4.3 Minimum Spanning Trees

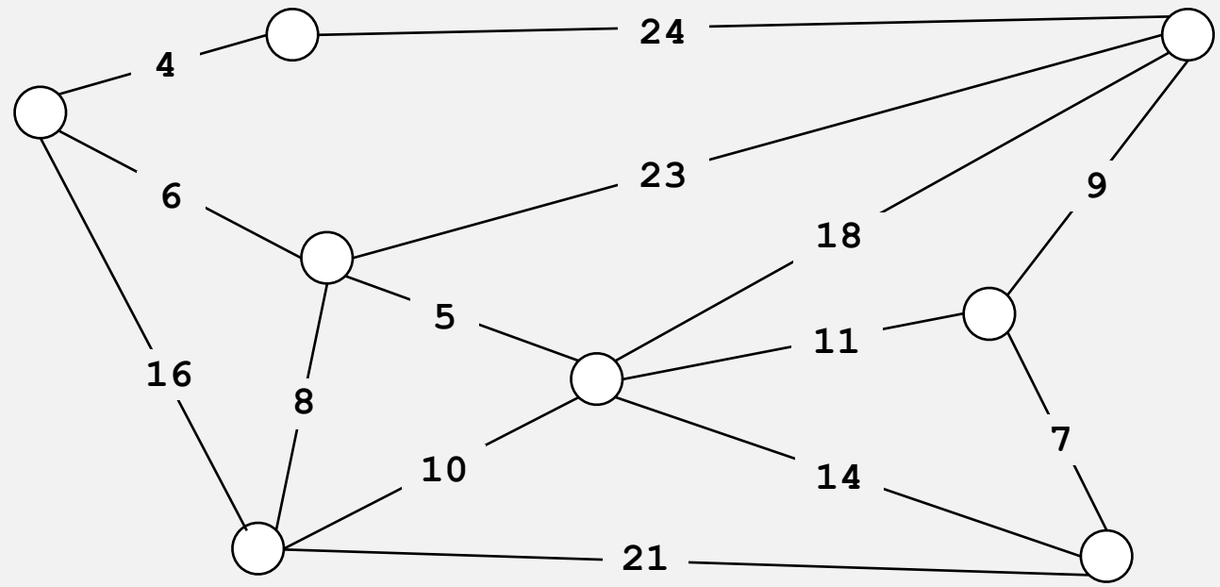
- edge-weighted graph API
- greedy algorithm
- Kruskal's algorithm
- Prim's algorithm
- advanced topics

Minimum spanning tree

Given. Undirected graph G with positive edge weights (connected).

Def. A spanning tree of G is a subgraph T that is connected and acyclic.

Goal. Find a min weight spanning tree.



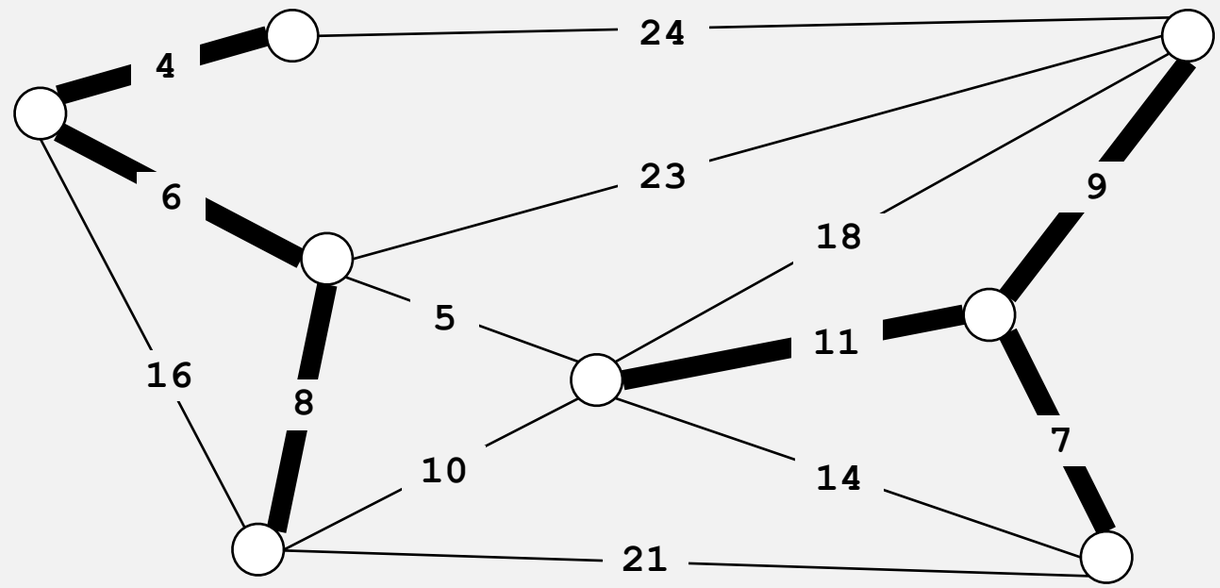
graph G

Minimum spanning tree

Given. Undirected graph G with positive edge weights (connected).

Def. A spanning tree of G is a subgraph T that is connected and acyclic.

Goal. Find a min weight spanning tree.



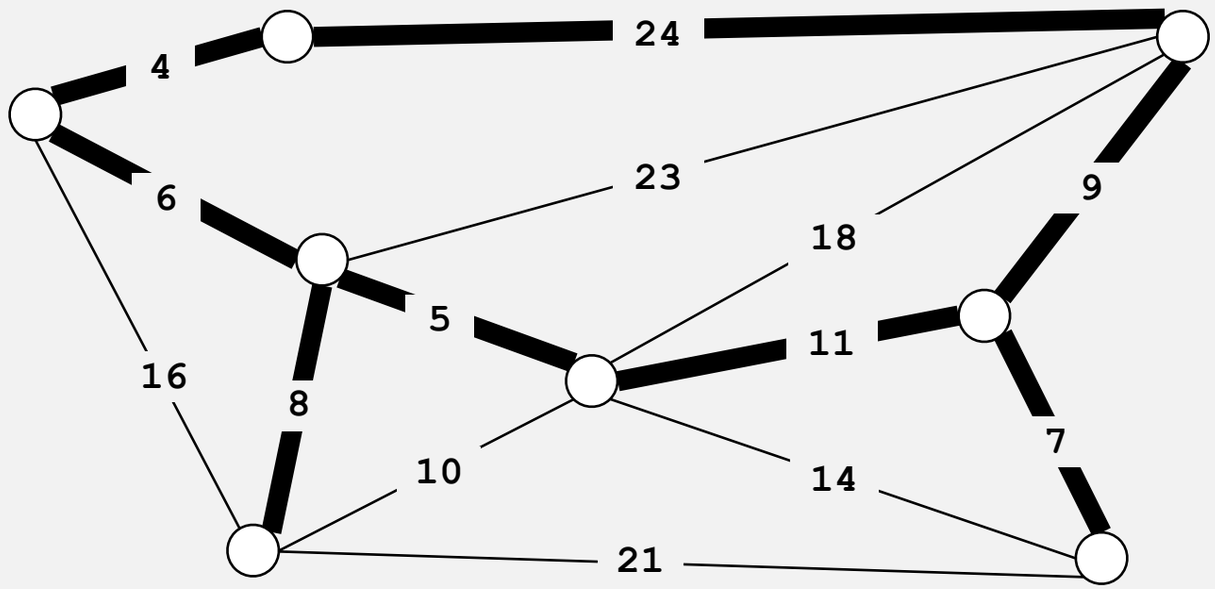
not connected

Minimum spanning tree

Given. Undirected graph G with positive edge weights (connected).

Def. A spanning tree of G is a subgraph T that is connected and acyclic.

Goal. Find a min weight spanning tree.



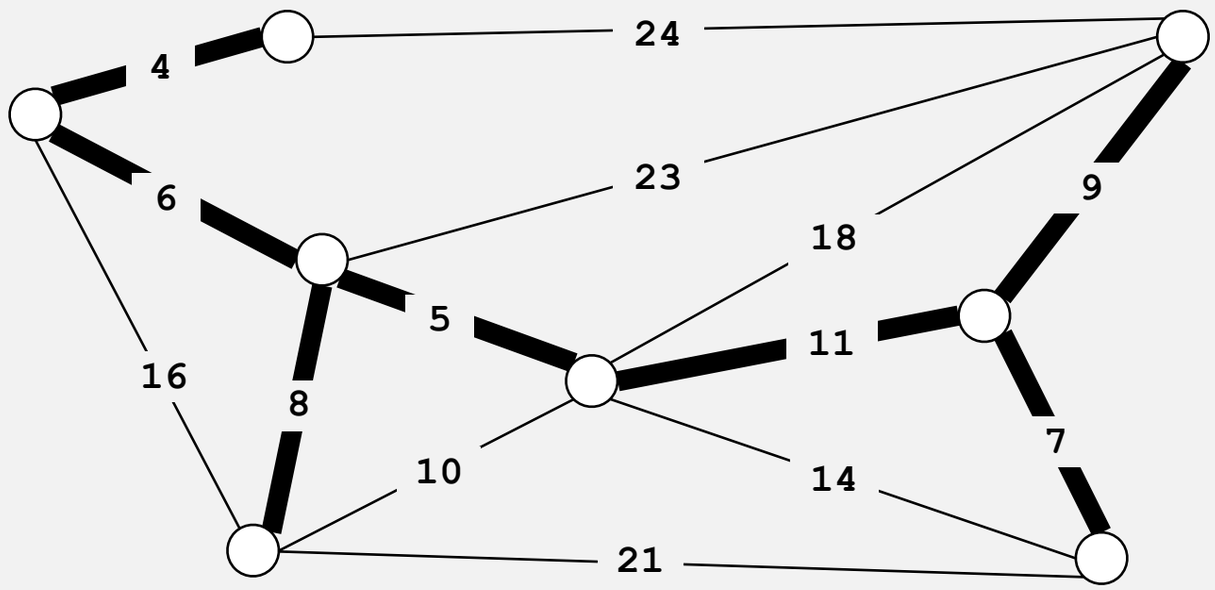
not acyclic

Minimum spanning tree

Given. Undirected graph G with positive edge weights (connected).

Def. A spanning tree of G is a subgraph T that is connected and acyclic.

Goal. Find a min weight spanning tree.



spanning tree T: cost = 50 = 4 + 6 + 8 + 5 + 11 + 9 + 7

Brute force. Try all spanning trees?

Applications

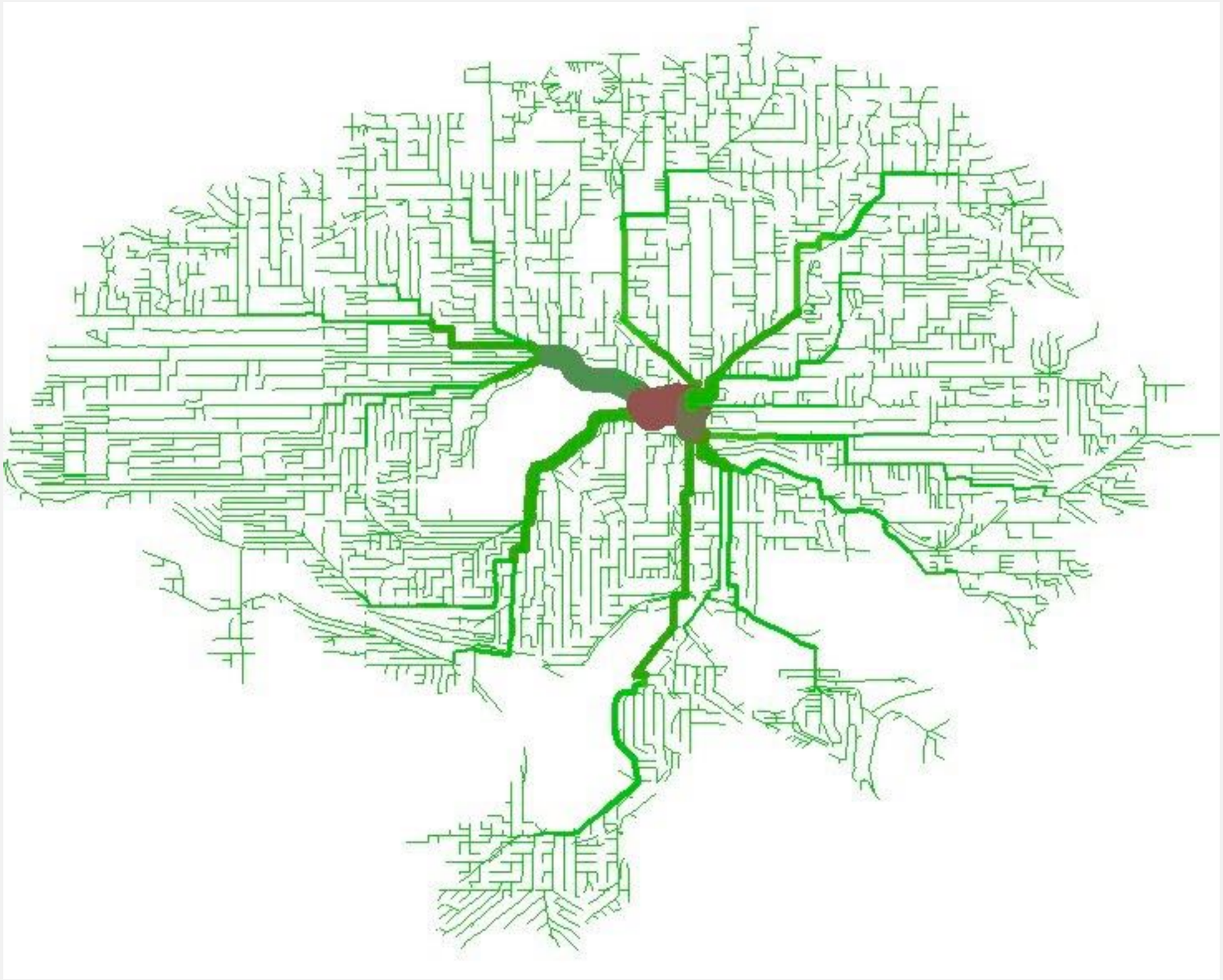
MST is fundamental problem with diverse applications.

