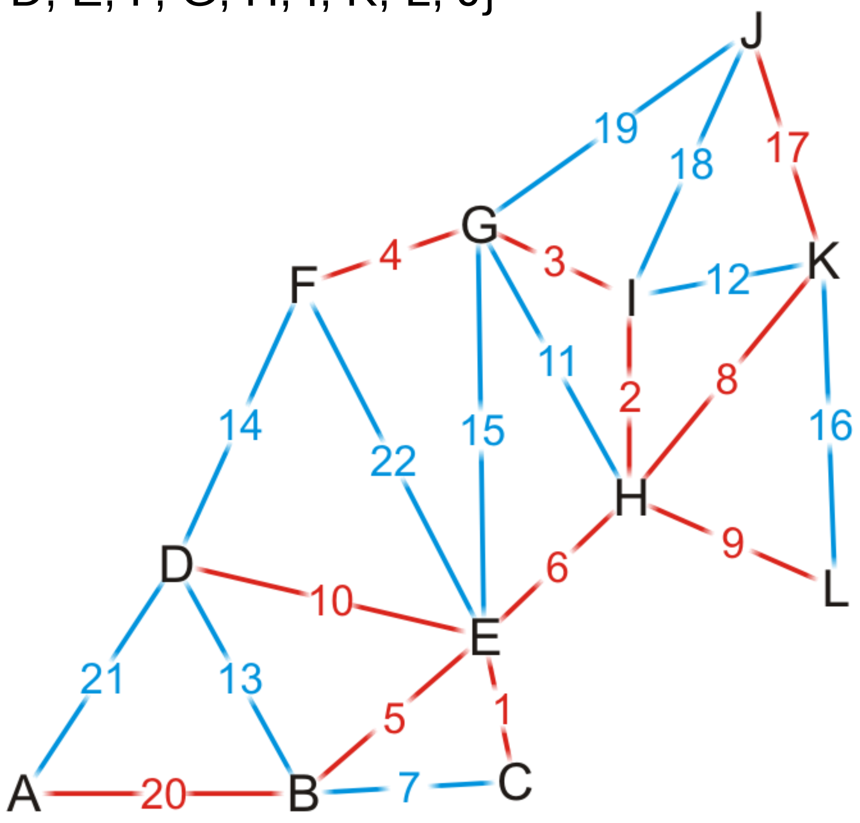


{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Example

We end when there is only one set, having added (A, B)

{A, B, C, D, E, F, G, H, I, K, L, J}



{C, E}
{H, I}
{G, I}
{F, G}
{B, E}
{E, H}
{B, C}
{H, K}
{H, L}
{D, E}
{G, H}
{I, K}
{B, D}
{D, F}
{E, G}
{K, L}
{J, K}
{J, I}
{J, G}
{A, B}
{A, D}
{E, F}

Implementation: Kruskal's Algorithm

Use the union-find data structure.

- Build set T of edges in the MST.
- Maintain set for each connected component.
- O(m log m) for sorting.

```
Kruskal(G, c) {
   Sort edges weights so that c₁ ≤ c₂ ≤ ... ≤ cₘ.
   T ← φ

   foreach (u ∈ V) make a set containing singleton u

   for i = 1 to m        are u and v in different connected components?
      (u,v) = eᵢ
      if (u and v are in different sets) {
         T ← T ∪ {eᵢ}
         merge the sets containing u and v
      }                  merge two components
   return T
}
```

Topological Sort

Topological Sort

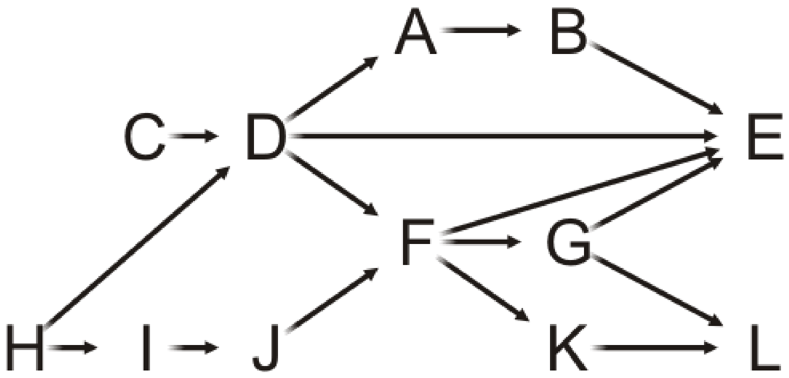
Idea:

- Given a DAG V, iterate:
  - Find a vertex v in V with in-degree zero
  - Let v be the next vertex in the topological sort
  - Continue iterating with the vertex-induced sub-graph $V \setminus \{v\}$

Example

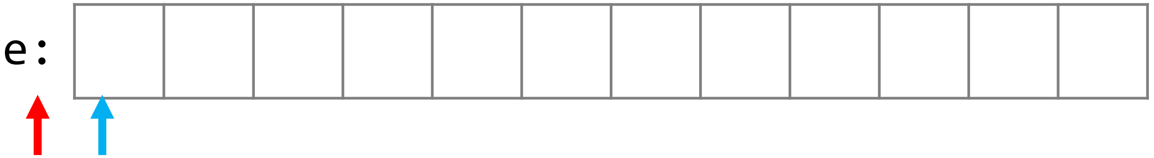
With the previous example, we initialize:

- The array of in-degrees
- The queue



A	1
B	1
C	0
D	2
E	4
F	2
G	1
H	0
I	1
J	1
K	1
L	2

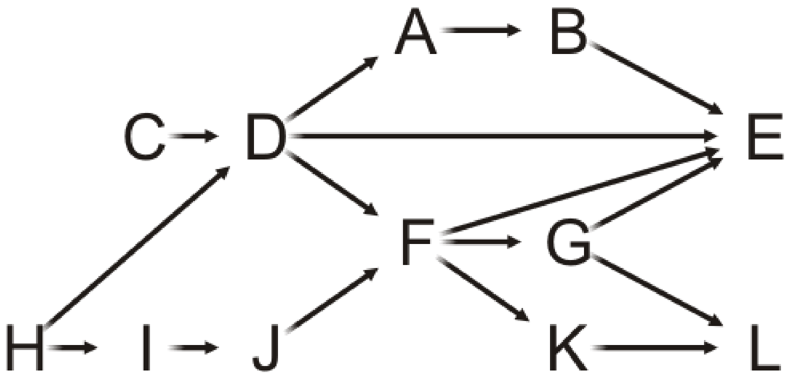
Queue:



The queue is empty

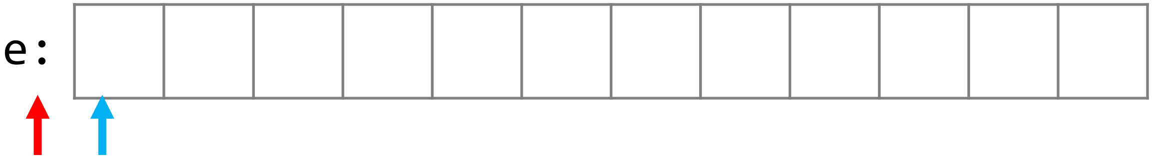
Example

Stepping through the array, push all source vertices into the queue



A	1
B	1
C	0
D	2
E	4
F	2
G	1
H	0
I	1
J	1
K	1
L	2

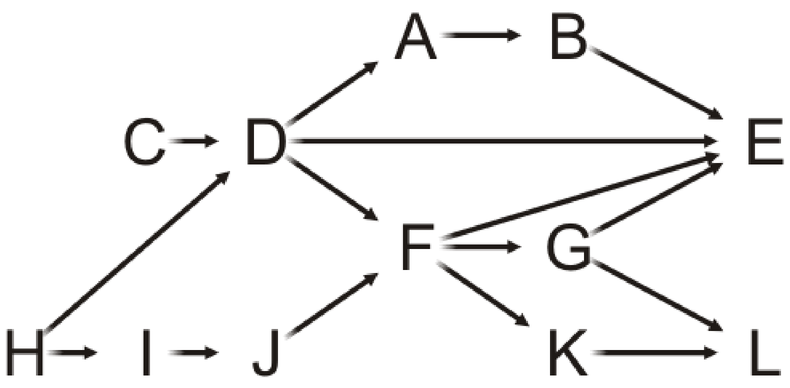
Queue:



The queue is empty

Example

Stepping through the table, push all source vertices into the queue



A	1
B	1
C	**0**
D	2
E	4
F	2
G	1
H	**0**
I	1
J	1
K	1
L	2

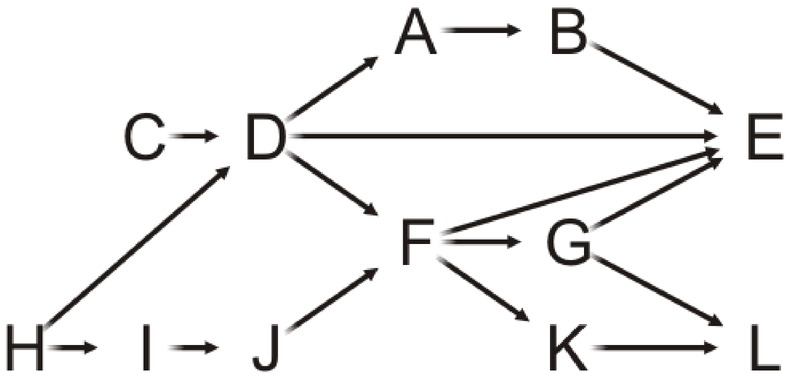
Queue:

C	**H**										

The queue is empty

Example

Pop the front of the queue



Queue:

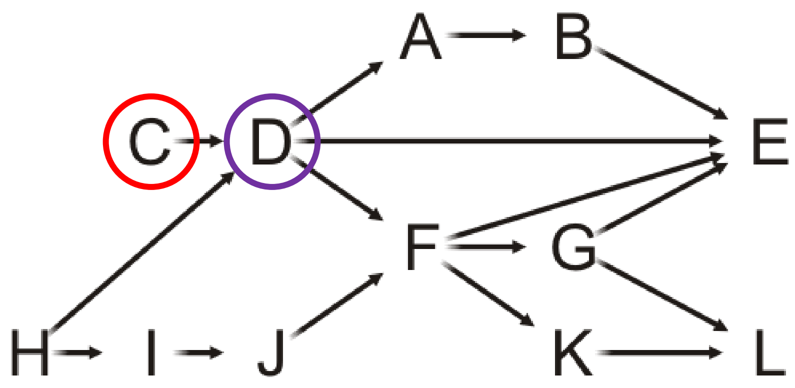
C	H										

A	1
B	1
C	0
D	2
E	4
F	2
G	1
H	0
I	1
J	1
K	1
L	2

Example

Pop the front of the queue

- C has one neighbor: D



Queue:

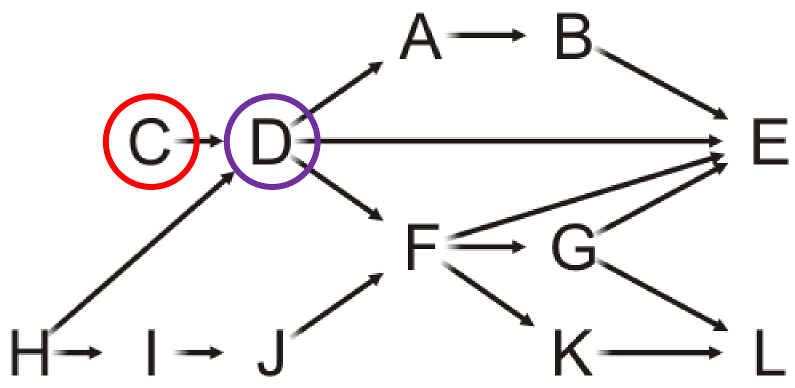
C	H										

A	1
B	1
C	**0**
D	**2**
E	4
F	2
G	1
H	0
I	1
J	1
K	1
L	2

Example

Pop the front of the queue

- C has one neighbor: D
- Decrement its in-degree



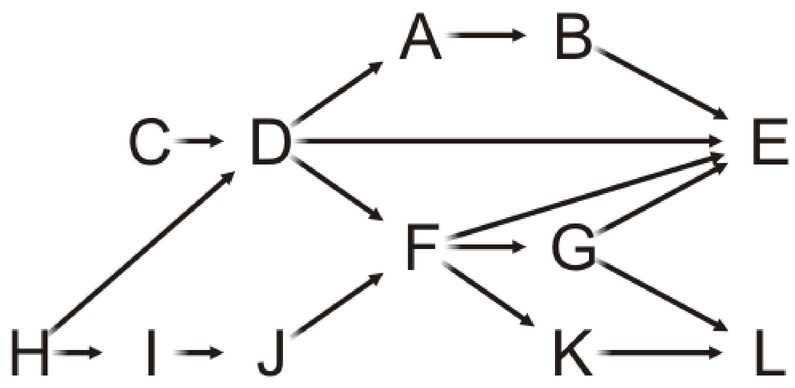
Queue:

C	H										

A	1
B	1
C	**0**
D	**1**
E	4
F	2
G	1
H	0
I	1
J	1
K	1
L	2

Example

Pop the front of the queue



Queue:

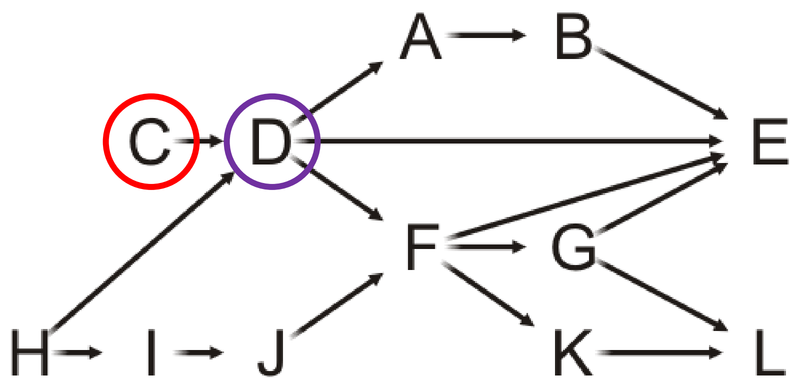
C	H										

A	1
B	1
C	0
D	1
E	4
F	2
G	1
H	0
I	1
J	1
K	1
L	2

Example

Pop the front of the queue

- C has one neighbor: D
- Decrement its in-degree



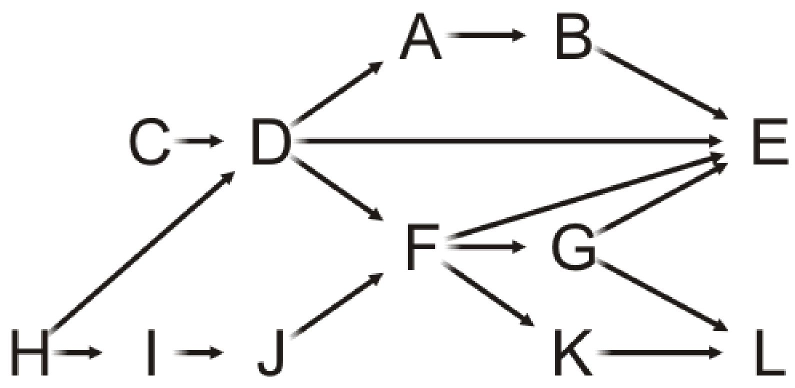
Queue:

C	H										

A	1
B	1
C	**0**
D	**1**
E	4
F	2
G	1
H	0
I	1
J	1
K	1
L	2

Example

Pop the front of the queue



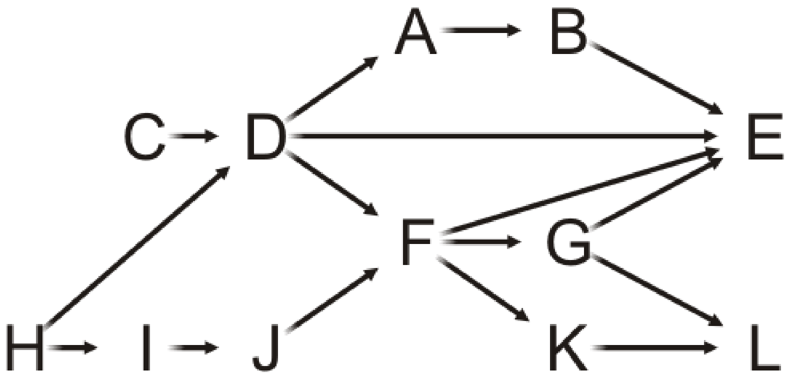
Queue:

C	H										

A	1
B	1
C	0
D	1
E	4
F	2
G	1
H	0
I	1
J	1
K	1
L	2

Example

The array used for the queue stores the topological sort



C	H	D	I	A	J	B	F	G	K	E	L

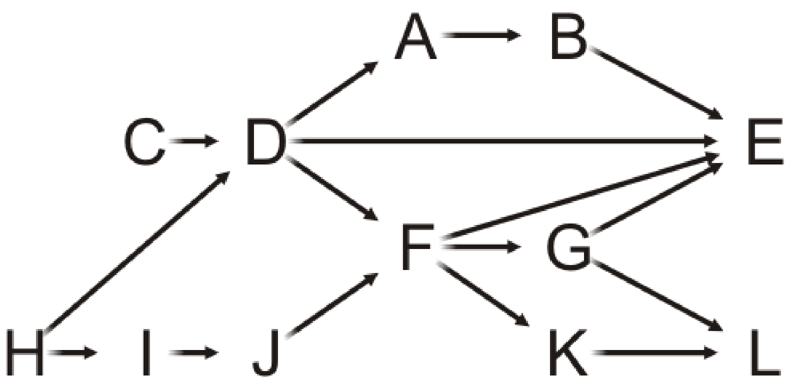
A	0
B	0
C	0
D	0
E	0
F	0
G	0
H	0
I	0
J	0
K	0
L	0

Example

The array used for the queue stores the topological sort

- Note the difference in order from our previous sort?

C, H, D, A, B, I, J, F, G, E, K, L

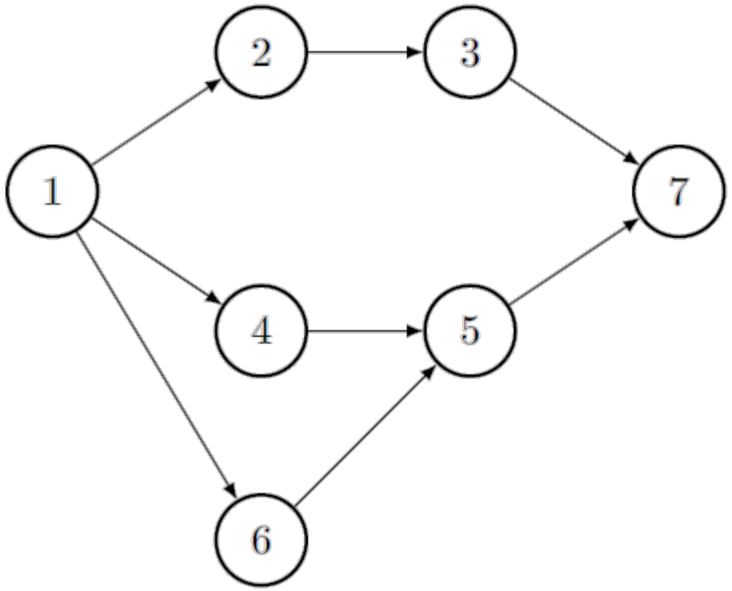


C	H	D	I	A	J	B	F	G	K	E	L

A	0
B	0
C	0
D	0
E	0
F	0
G	0
H	0
I	0
J	0
K	0
L	0

Question

- Count the number of possible topological sorts for the following graph. Show your detailed counting method. Do NOT enumerate.



Answer:

Consider this question as placing 5 balls into 5 boxes. We first place 2, 3; and we have $C_5^2 = 10$ ways to do this.

Then the rest of the boxes will be given to 4, 5, 6; the order is either 4, 6, 5 or 6, 4, 5.