#### **Notes on Minimum Spanning Tree problem**

#### A Priority queue trouble

Usual theoretical and conceptual descriptions of Prim's (Dijkstra's, etc.) algorithm say: "... store nodes in a priority queue".

Technically, this is nearly impossible to do.

The graph and the nodes are defined separately and stored elsewhere in the memory. The node has no reference to its position in the queue.

The programmer does not know where is the node in the queue.

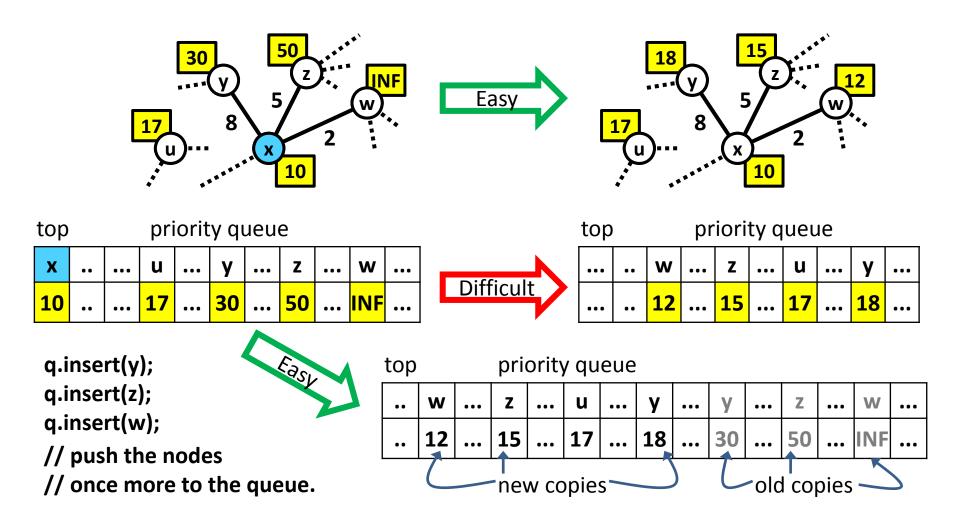
So, how to move a node inside the queue according to the algorithm demands? Standard solution:

Do not move a node, enqueue a "copy of a node", possibly more times.

When a copy of the node with the smallest value (=highest priority) among all its copies appears at the top of the queue

it does its job exactly according to the algorithm prescription.

From that moment on, all other copies of the node which are still in the queue become useless and must be ignored. The easiest way to ignore a copy is to check it when it later appears at the top of the queue: If the node is already closed, ignore the copy, pop it and process the next top of the queue. If the node is still open, process it according to the algorithm.



The older copies of nodes will get to the top of the queue later than the new copies (which have higher priority). The older copy gets to the top when the node had been processed and closed earlier. Thus:

If the node at the top of the queue is closed just pop it and do not process it any more.

#### **Example of Prim's algorithm implementation using standard library priority queue**

```
void MST_Prim(Graph g, int start, final int [] dist, int [] pred ) {
  // allocate structures
  int currnode = start, currdist, neigh;
  int INF = Integer.MAX VALUE;
  boolean [] closed = new boolean[g.N];
  PriorityQueue <Integer> pg
    = new PriorityQueue<Integer>( g.N,
          new Comparator<Integer>() {
            @Override
            public int compare(Integer n1, Integer n2) {
              if( dist[n1] < dist[n2] ) return -1;</pre>
              if( dist[n1] > dist[n2] ) return 1;
              return 0;
          } } );
  // initialize structures
  pq.add( start );
  for( int i = 0; i < g.N; i++ ) pred[i] = i;</pre>
  Arrays.fill( dist, INF );
  Arrays.fill( closed, false );
  dist[start] = 0;
```

#### Example of Prim's algorithm implementation using standart library priority queue

```
for( int i = 0; i < g.N; i++ ) {</pre>
 // take the closest node and skip the closed ones
 while( closed[currnode = pq.poll()] == true );
  // and expand the closest node
  for( int j = 0; j < g.dg[currnode]; j++ ){</pre>
   neigh = g.edge[currnode][j];
    if( !closed[neigh] &&
       ( dist[neigh] > g.w[currnode][j]) ) {
         dist[neigh] = g.w[currnode][j];
         pred[neigh] = currnode;
         pq.add(neigh);
  } // for j
  closed[currnode] = true;
 // for i
  MST Prim
```

#### A very small change produces Dijkstra's algorithm:

```
if(!closed[neigh] &&
   ( dist[neigh] > g.w[currnode][j] + dist[currnode]) ) {
    dist[neigh] = g.w[currnode][j] + dist[currnode];
```

## Prim and Dijkstra Algorithms compared

The only difference:

```
void Dijkstra( Graph G, function weight, Node startnode )
for each u in G.V: u.dist = INFINITY; u.parent = NIL
startnode.dist = 0; PriorityQueue Q = G.V

while !Q.isEmpty()
    u = Extract-Min(Q)
    for each v in G.Adj[u]
    if (v in Q) and v.dist > weight(u,v) + u.dist
    v.parent = u
    v.dist = weight(u,v) + u.dist
```

```
void MST_Prim( Graph G, function weight, Node startnode )
for each u in G.V: u.dist = INFINITY; u.parent = NIL
startnode.dist = 0; PriorityQueue Q = G.V

while !Q.isEmpty()
    u = Extract-Min(Q)
    for each v in G.Adj[u]
    if (v in Q) and v.dist > weight(u,v)
    v.parent = u
    v.dist = weight(u,v)
```

## Disjoint-set data structure alias Union-Find structure

A graph with spanning trees of its subgraphs A, B, C.

Each subgraph is represented by one tree in the Union-Find structure.

Union-Find structure with trees corrresponding to subgraphs A, B, C.

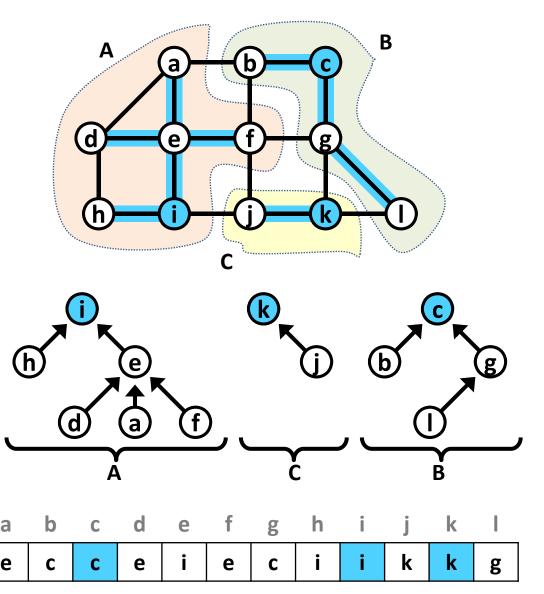
Tree roots represent the subgraphs and/or their spanning trees.

node

boss

Implementation of the Union-Find structure.

A tree root is called here "boss", for brevity.