- Cluster analysis.
- Max bottleneck paths.
- Real-time face verification.
- LDPC codes for error correction.
- Image registration with Renyi entropy.
- Find road networks in satellite and aerial imagery.
- Reducing data storage in sequencing amino acids in a protein.
- Model locality of particle interactions in turbulent fluid flows.
- Autoconfig protocol for Ethernet bridging to avoid cycles in a network.
- Approximation algorithms for NP-hard problems (e.g., TSP, Steiner tree).
- Network design (communication, electrical, hydraulic, cable, computer, road).

`http://www.ics.uci.edu/~eppstein/gina/mst.html`

Network design

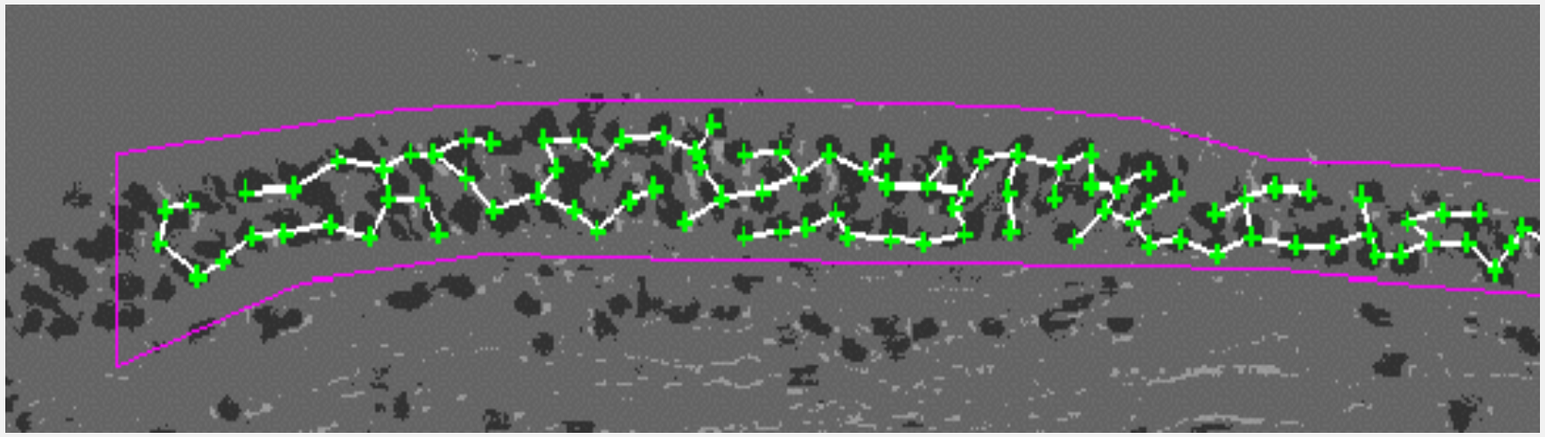
MST of bicycle routes in North Seattle

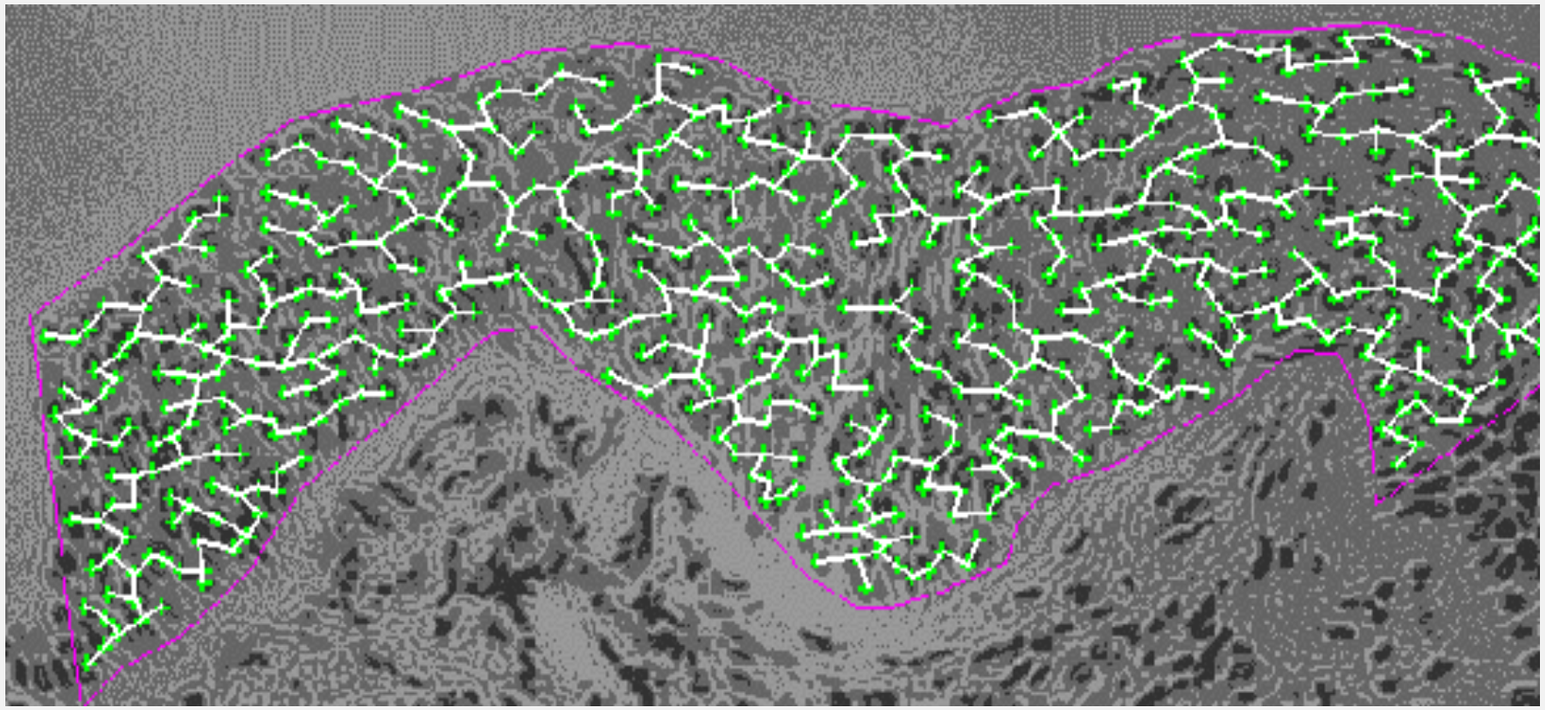


http://www.flickr.com/photos/ewedistrict/21980840

Medical image processing

MST describes arrangement of nuclei in the epithelium for cancer research

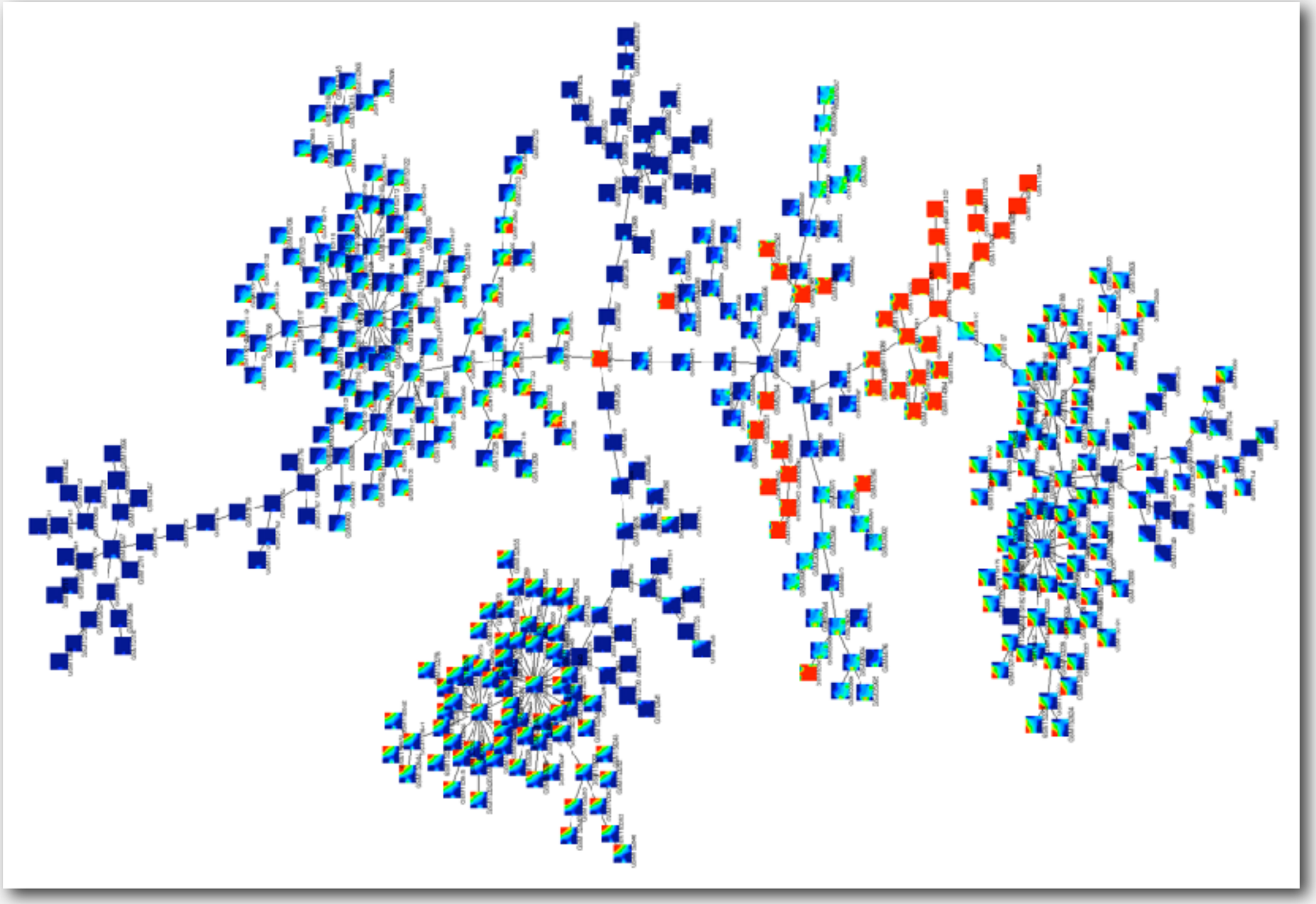




http://www.bccrc.ca/ci/ta01_archlevel.html

Genetic research

MST of tissue relationships measured by gene expression correlation coefficient



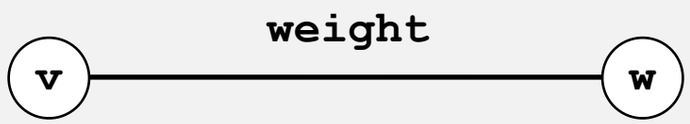
http://riodb.ibase.aist.go.jp/CELLPEDIA

- **edge-weighted graph API**
- greedy algorithm
- Kruskal's algorithm
- Prim's algorithm
- advanced topics

Weighted edge API

Edge abstraction needed for weighted edges.

`public class Edge implements Comparable<Edge>`	
`Edge(int v, int w, double weight)`	*create a weighted edge v-w*
`int either()`	*either endpoint*
`int other(int v)`	*the endpoint that's not v*
`int compareTo(Edge that)`	*compare this edge to that edge*
`double weight()`	*the weight*
`String toString()`	*string representation*



Idiom for processing an edge `e`: `int v = e.either(), w = e.other(v);`

Weighted edge: Java implementation

```
public class Edge implements Comparable<Edge>
{
   private final int v, w;
   private final double weight;

   public Edge(int v, int w, double weight)
   {
      this.v = v;
      this.w = w;
      this.weight = weight;
   }

   public int either()
   {  return v;  }

   public int other(int vertex)
   {
      if (vertex == v) return w;
      else return v;
   }

   public int compareTo(Edge that)
   {
      if      (this.weight < that.weight) return -1;
      else if (this.weight > that.weight) return +1;
      else                                return  0;
   }
}
```

constructor

either endpoint

other endpoint

compare edges by weight

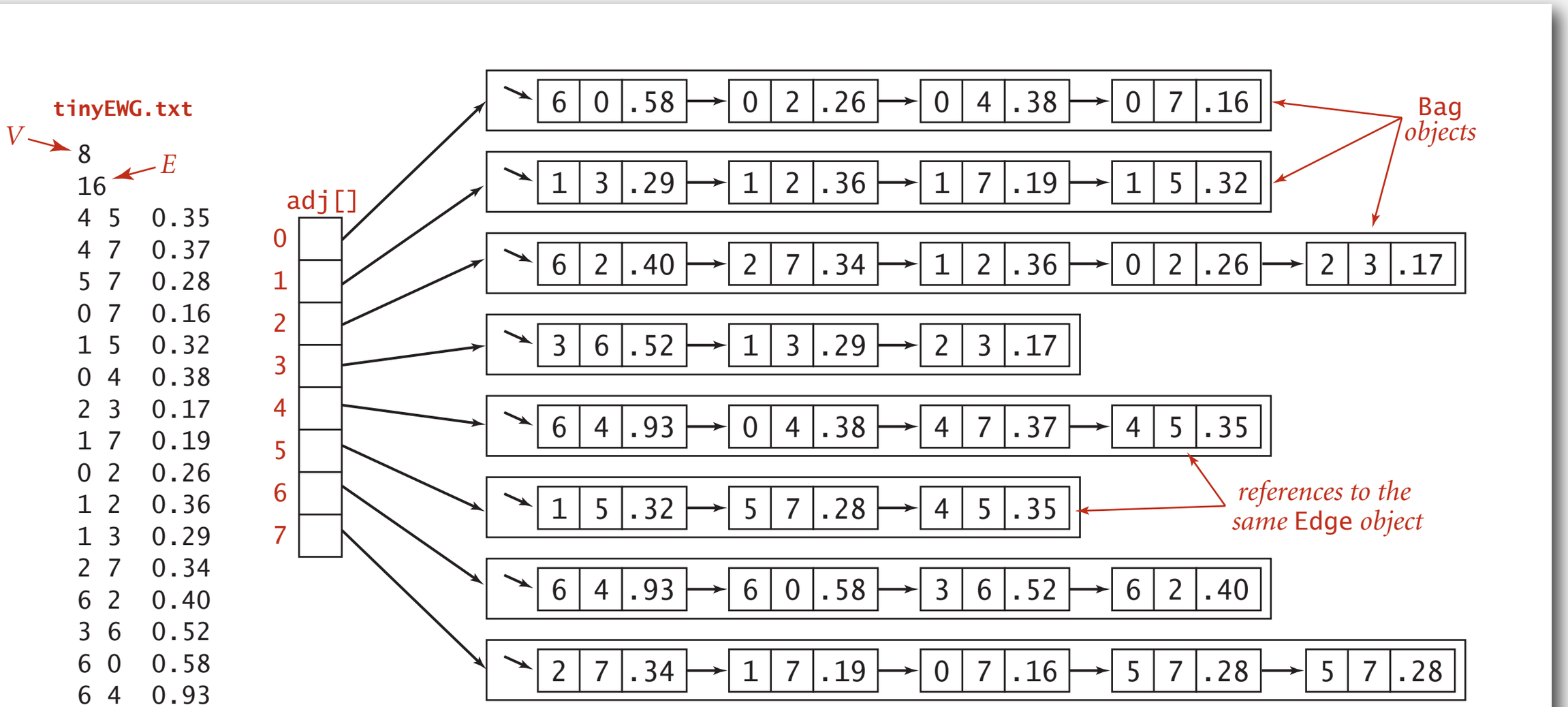
Edge-weighted graph API

public class EdgeWeightedGraph		
	`EdgeWeightedGraph(int V)`	*create an empty graph with V vertices*
	`EdgeWeightedGraph(In in)`	*create a graph from input stream*
`void`	`addEdge(Edge e)`	*add weighted edge e*
`Iterable<Edge>`	`adj(int v)`	*edges incident to v*
`Iterable<Edge>`	`edges()`	*all of this graph's edges*
`int`	`V()`	*return number of vertices*
`int`	`E()`	*return number of edges*
`String`	`toString()`	*string representation*

Conventions. Allow self-loops and parallel edges.

Edge-weighted graph: adjacency-list representation

Maintain vertex-indexed array of `Edge` lists (use `Bag` abstraction).



Edge-weighted graph: adjacency-lists implementation

```
public class EdgeWeightedGraph
{
   private final int V;
   private final Bag<Edge>[] adj;

   public EdgeWeightedGraph(int V)
   {
      this.V = V;
      adj = (Bag<Edge>[]) new Bag[V];
      for (int v = 0; v < V; v++)
         adj[v] = new Bag<Edge>();
   }

   public void addEdge(Edge e)
   {
      int v = e.either(), w = e.other(v);
      adj[v].add(e);
      adj[w].add(e);
   }

   public Iterable<Edge> adj(int v)
   {  return adj[v];  }
}
```

same as **`Graph`**, but adjacency lists of **`Edge`**s instead of integers

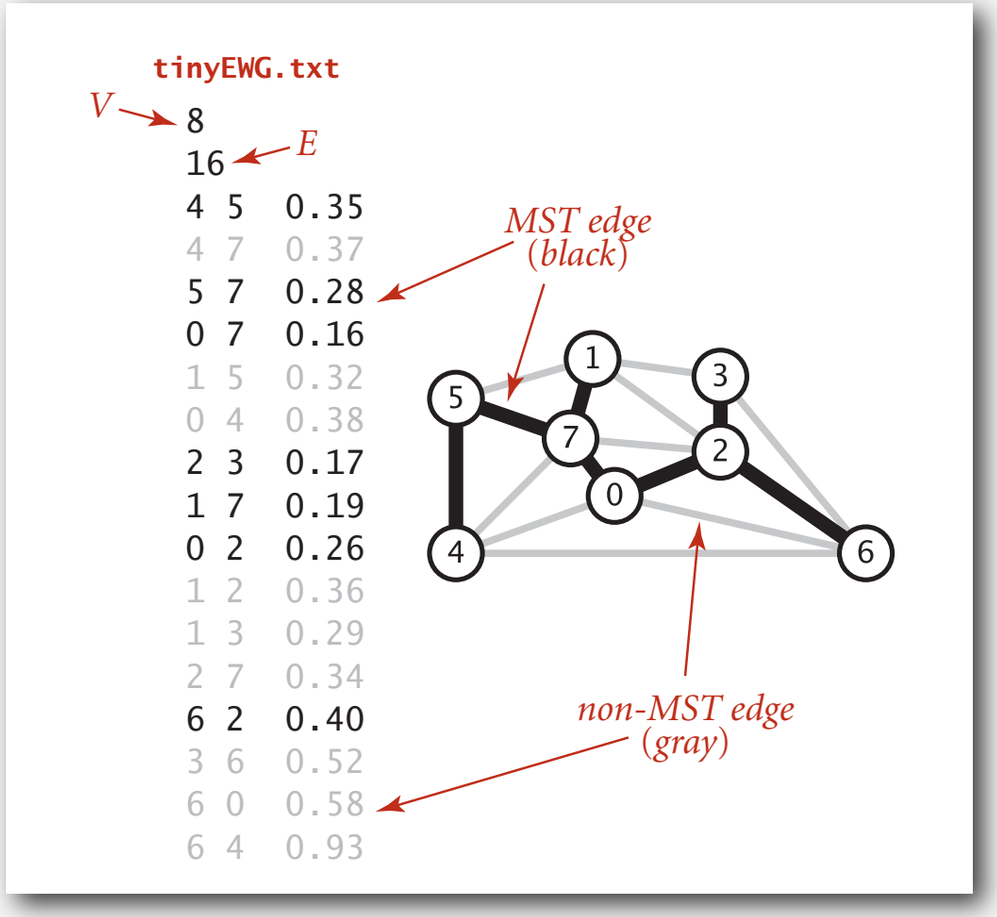
constructor

add edge to both adjacency lists

Minimum spanning tree API

Q. How to represent the MST?

public class MST	
	`MST(EdgeWeightedGraph G)` *constructor*
`Iterable<Edge>`	`edges()` *edges in MST*
`double`	`weight()` *weight of MST*




```
% java MST tinyEWG.txt
0-7 0.16
1-7 0.19
0-2 0.26
2-3 0.17
5-7 0.28
4-5 0.35
6-2 0.40
1.81
```

Minimum spanning tree API

Q. How to represent the MST?

public class MST	
	`MST(EdgeWeightedGraph G)` *constructor*
`Iterable<Edge>`	`edges()` *edges in MST*
`double`	`weight()` *weight of MST*

```
public static void main(String[] args)
{
   In in = new In(args[0]);
   EdgeWeightedGraph G = new EdgeWeightedGraph(in);
   MST mst = new MST(G);
   for (Edge e : mst.edges())
      StdOut.println(e);
   StdOut.println(mst.weight());
}
```

```
% java MST tinyEWG.txt
0-7 0.16
1-7 0.19
0-2 0.26
2-3 0.17
5-7 0.28
4-5 0.35
6-2 0.40
1.81
```

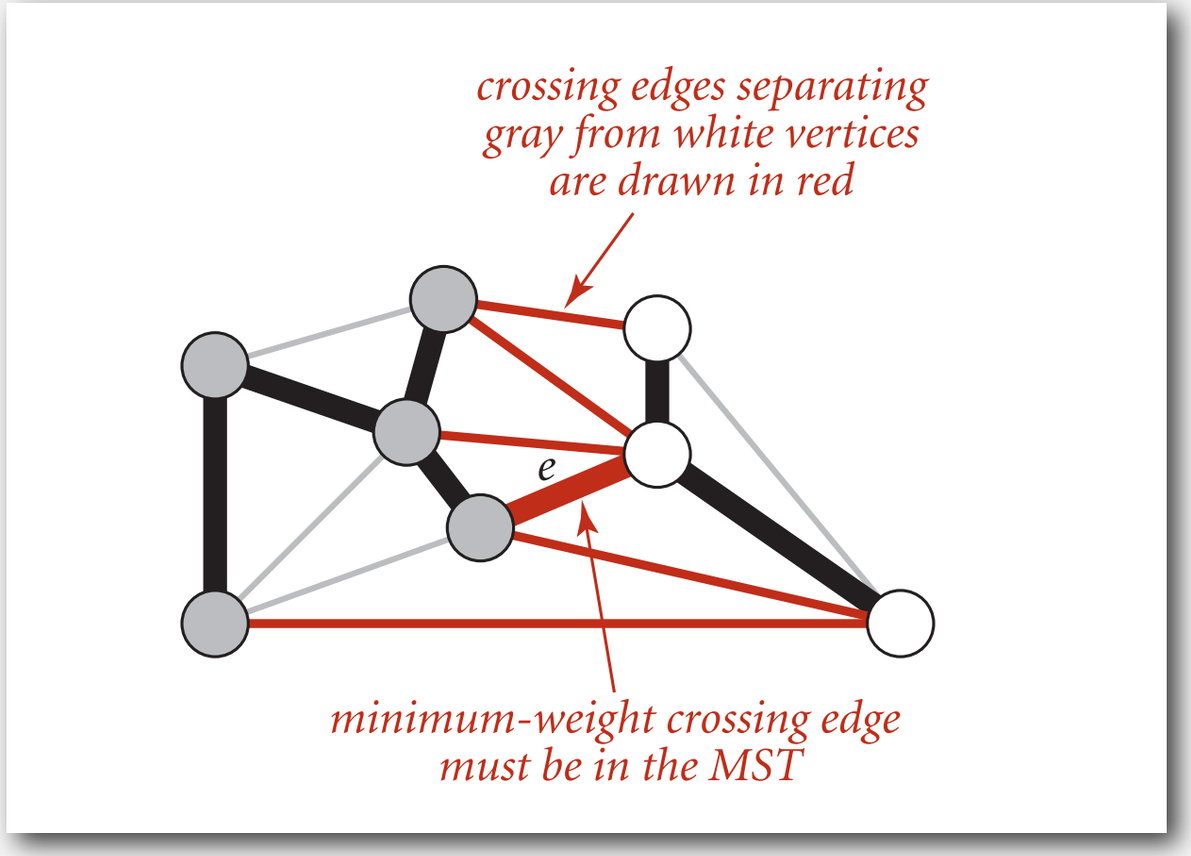
- edge-weighted graph API
- **greedy algorithm**
- Kruskal's algorithm
- Prim's algorithm
- advanced topics

Cut property

Simplifying assumptions. Edge weights are distinct; graph is connected.