#### Disjoint-set data structure alias Union-Find structure

Only 3 operations are needed:

```
Initialize()
```

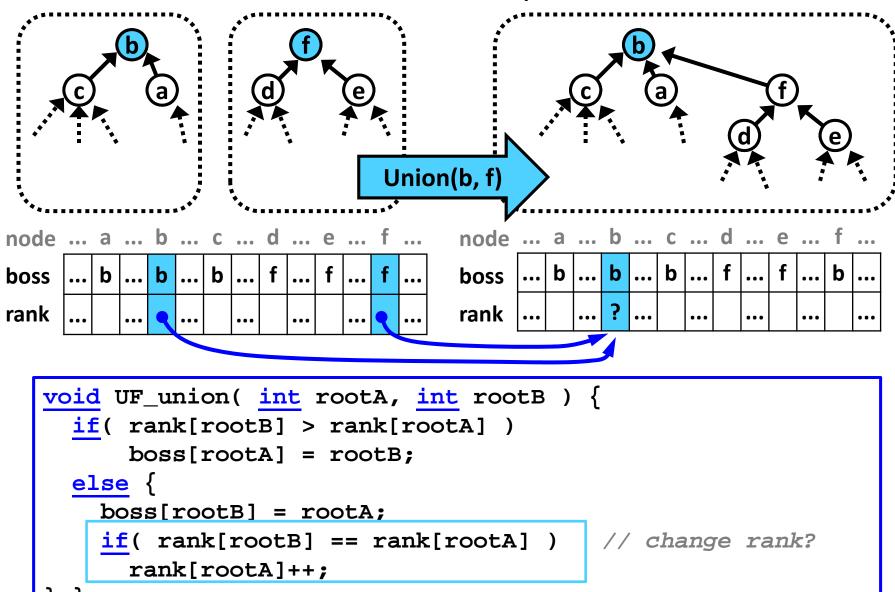
```
int [] boss; // = representative = tree root
int [] rank;

void UF_init( int n ) {
  boss = new int [n];
  rank = new int [n];
  for( int i = 0; i < n; i++ ) {
    boss[i] = i; // everybody's their own boss
    rank[i] = 0; // initial rank is 0
} }</pre>
```

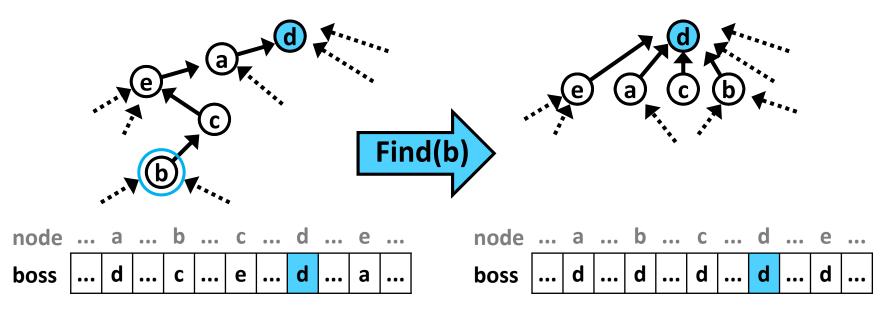
Easy experiment, try it at home:

When the end nodes of the inspected edges are chosen uniformly randomly then the average depth of a queried node in the Union-Find forest is less than 2.

#### Union with rank comparison

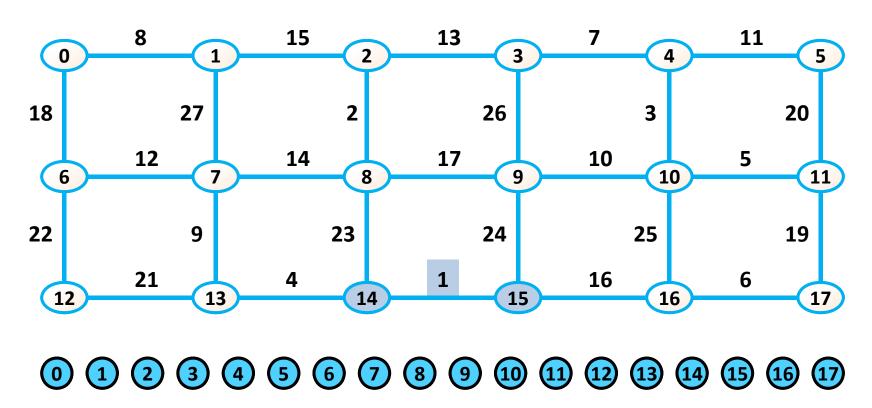


#### Find with path compression

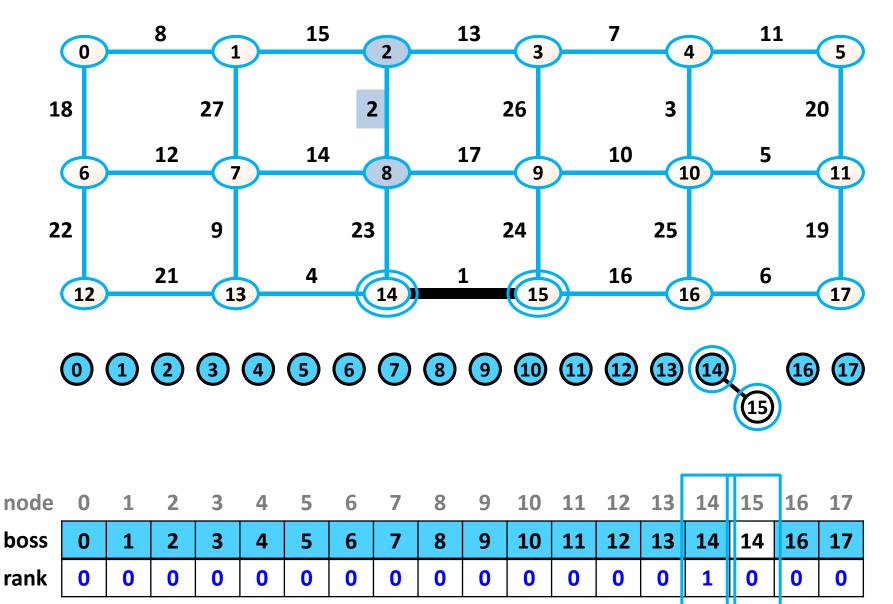


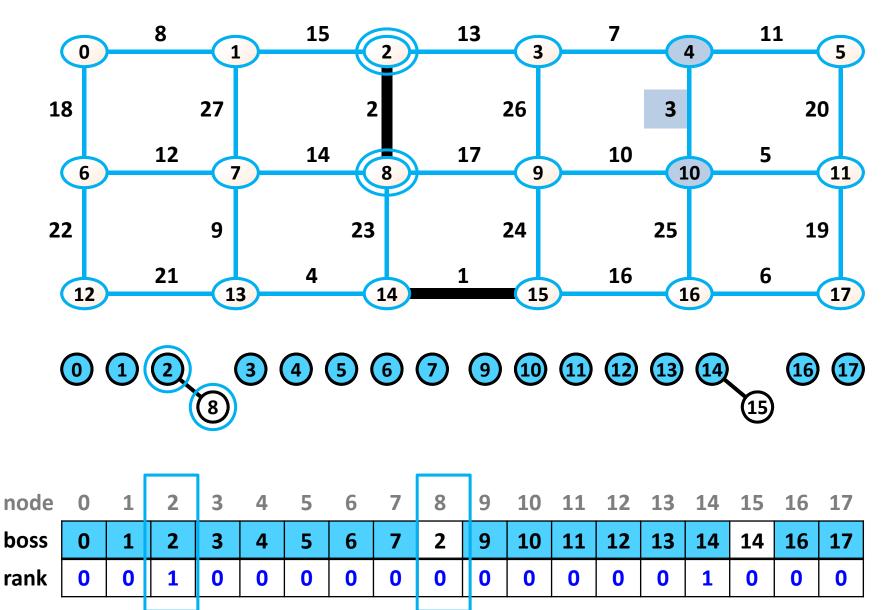
```
int UF_find( int a ) {
  int parent = boss[a];
  if( parent != a )
    boss[a] = UF_find( parent ); // path compression
  return boss[a];
}
```