Def. A cut in a graph is a partition of its vertices into two (nonempty) sets.
A crossing edge connects a vertex in one set with a vertex in the other.

Cut property. Given any cut, the crossing edge of min weight is in the MST.



Cut property: correctness proof

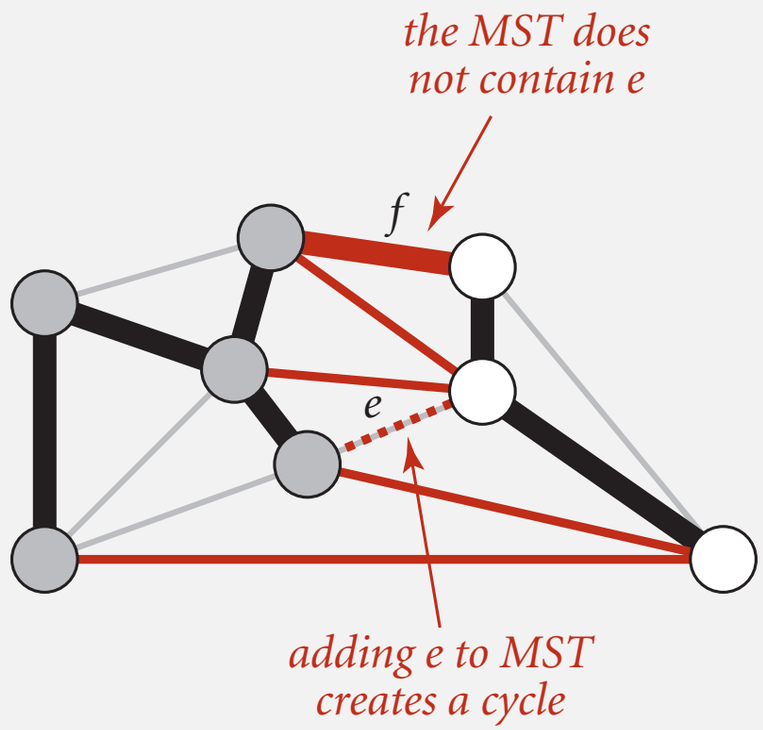
Simplifying assumptions. Edge weights are distinct; graph is connected.

Def. A cut in a graph is a partition of its vertices into two (nonempty) sets.
A crossing edge connects a vertex in one set with a vertex in the other.

Cut property. Given any cut, the crossing edge of min weight is in the MST.

Pf. Let e be the min-weight crossing edge in cut.

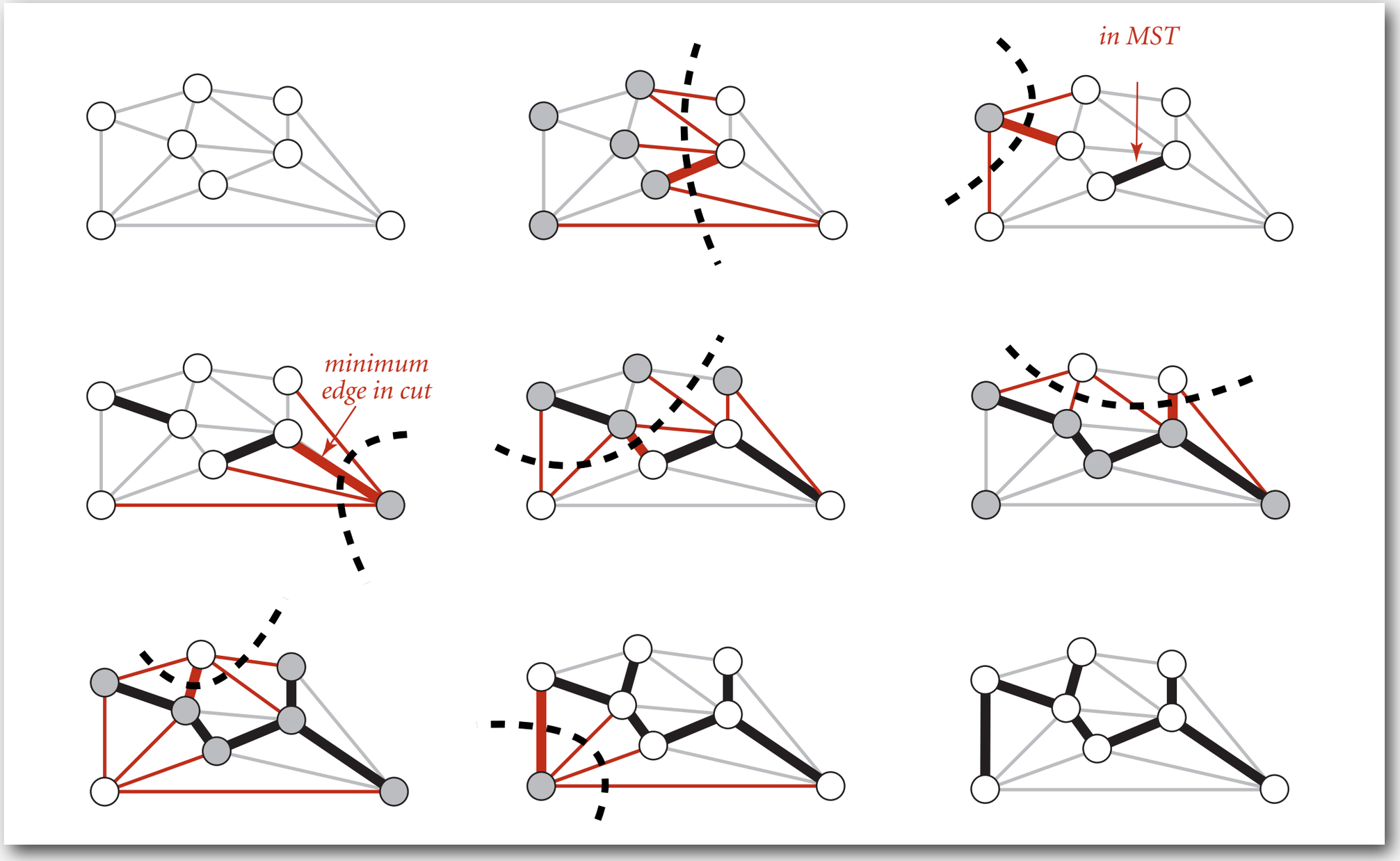
- Suppose e is not in the MST.
- Adding e to the MST creates a cycle.
- Some other edge f in cycle must be a crossing edge.
- Removing f and adding e is also a spanning tree.
- Since weight of e is less than the weight of f, that spanning tree is lower weight.
- Contradiction. ▪



Greedy MST algorithm

Proposition. The following algorithm computes the MST:

- Start with all edges colored gray.
- Find a cut with no black crossing edges, and color its min-weight edge black.
- Continue until $V - 1$ edges are colored black.



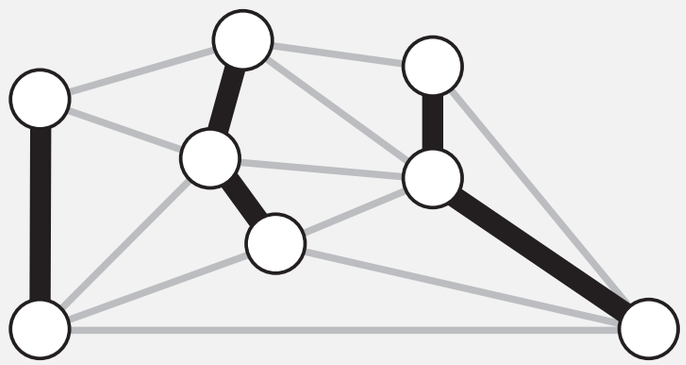
Greedy MST algorithm

Proposition. The following algorithm computes the MST:

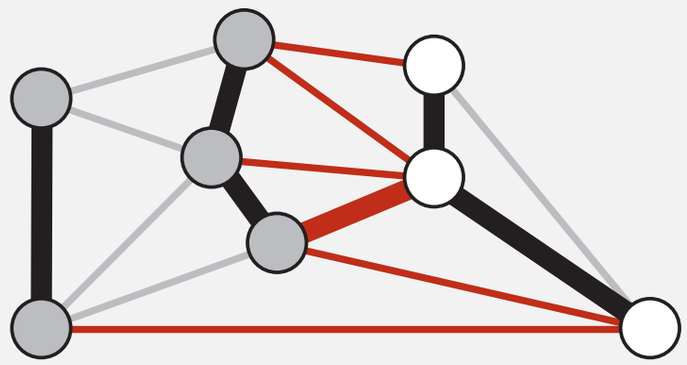
- Start with all edges colored gray.
- Find a cut with no black crossing edges, and color its min-weight edge black.
- Continue until $V - 1$ edges are colored black.

Pf.

- Any edge colored black is in the MST (via cut property).
- If fewer than $V - 1$ black edges, there exists a cut with no black crossing edges. (consider cut whose vertices are one connected component)



fewer than V-1 edges colored black



a cut with no black crossing edges

Greedy MST algorithm

Proposition. The following algorithm computes the MST:

- Start with all edges colored gray.
- Find a cut with no black crossing edges, and color its min-weight edge black.
- Continue until $V - 1$ edges are colored black.

Efficient implementations. How to choose cut? How to find min-weight edge?

Ex 1. Kruskal's algorithm. [stay tuned]

Ex 2. Prim's algorithm. [stay tuned]

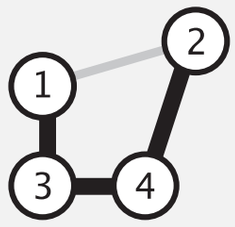
Ex 3. Borüvka's algorithm.

Removing two simplifying assumptions

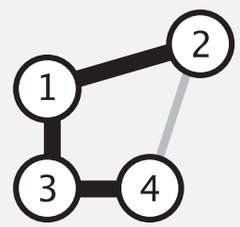
Q. What if edge weights are not all distinct?

A. Greedy MST algorithm still correct if equal weights are present!

(our correctness proof fails, but that can be fixed)



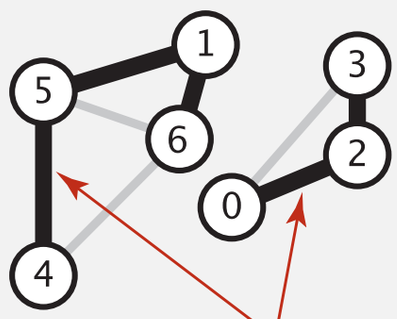

```
1 2  1.00
1 3  0.50
2 4  1.00
3 4  0.50
```




```
1 2  1.00
1 3  0.50
2 4  1.00
3 4  0.50
```

Q. What if graph is not connected?

A. Compute minimum spanning forest = MST of each component.



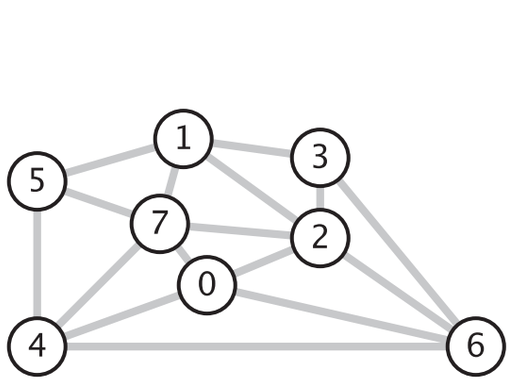

```
4 5  0.61
4 6  0.62
5 6  0.88
1 5  0.11
2 3  0.35
0 3  0.6
1 6  0.10
0 2  0.22
```

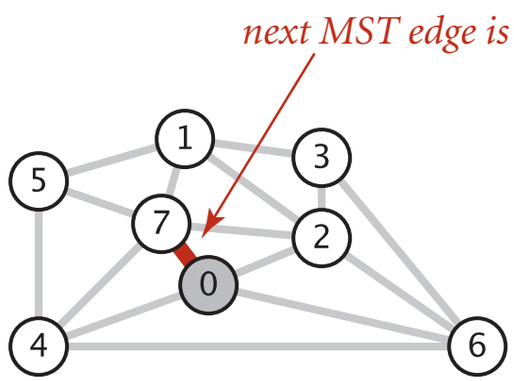
- edge-weighted graph API
- greedy algorithm
- **Kruskal's algorithm**
- Prim's algorithm
- advanced topics

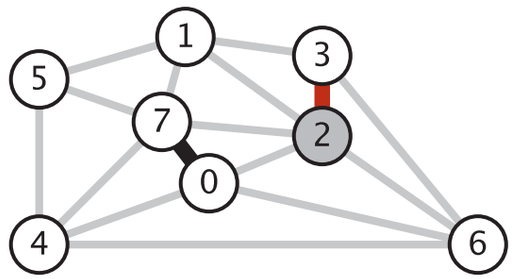
Kruskal's algorithm

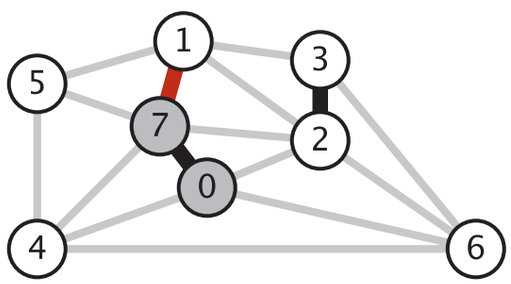
Kruskal's algorithm. [Kruskal 1956] Consider edges in ascending order of weight.