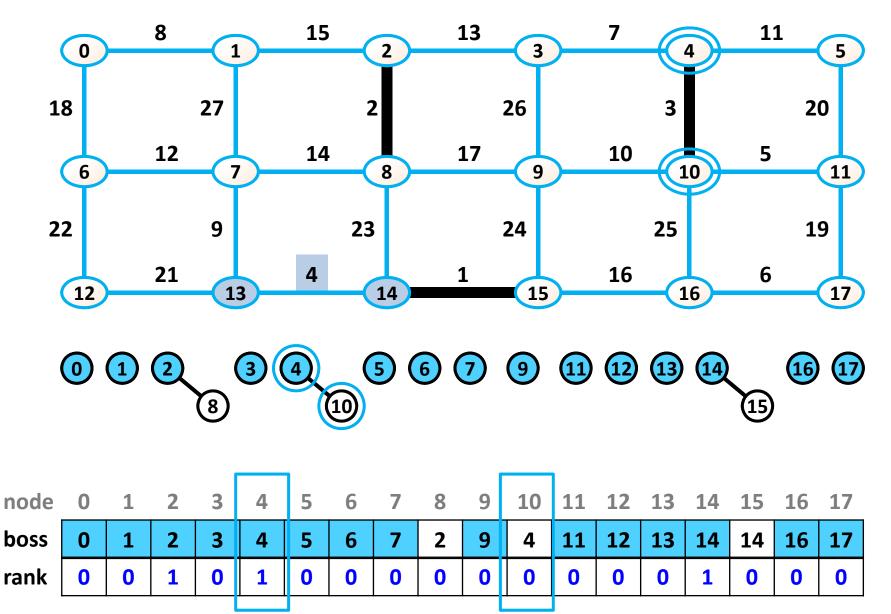
```
int find( int a ) // for C experts
{ return( boss[a] == a ? a : (boss[a] = find(boss[a])) );}
```

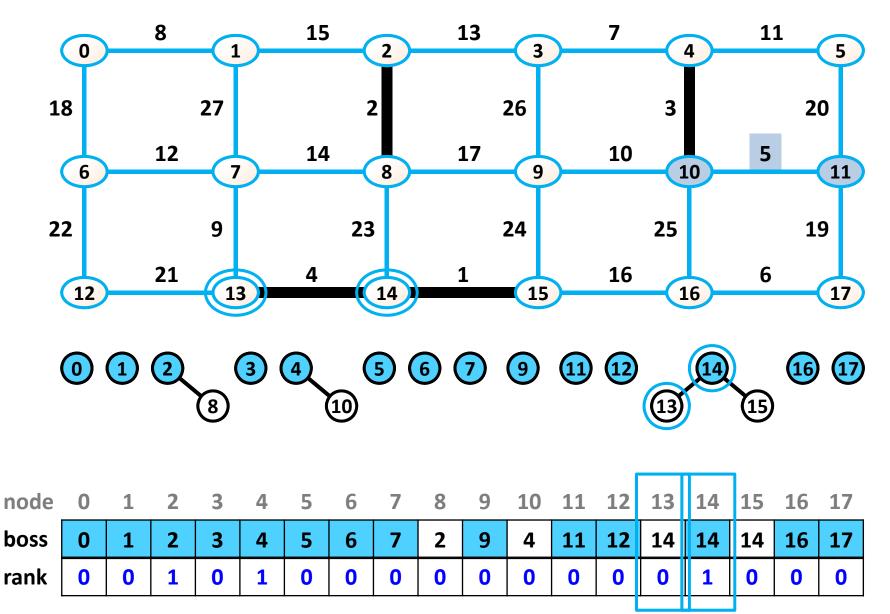


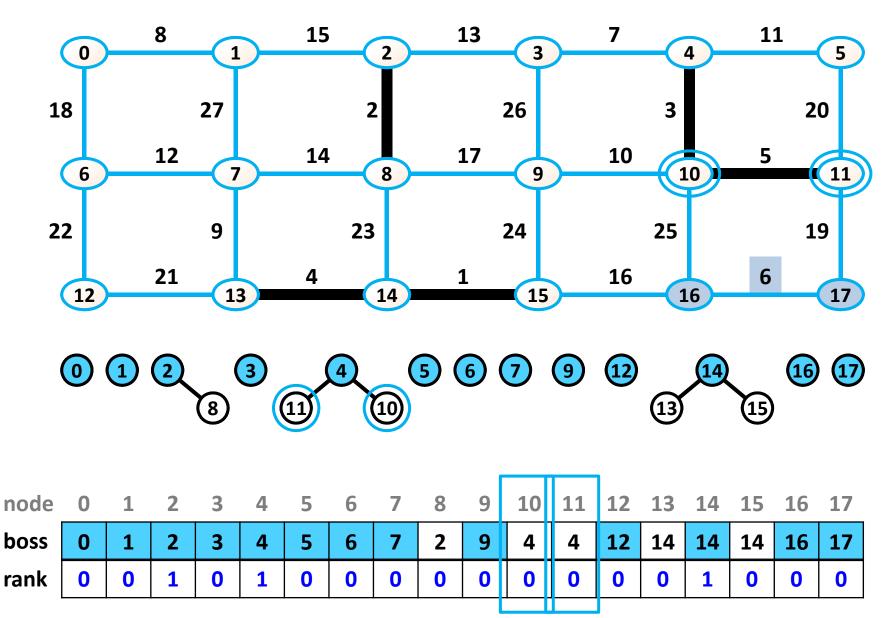
node	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
boss	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
rank	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