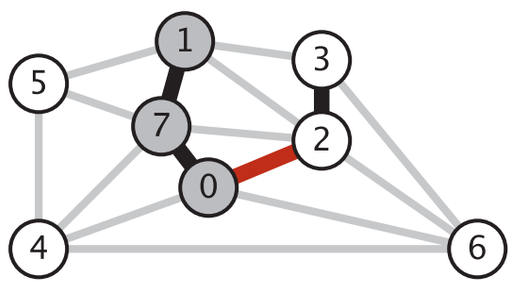
Add the next edge to the tree T unless doing so would create a cycle.

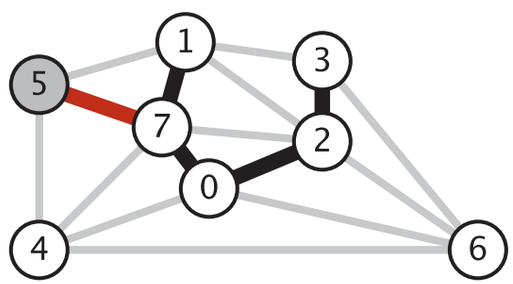


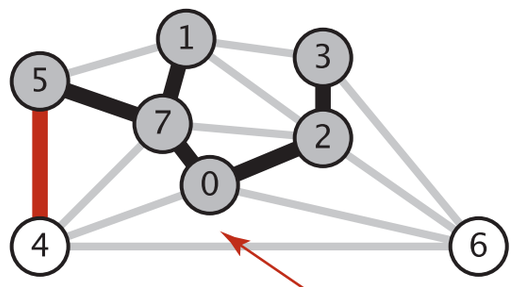


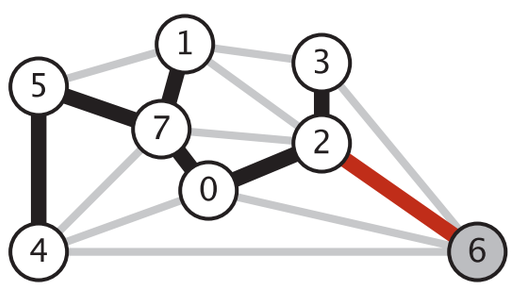


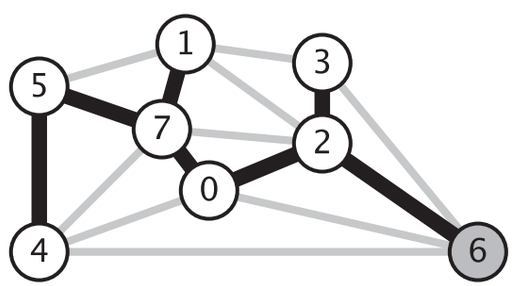












grey vertices are a cut defined by the vertices connected to one of the red edge's vertices

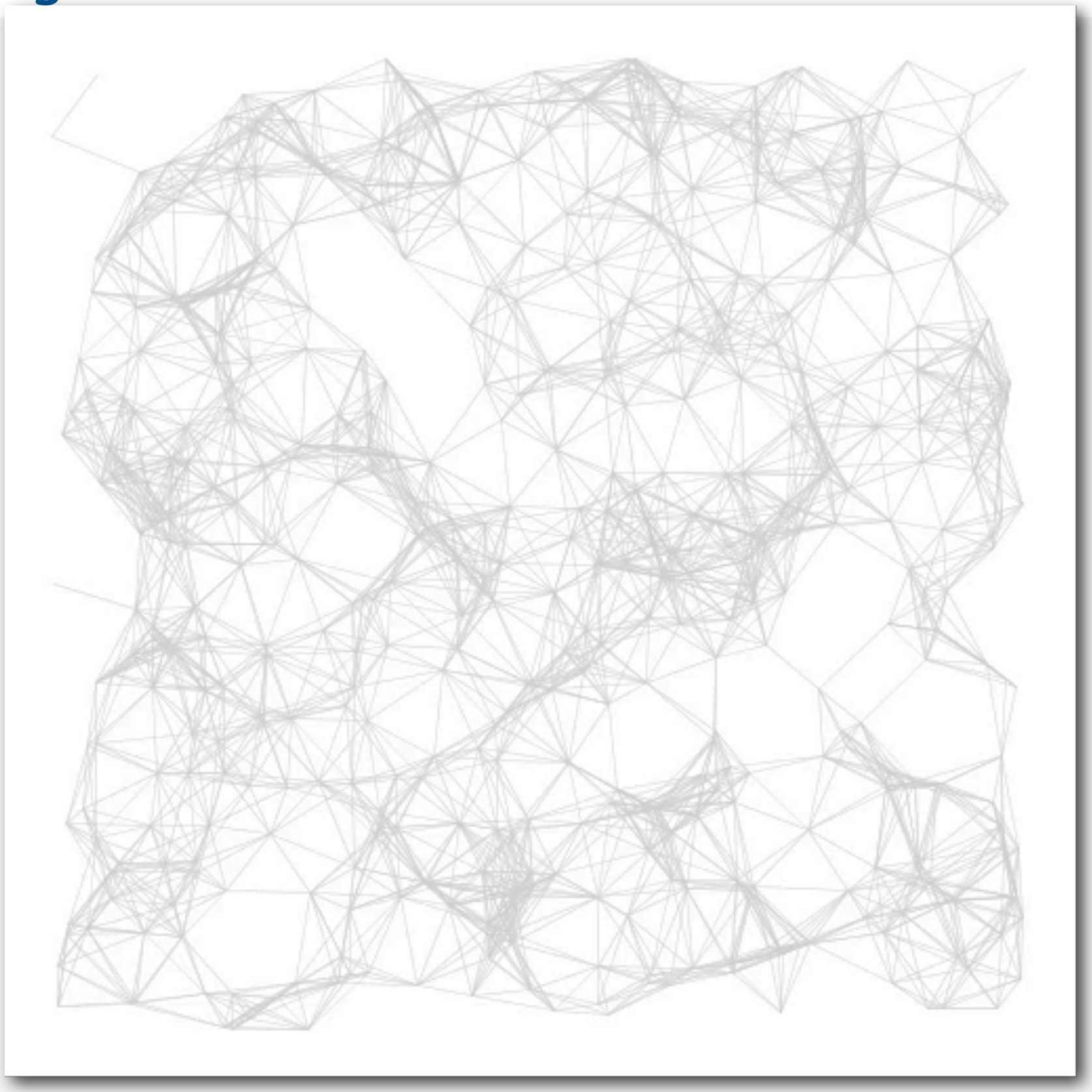
graph edges sorted by weight

MST edge (black)

```
0-7  0.16
2-3  0.17
1-7  0.19
0-2  0.26
5-7  0.28
1-3  0.29
1-5  0.32
2-7  0.34
4-5  0.35
1-2  0.36
4-7  0.37
0-4  0.38
6-2  0.40
3-6  0.52
6-0  0.58
6-4  0.93
```

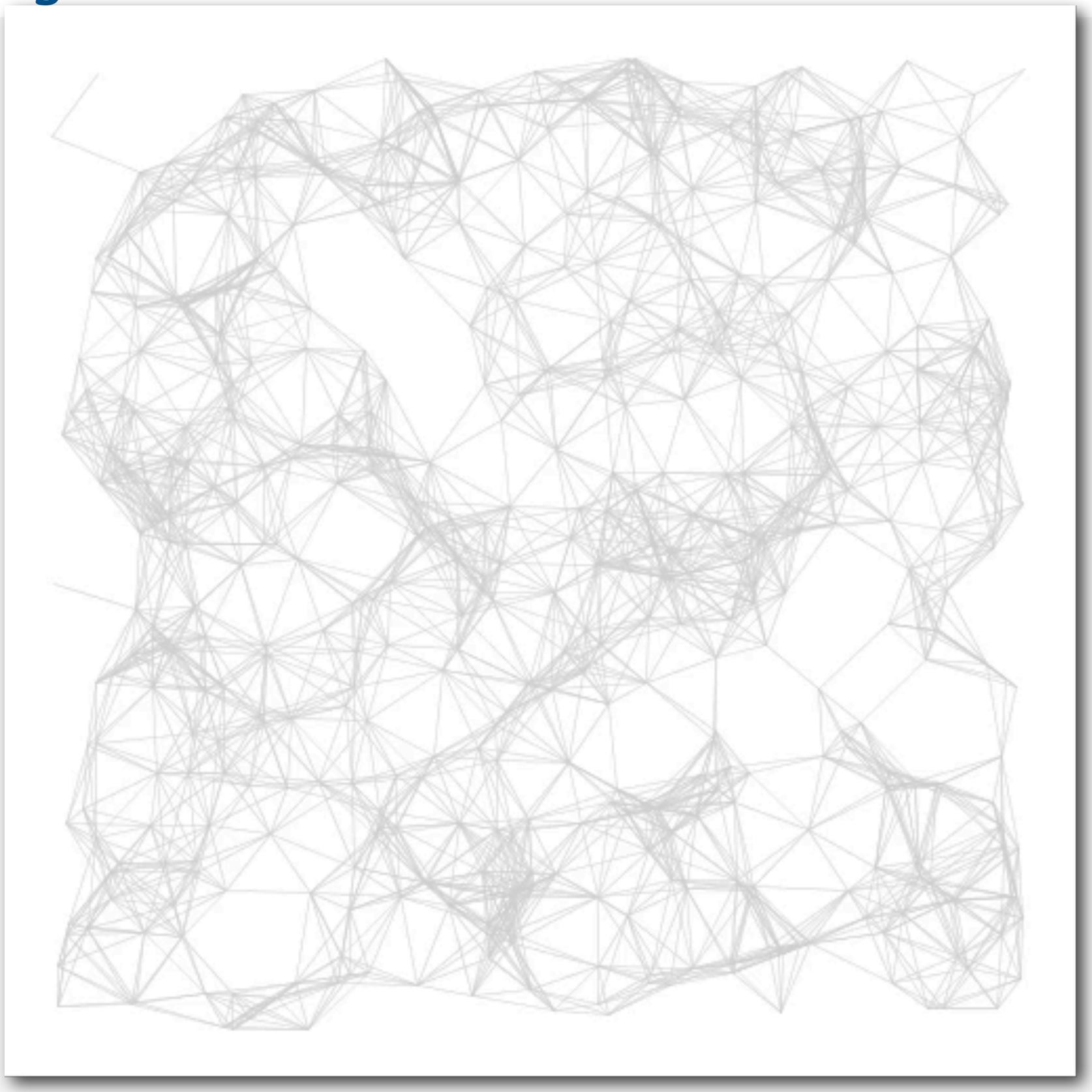
obsolete edge (gray)