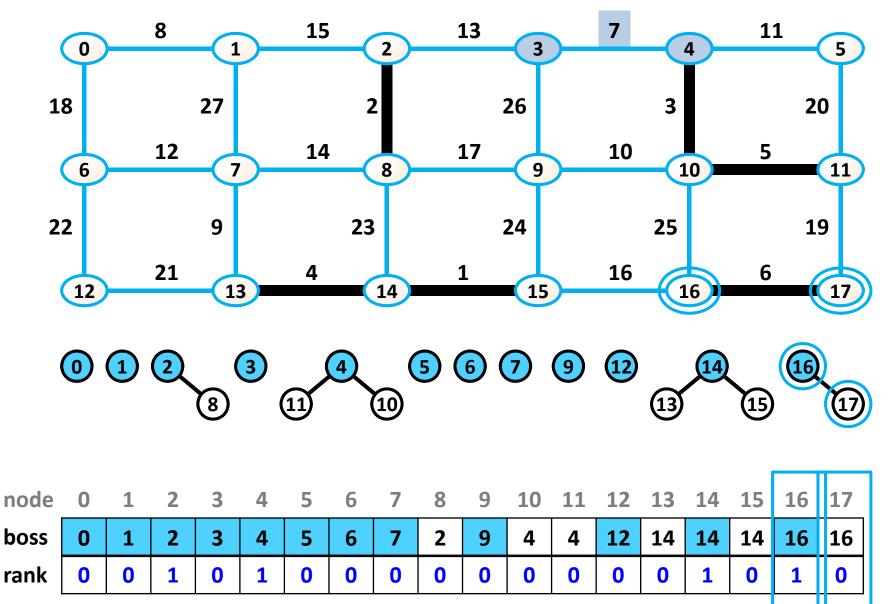


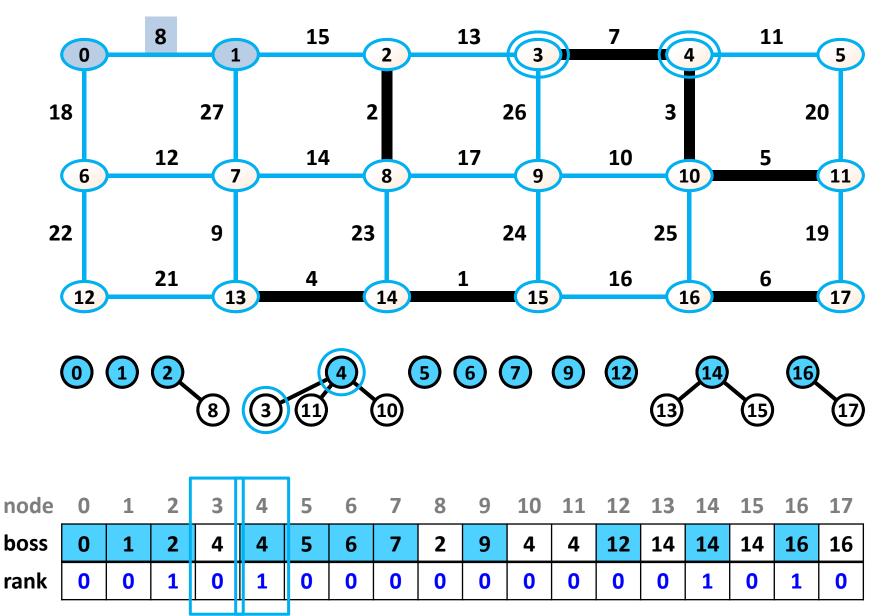


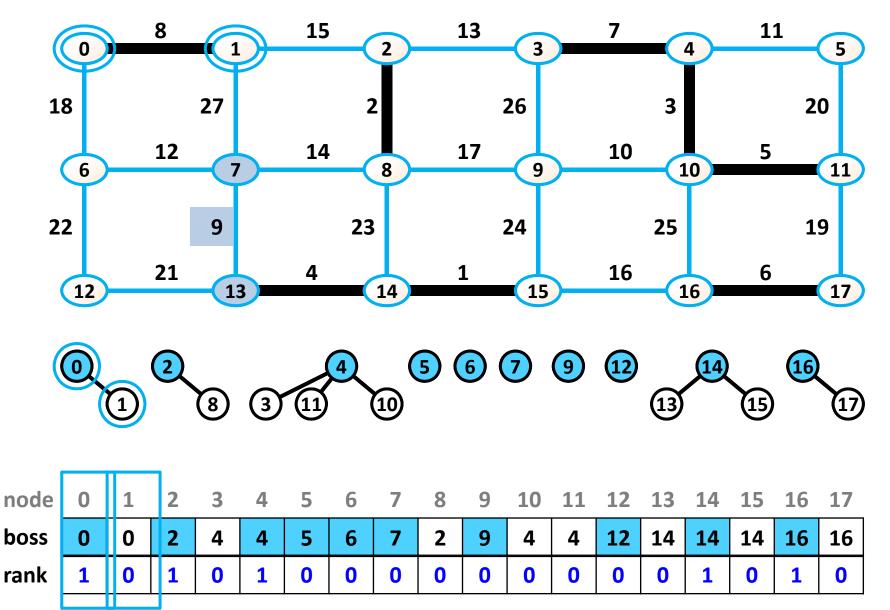


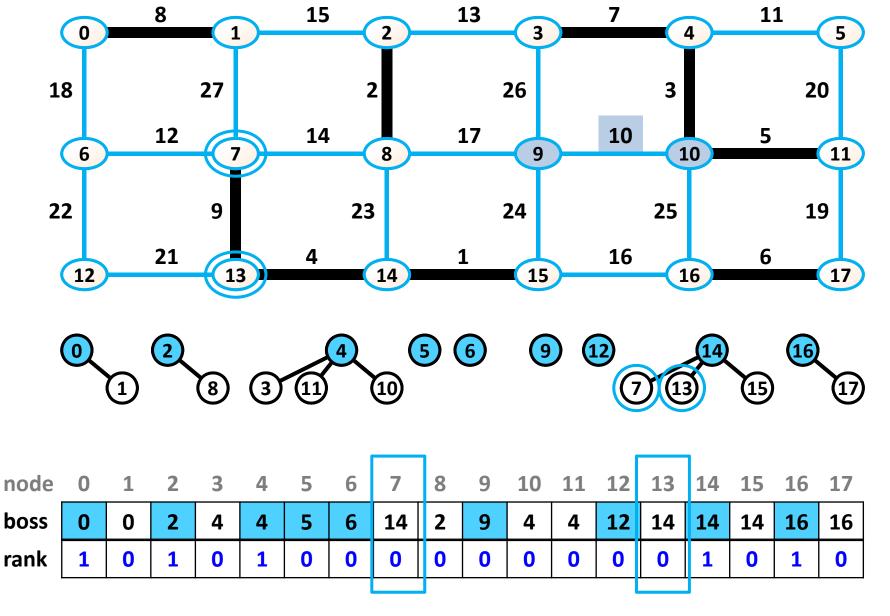


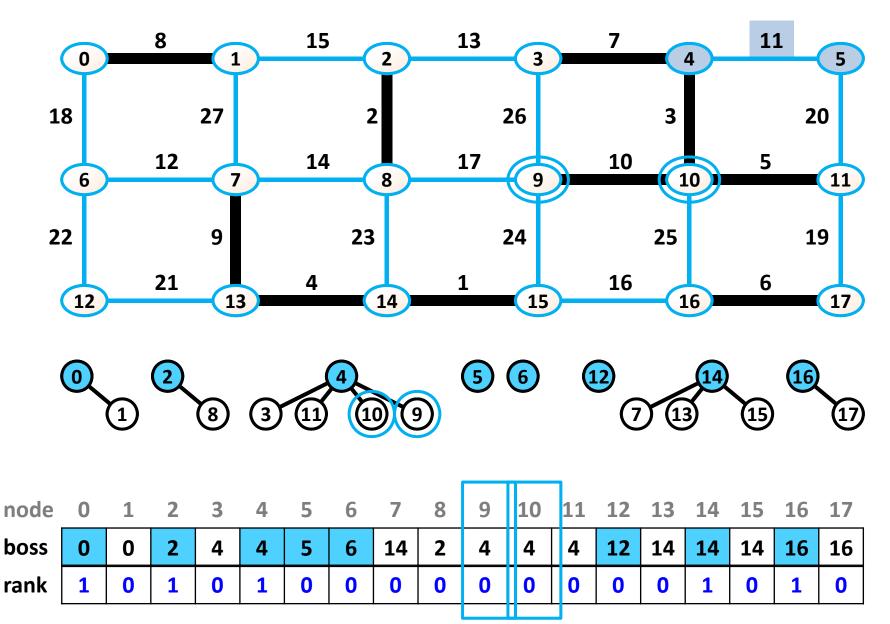


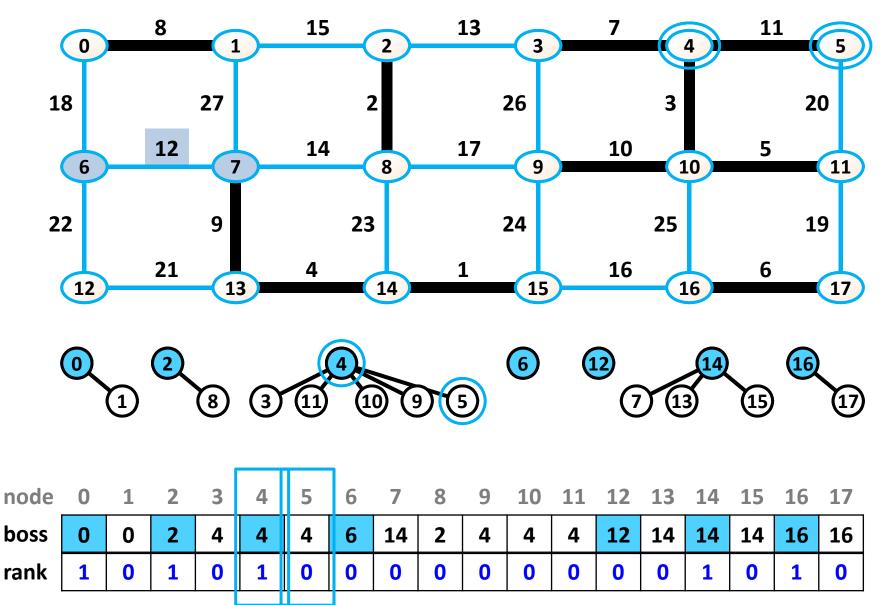


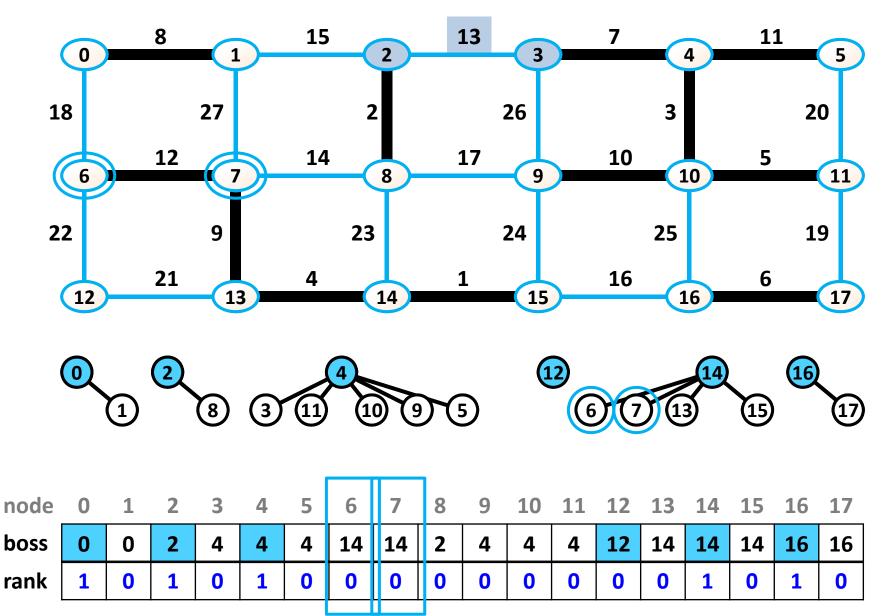


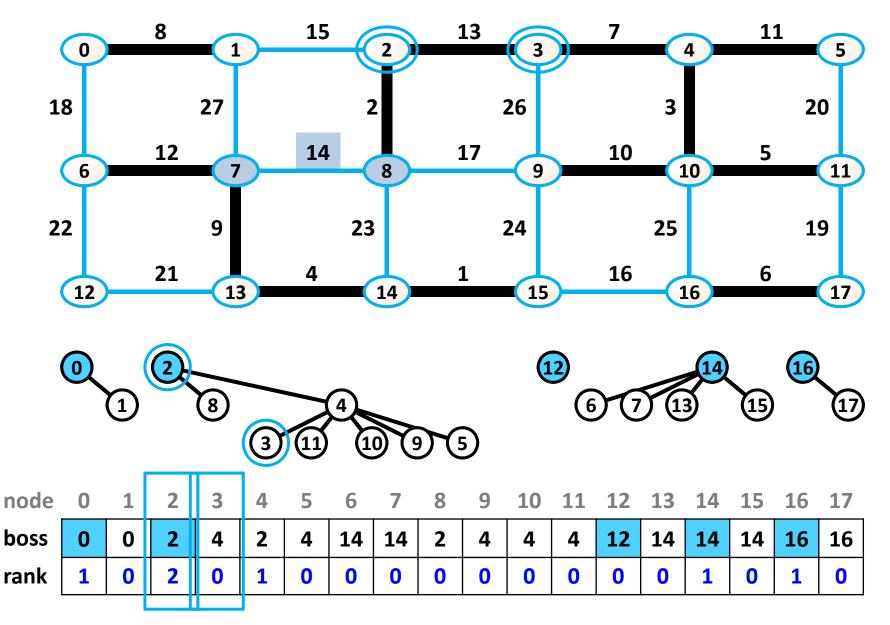


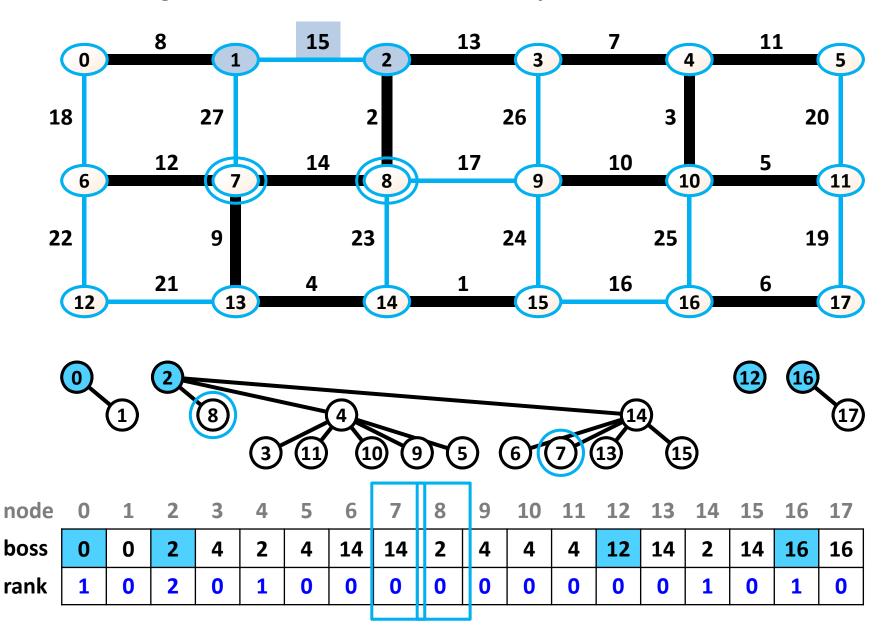


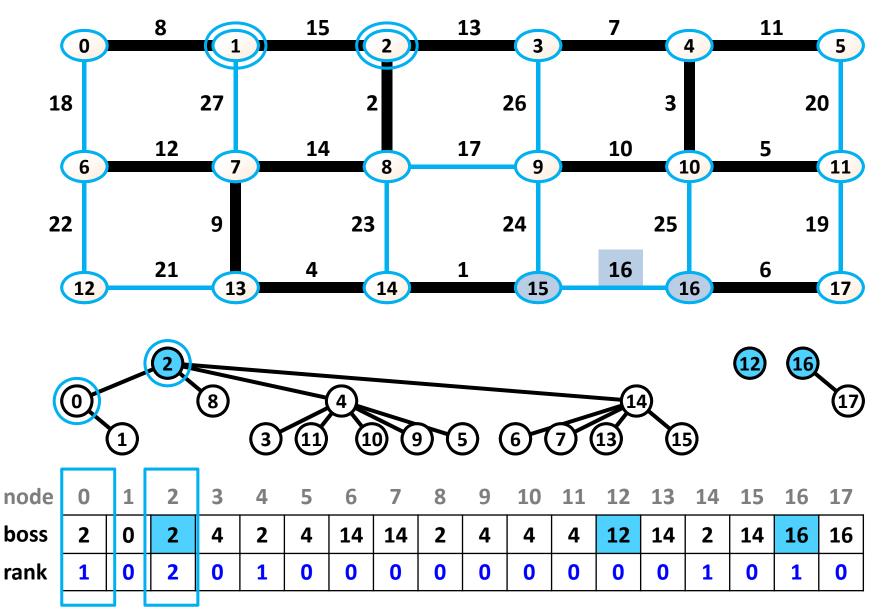


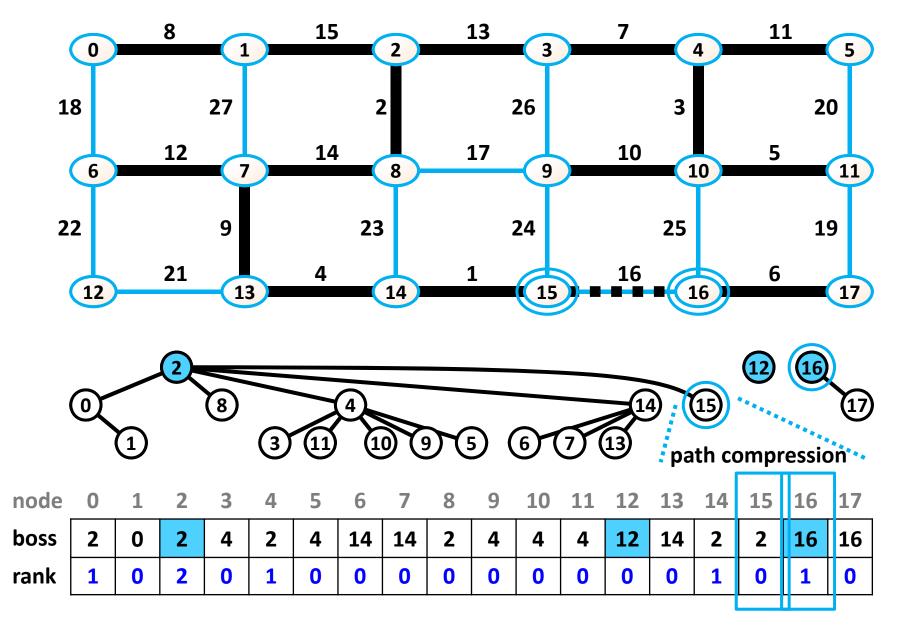


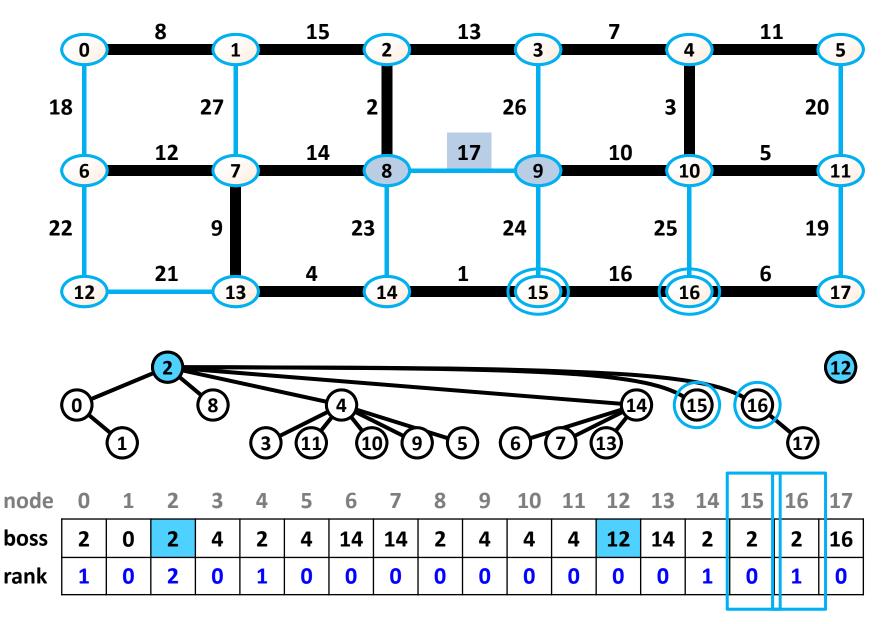


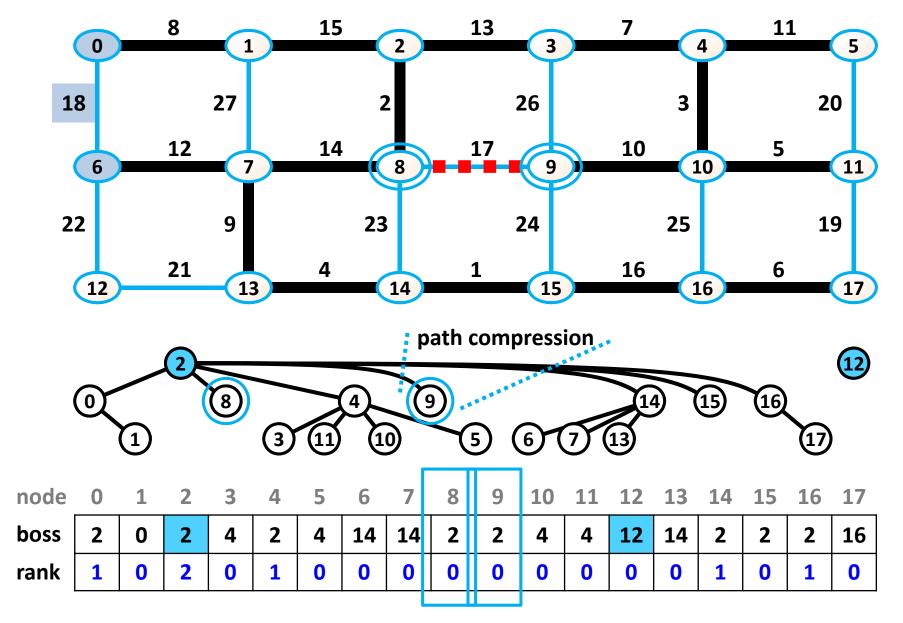


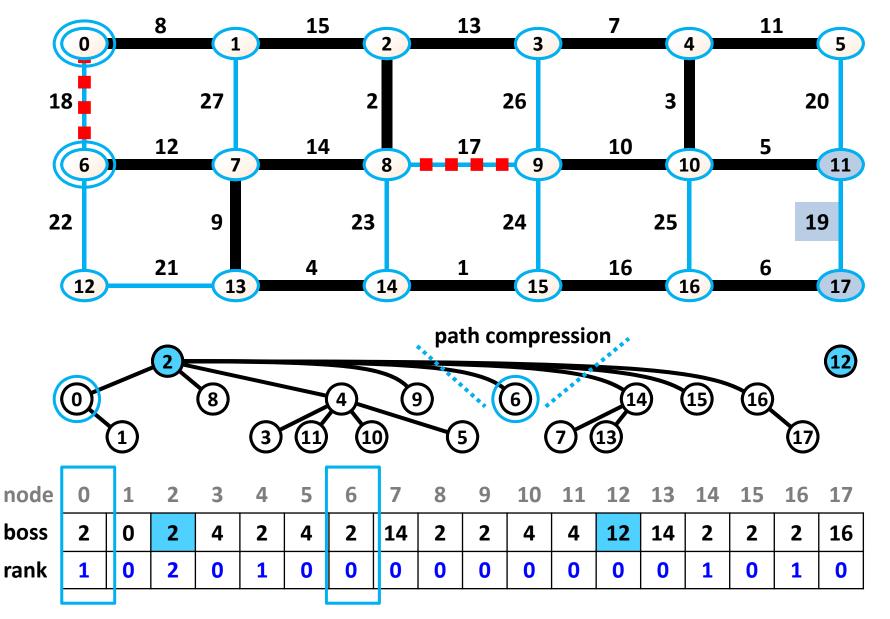


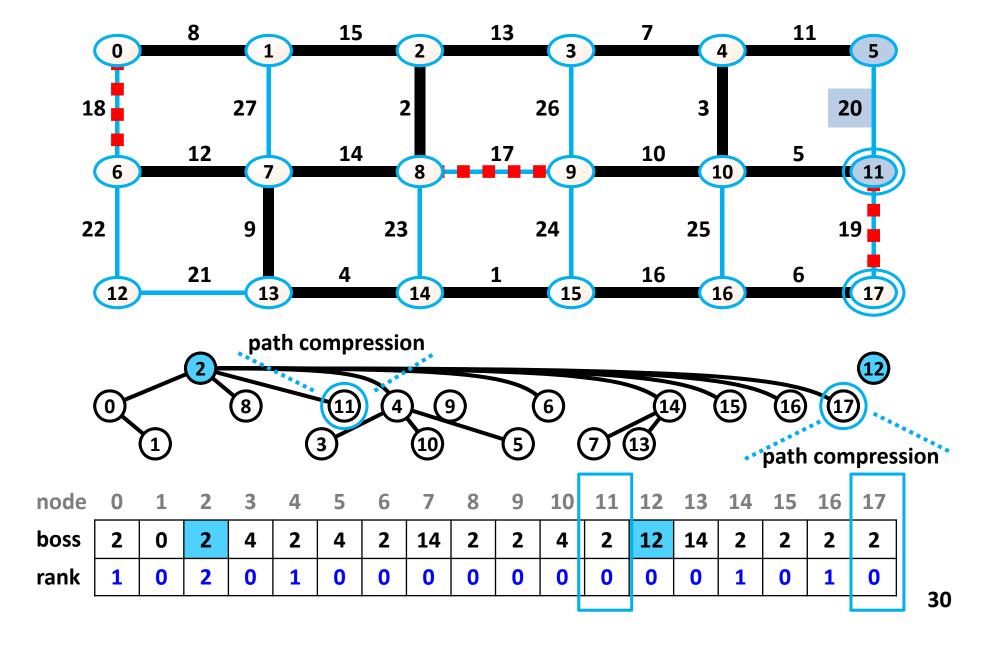


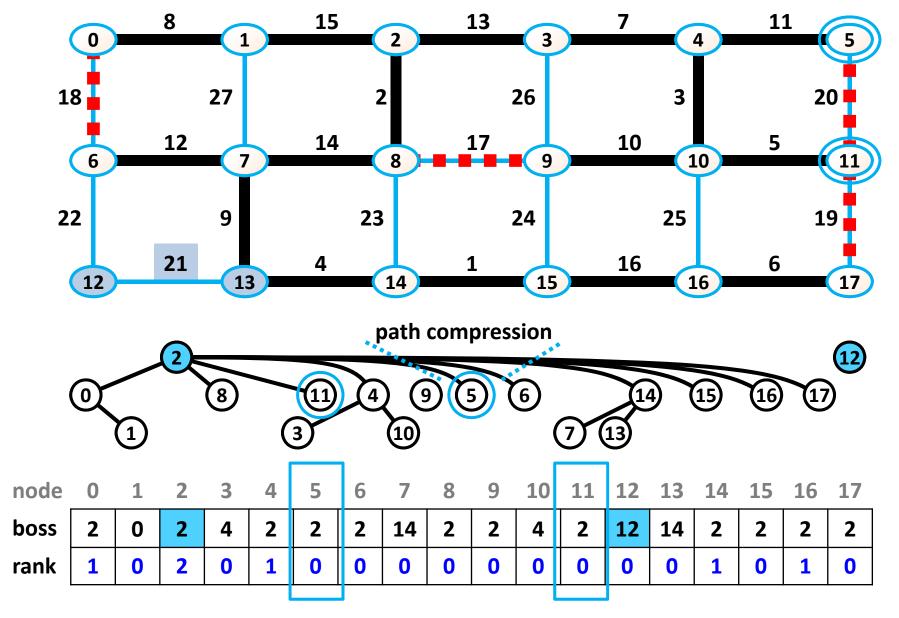


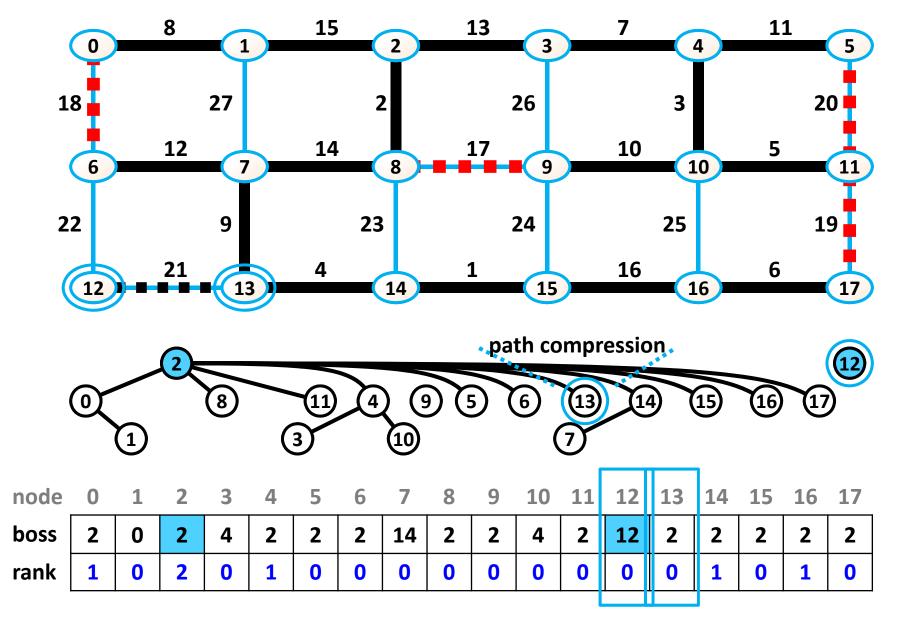


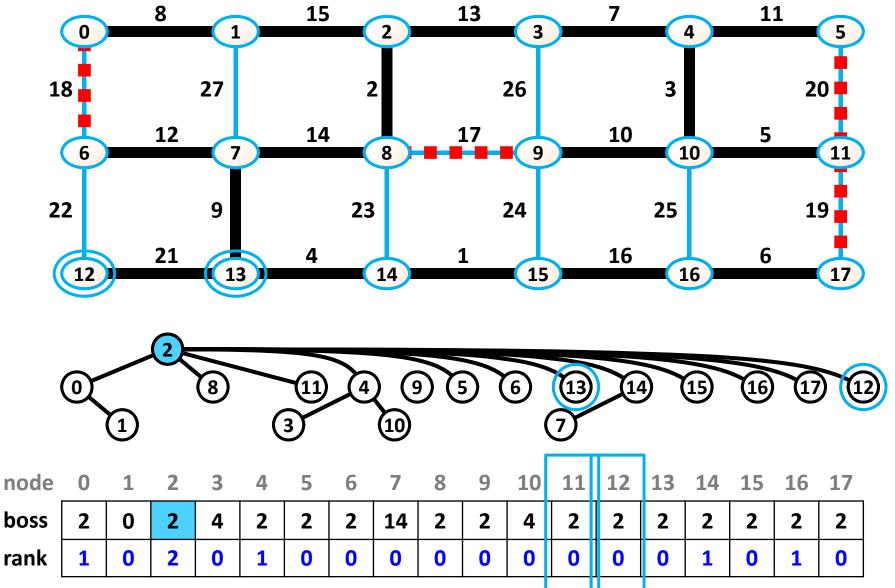












#### Kruskal's algorithm running time

When both union by rank and path compression are used, the running time spent on Union-Find operations in a graph with N nodes and M edges is  $O(M \cdot \alpha(N))$ , where  $\alpha(N)$  is the inverse function of f(x) = A(x, x), where A(x, y), is the Ackermann function, known to grow quite fast. In fact, for any graph representable in any conceivable machine  $\alpha(N) < 4$ .

Thanks to the inverse Ackermann function, all Union-Find operations run in amortized constant time in all practical situations.

It means that the speed bottleneck in Kruskal's algorithm is the initial edge sorting. Sorting can be done in linear time when:

- edge values are strings (does not happen too often), apply Radix sort,
- edge values are integers in some (relatively) moderate range (e.g 0..10 000, etc.),
   apply Counting sort,
- edge values are floats (more or less) uniformly distributed over some interval, apply **Bucket sort**.

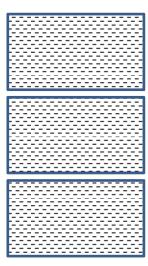
#### Conclusion

In many practical situations, a careful implementation of Kruskal algorithm may run in  $\Theta(M)$  time.

$$A(m,n) = \begin{cases} n+1 & \text{if } m = 0, \\ A(m-1,1) & \text{if } m > 0, n = 0, \\ A(m-1,A(m,n-1)) & \text{if } m > 0, n > 0. \end{cases}$$

$m \setminus n$	0	1	2	3	4	5	n		
0	1	2	3	4	5	6	n+1		
1	2	3	4	5	6	7	n+2		
2	3	5	7	9	11	13	2n - 3		
3	5	$13$ $= 2^4 - 3$	$= 2^5 - 3$ = 29	$= 2^6 - 3$ = 61	$= 2^7 - 3$ = 125	$= 2^8 - 3$ = 253	$2^{n+3}-3$		
4	$   \begin{array}{r}     13 \\     = 2^{2^2} - 3   \end{array} $	$65533 = 2^{2^{2^2}} - 3$		$(##) = 2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2$	$(###)$ $= 2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{3}}}}}}}}}}$	$(####)$ $= 2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2$	$2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2^{2$		
5	$65533 \\ = 2^{2^{2^2}} - 3$	,2	} 65536		22	2	too big to fit here		
6	A(6,0) = $A(5,1)$	22222222	-3	too big to fit here					
				222222222					

- 1. 0-th value in each line = 1st value in the previous line.
- 2. Value X in the j-th column (j > 0) on the current line is equal to the value on the previous line which column index is equal to the value written to left of the value X on the current line (= in the (j-1)-th column).



$$A(4,3) = (\#\#) = 2^{(\#)} - 3 = ?$$

$$A(4,4) = (\#\#\#) = 2^{2^{(\#)}} -3 = ??$$

Informal discussions concerning big integers:

http://waitbutwhy.com/2014/11/1000000-grahams-number.html http://www.scottaaronson.com/writings/bignumbers.html

92718694611585137933864756997485686700798239606043934788508616492603049450617434123658283521448067266768418070837548622114082365798029612000

63129066885434542686978844774298177749371011761465162418361668025481529633530849084994300676365480610294009469375060984558855804397048591444042671493660849802382746805759825913310069199419046519065311719089260779491192179464073551296338645230356733455880333131970803654571847915508439490325290526337532316509087681336614242398309530806549661879381949120033919489494065132398816642080088395554942237096734840072642705701113005282797038580359815182929600305682612091950943737325454171056383887047528950563961029843641360935641632589408137981511693338619797339821 030950313341050704760159987985472529190665222479319715440331794836837373220821885773341623856441380700541913530245943913502554531886454796252260251762928374330465102361057583514550739443339610216229675461415781127197001738611494279501411253280621254775810512972088465263158094806633687670147310733540717710876615935856814098212967730759197382973441445256688770855324570888958320993823432102718224114763732791357568615421005853292720149392350052584514670698262854825788326739873522045722823929020714482221988558710289699193587307427781515975762076402395124386022644889967589828812925480076425186586490241111127301357197181381602583178506932244007998656635371544088454866393181708395735780799059730839094881804060935959190907473960904410150516321749681412100765719177483767355751000733616922386537429079457803200042337452807566153042929014495