Kruskal's algorithm visualization



https://www.cs.purdue.edu/homes/cs251/slides/media/Kruskal500.mov

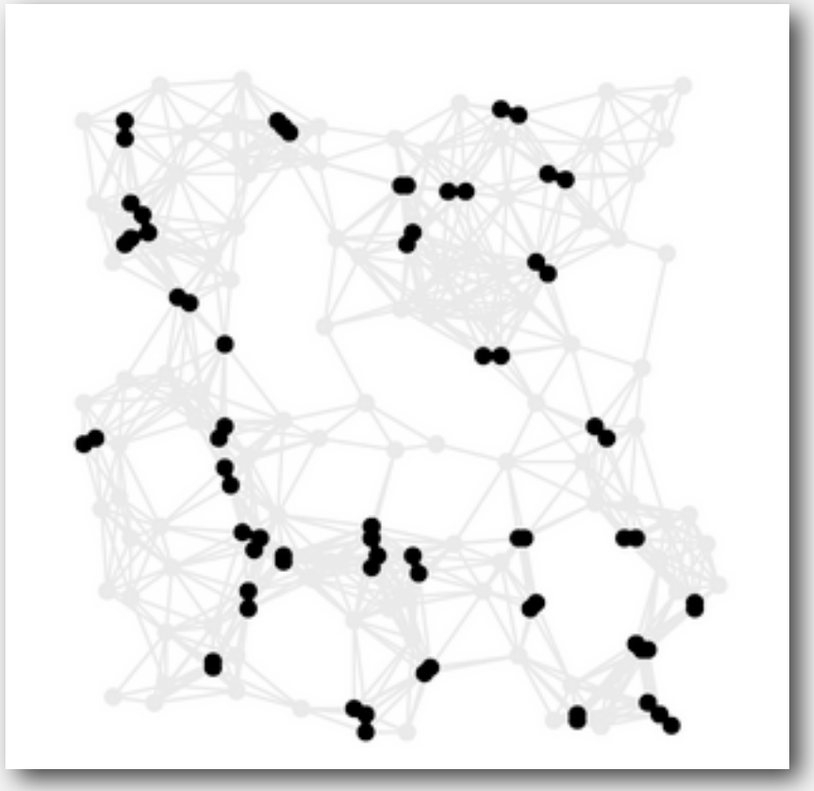
Kruskal's algorithm visualization

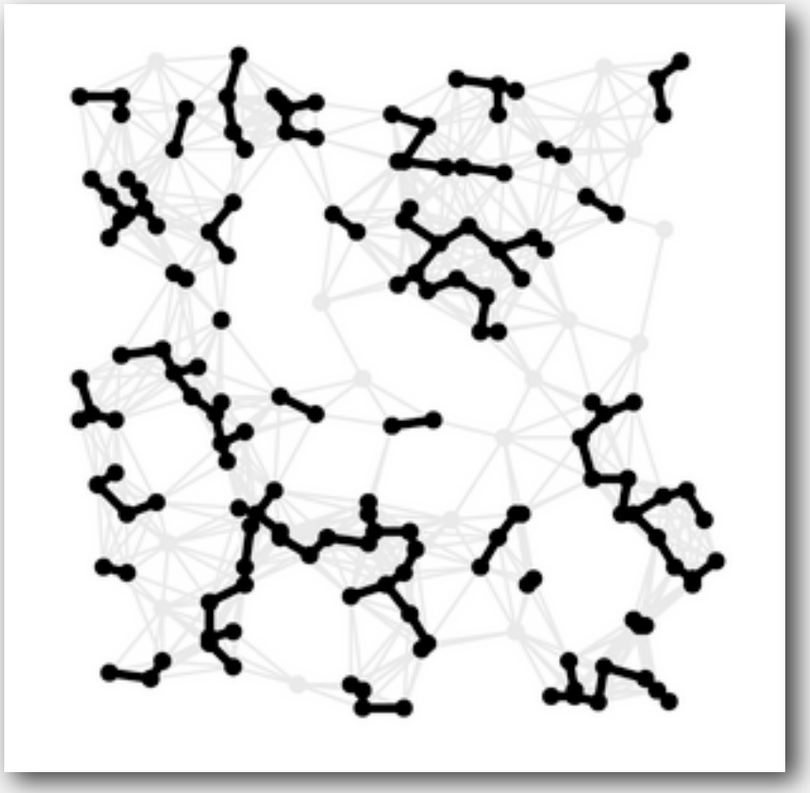


https://www.cs.purdue.edu/homes/cs251/slides/media/Kruskal500.mov

Kruskal's algorithm visualization

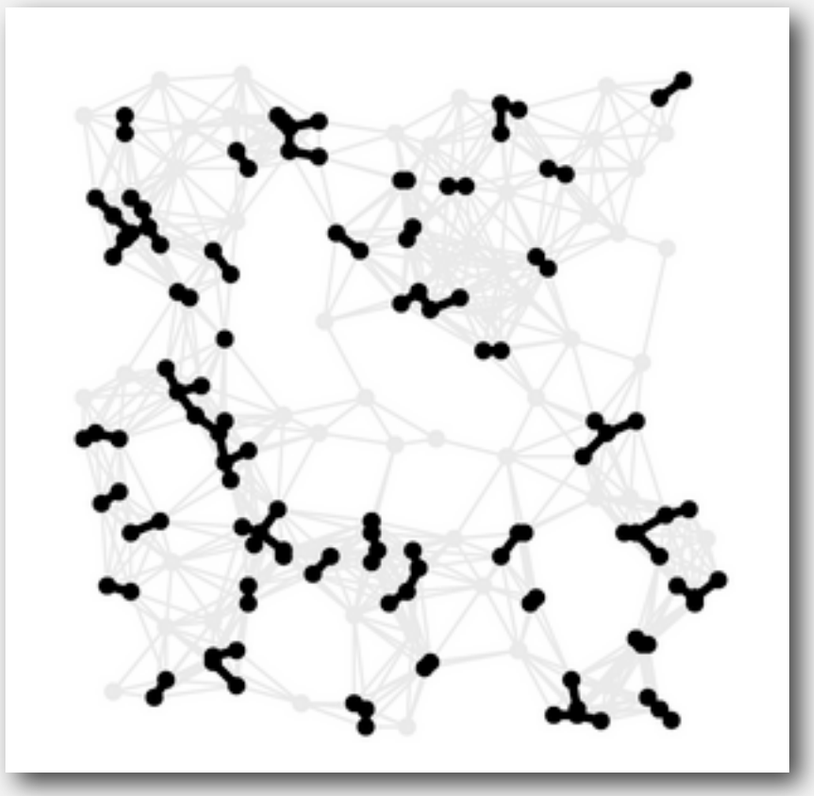
25%

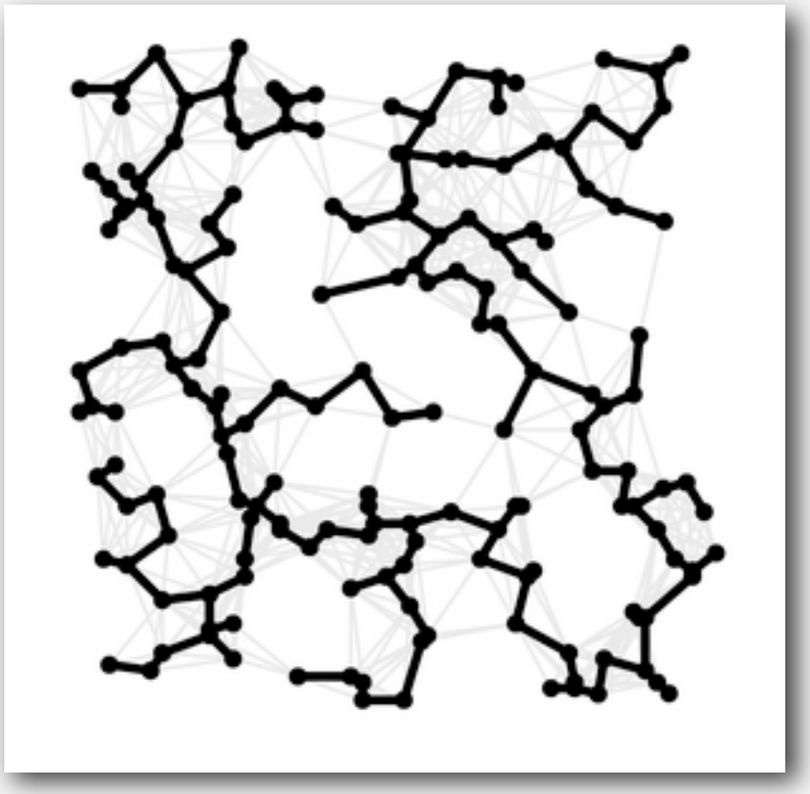




75%

50%





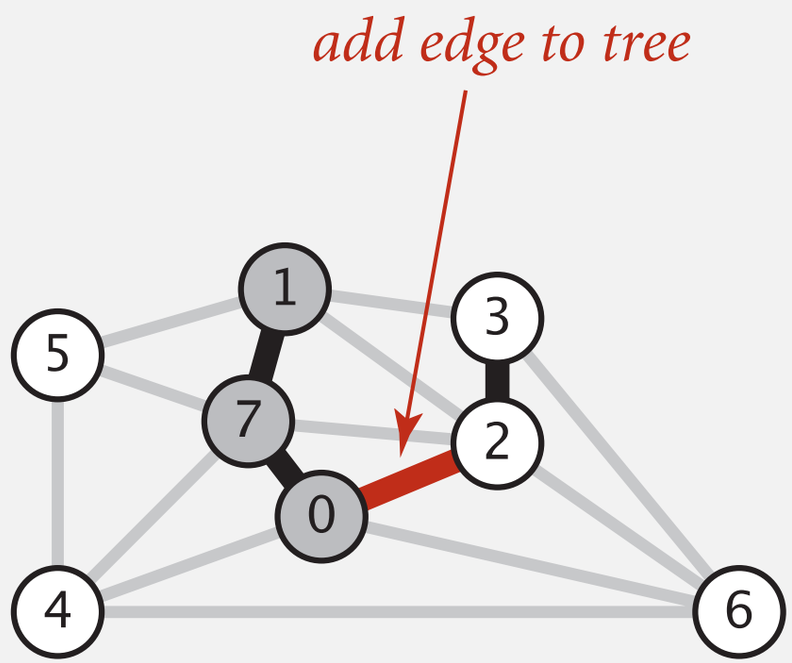
100%

Kruskal's algorithm: proof of correctness

Proposition. Kruskal's algorithm computes the MST.

Pf. Kruskal's algorithm is a special case of the greedy MST algorithm.

- Suppose Kruskal's algorithm colors edge $e = v$–w black.
- Cut = set of vertices connected to v (or to w) in tree T.
- No crossing edge is black.
- No crossing edge has lower weight. Why?

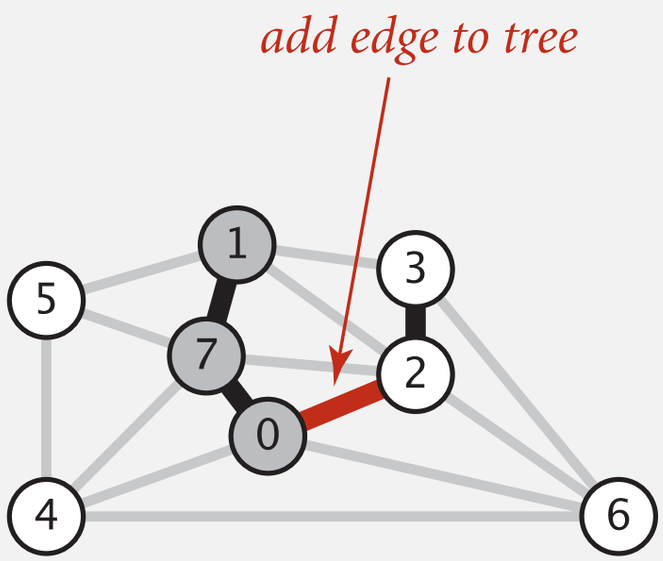


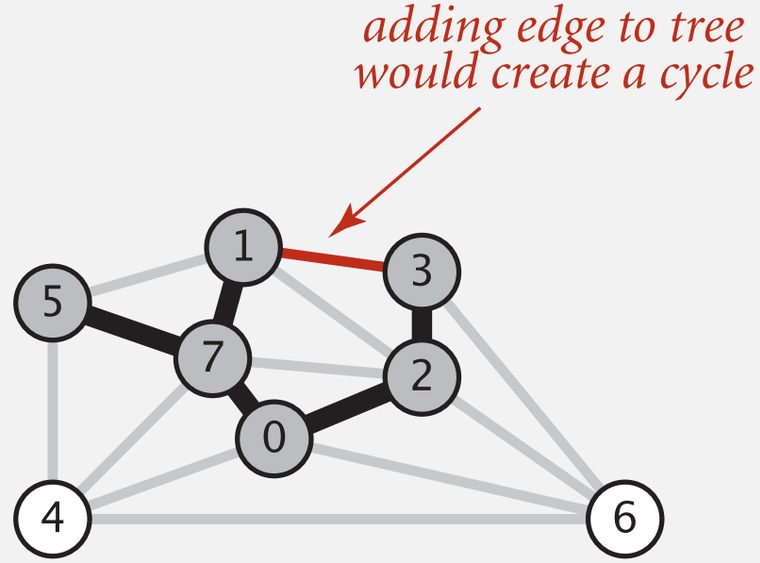
Kruskal's algorithm: implementation challenge

Challenge. Would adding edge v–w to tree T create a cycle? If not, add it.

How difficult?

- $O(E + V)$ time.
- $O(V)$ time.
- $O(\log V)$ time.
- $O(\log^* V)$ time.
- Constant time.



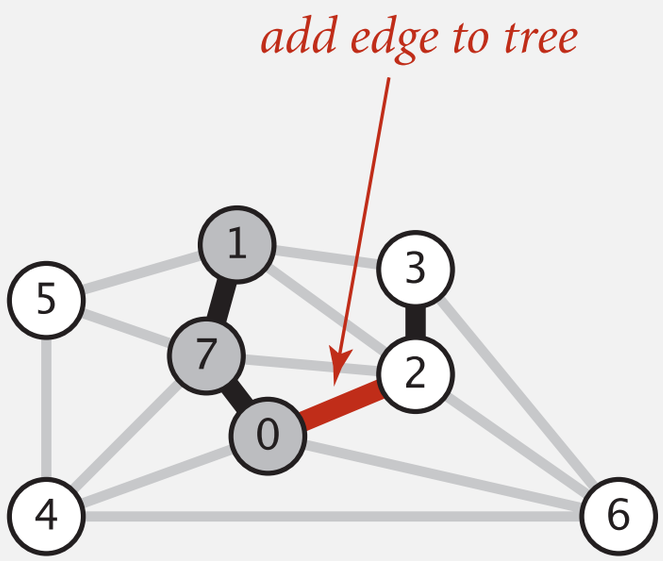


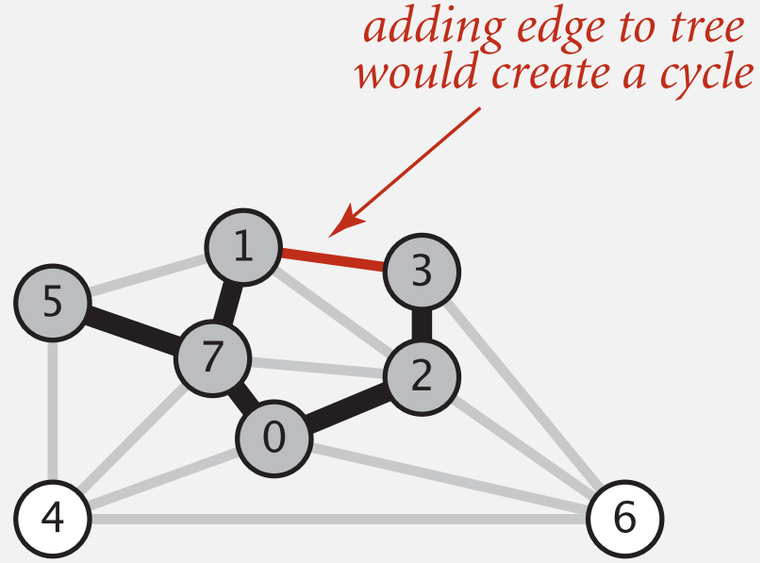
Kruskal's algorithm: implementation challenge

Challenge. Would adding edge v–w to tree T create a cycle? If not, add it.

How difficult?

- $O(E + V)$ time.
- $O(V)$ time. ← run DFS from v, check if w is reachable (T has at most V – 1 edges)
- $O(\log V)$ time.
- $O(\log^* V)$ time.
- Constant time.



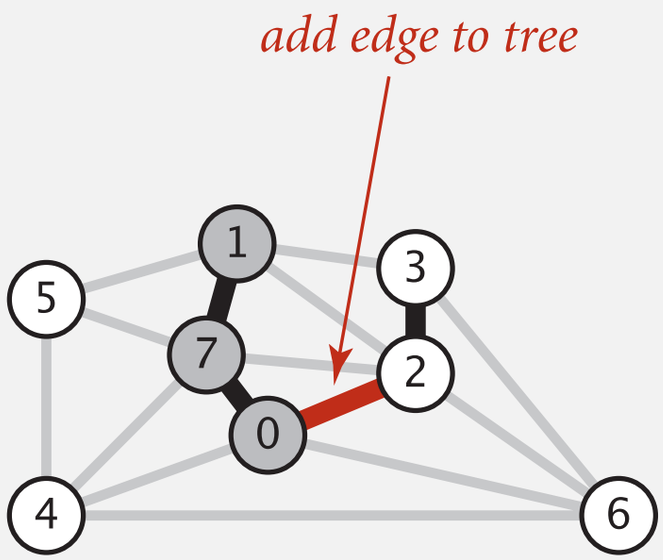


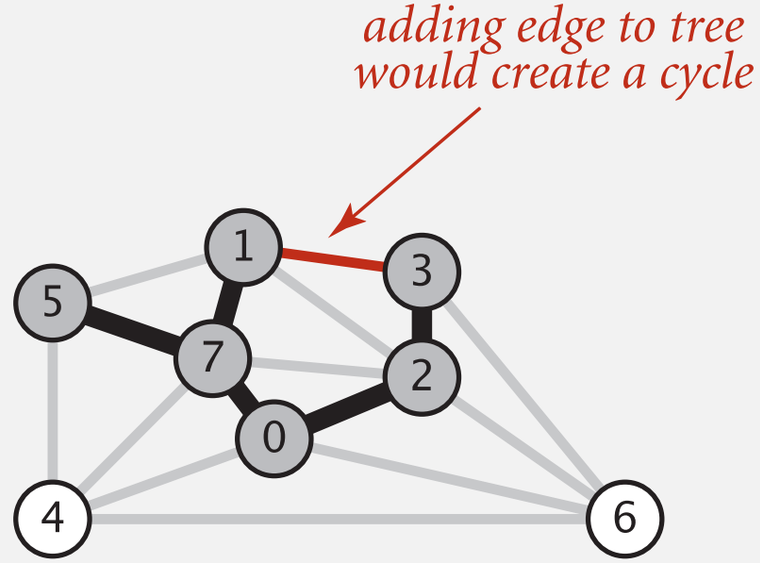
Kruskal's algorithm: implementation challenge

Challenge. Would adding edge v–w to tree T create a cycle? If not, add it.

How difficult?

- $O(E + V)$ time.
- $O(V)$ time. ← run DFS from v, check if w is reachable (T has at most V – 1 edges)
- $O(\log V)$ time.
- $O(\log^* V)$ time. ← use the union-find data structure !
- Constant time.



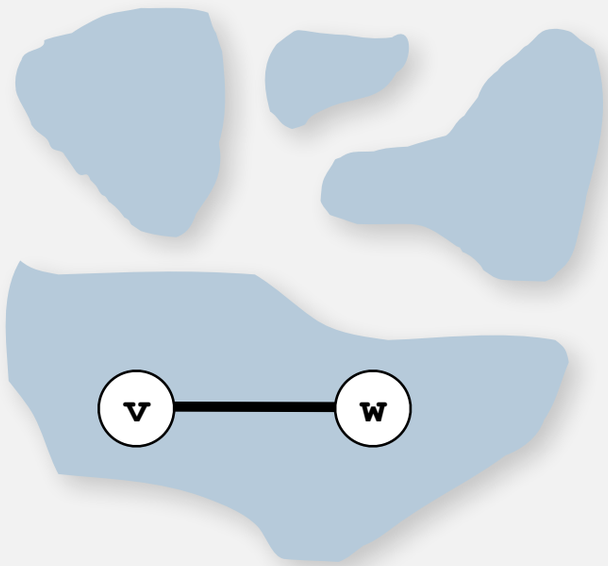


Kruskal's algorithm: implementation challenge

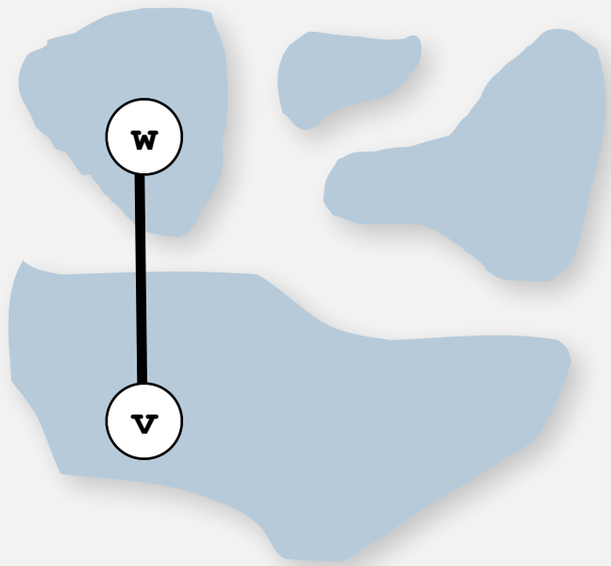
Challenge. Would adding edge v–w to tree T create a cycle? If not, add it.

Efficient solution. Use the union-find data structure.

- Maintain a set for each connected component in T.
- If v and w are in same set, then adding v–w would create a cycle.
- To add v–w to T, merge sets containing v and w.



Case 1: adding v–w creates a cycle



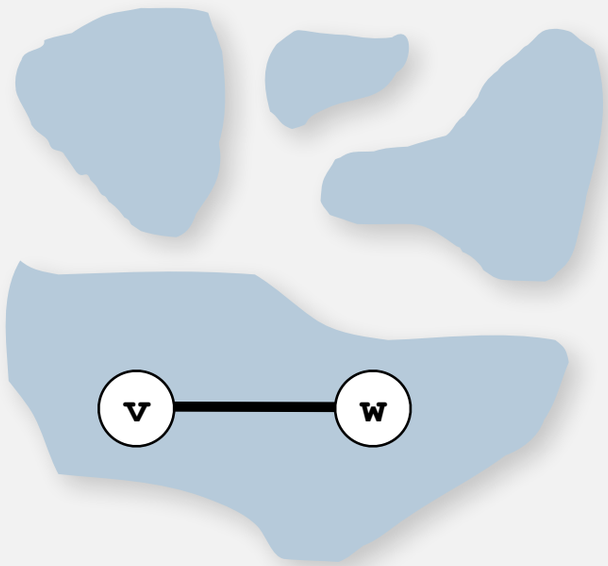
Case 2: add v–w to T and merge sets containing v and w

Kruskal's algorithm: implementation challenge

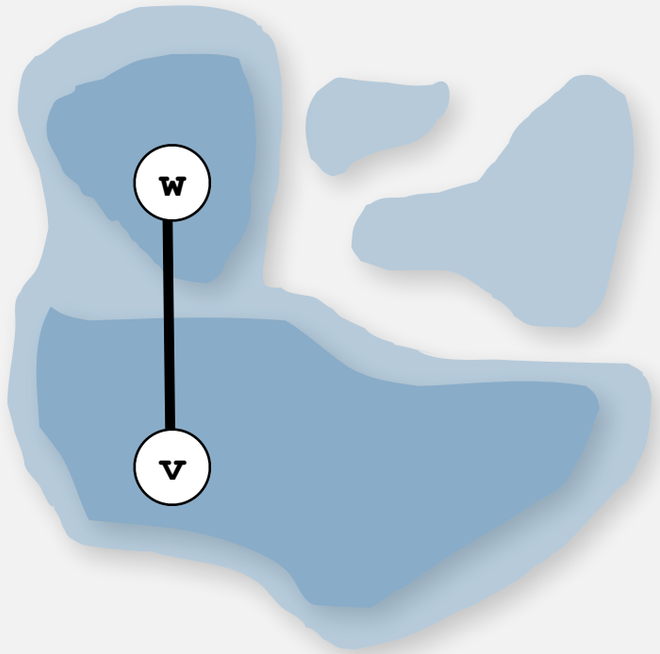
Challenge. Would adding edge v–w to tree T create a cycle? If not, add it.

Efficient solution. Use the union-find data structure.

- Maintain a set for each connected component in T.
- If v and w are in same set, then adding v–w would create a cycle.
- To add v–w to T, merge sets containing v and w.



Case 1: adding v–w creates a cycle



Case 2: add v–w to T and merge sets containing v and w

Kruskal's algorithm: Java implementation

```
public class KruskalMST
{
   private Queue<Edge> mst;
   private MinPQ<Edge> pq;

   public KruskalMST(EdgeWeightedGraph G)
   {
      mst = new Queue<Edge>();
      pq = new MinPQ<Edge>(G.edges());
      UnionFind uf = new UnionFind(G.V());
      while (!pq.isEmpty() && mst.size() < G.V()-1)
      {
         Edge e = pq.delMin();
         int v = e.either(), w = e.other(v);
         if (!uf.find(v, w))
         {
            uf.union(v, w);
            mst.enqueue(e);
         }
      }
   }

   public Iterable<Edge> edges()
   {  return mst;  }
}
```

build priority queue

greedily add edges to MST

edge v–w does not create cycle

merge sets

add edge to MST

Kruskal's algorithm running time

Proposition. Kruskal's algorithm computes MST in $O(E \log E)$ time.

Pf.

operation	frequency	time per op
build pq	1	E
del min	E	log E
union	V	log* V †
find	E	log* V †

† amortized bound using weighted quick union with path compression

recall: log* V ≤ 5 in this universe

Remark. If edges are already sorted, order of growth is $E \log^* V$.

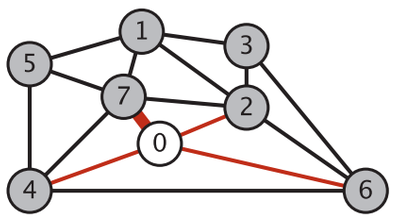
- edge-weighted graph API
- greedy algorithm
- Kruskal's algorithm
- **Prim's algorithm**
- advanced topics

Prim's algorithm example

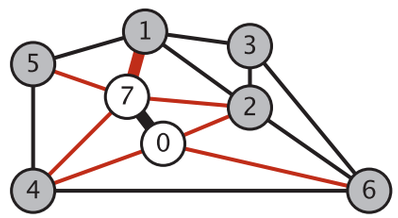
Prim's algorithm. [Jarník 1930, Dijkstra 1957, Prim 1959]

Start with vertex 0 and greedily grow tree T. At each step,

add to T the min weight edge with exactly one endpoint in T.

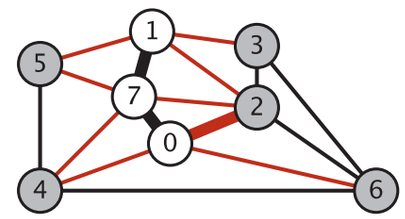


```
0-7 0.16
0-2 0.26
0-4 0.38
6-0 0.58
```

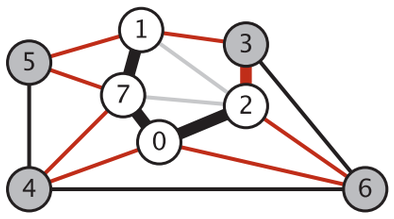


edges with exactly one endpoint in T (sorted by weight)

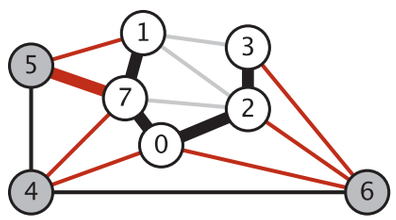
```
1-7 0.19
0-2 0.26
5-7 0.28
2-7 0.34
4-7 0.37
0-4 0.38
6-0 0.58
```



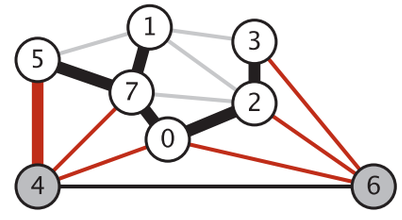
```
0-2 0.26
5-7 0.28
1-3 0.29
1-5 0.32
2-7 0.34
1-2 0.36
4-7 0.37
0-4 0.38
0-6 0.58
```



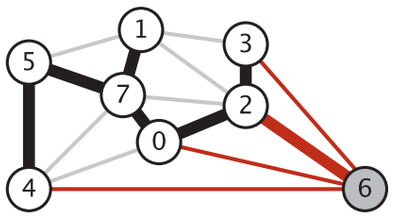
```
2-3 0.17
5-7 0.28
1-3 0.29
1-5 0.32
4-7 0.37
0-4 0.38
6-2 0.40
6-0 0.58
```



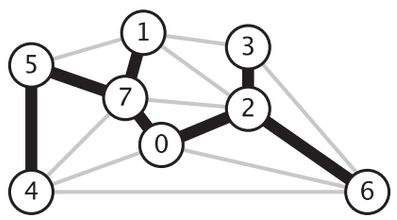
```
5-7 0.28
1-5 0.32
4-7 0.37
0-4 0.38
6-2 0.40
3-6 0.52
6-0 0.58
```



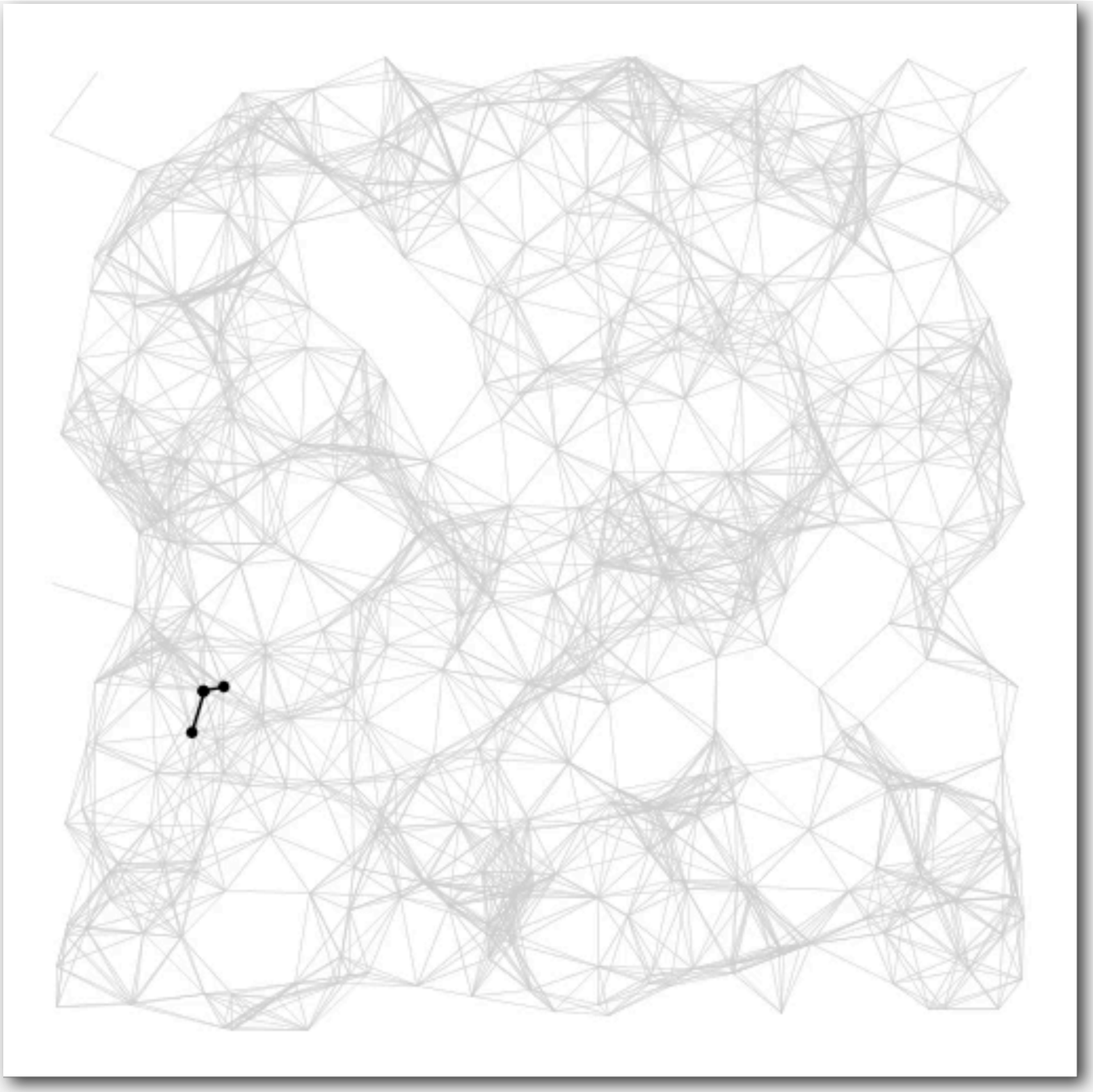
```
4-5 0.35
4-7 0.37
0-4 0.38
6-2 0.40
3-6 0.52
6-0 0.58
```



```
6-2 0.40
3-6 0.52
6-0 0.58
6-4 0.93
```

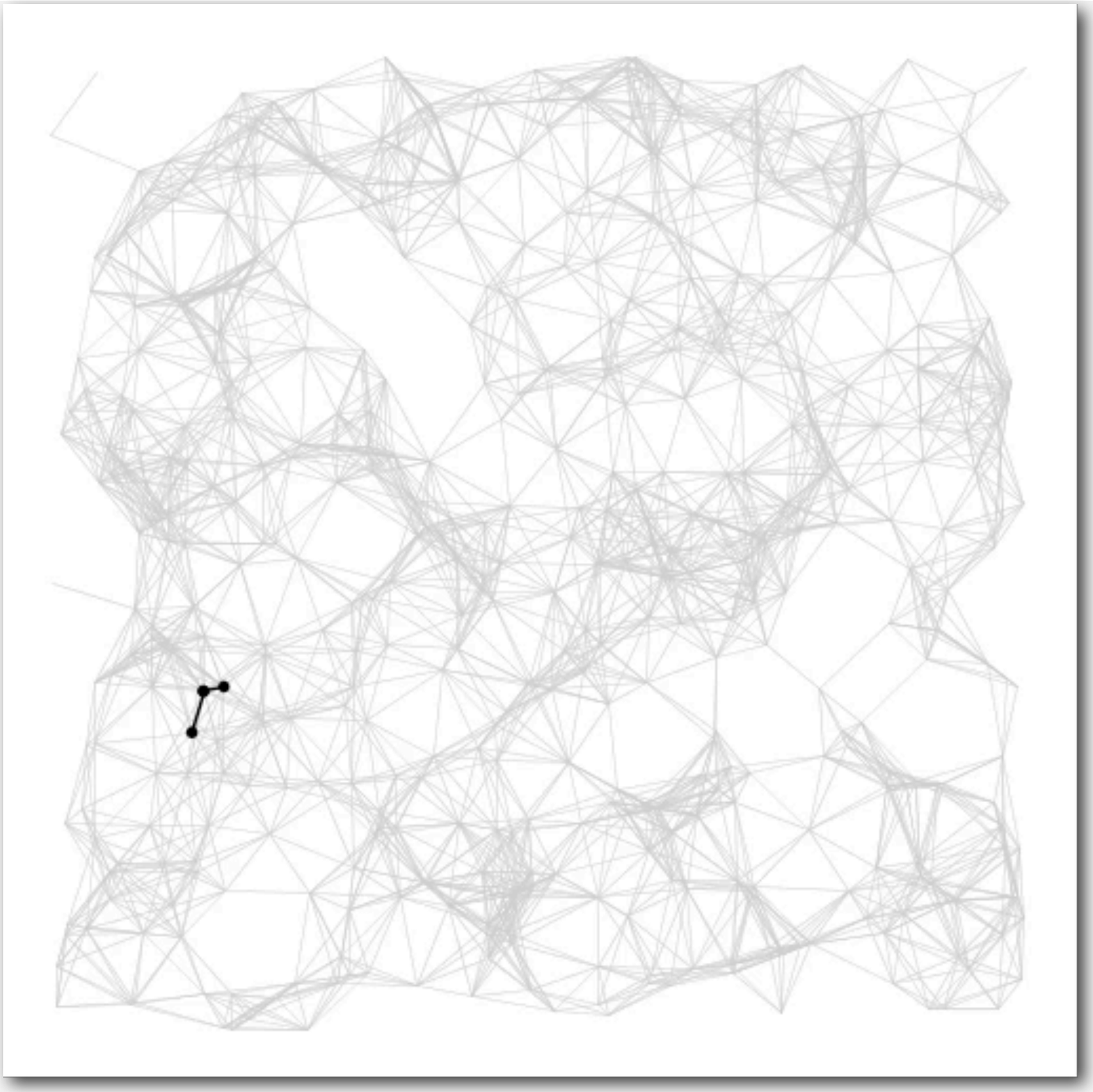


Prim's algorithm: visualization



https://www.cs.purdue.edu/homes/cs251/slides/media/Prim500.mov

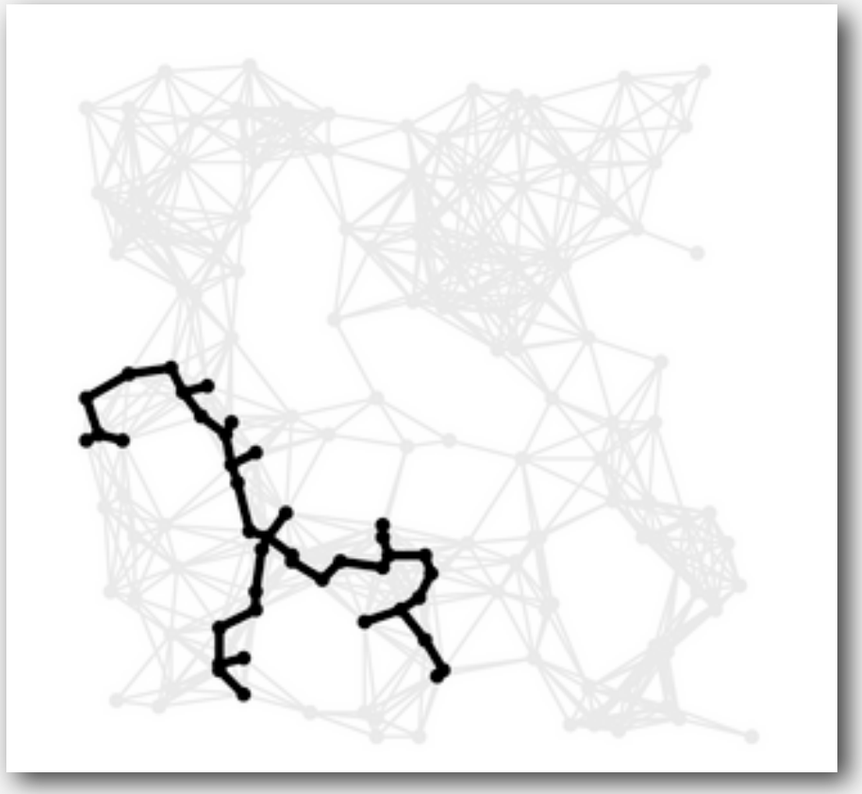
Prim's algorithm: visualization



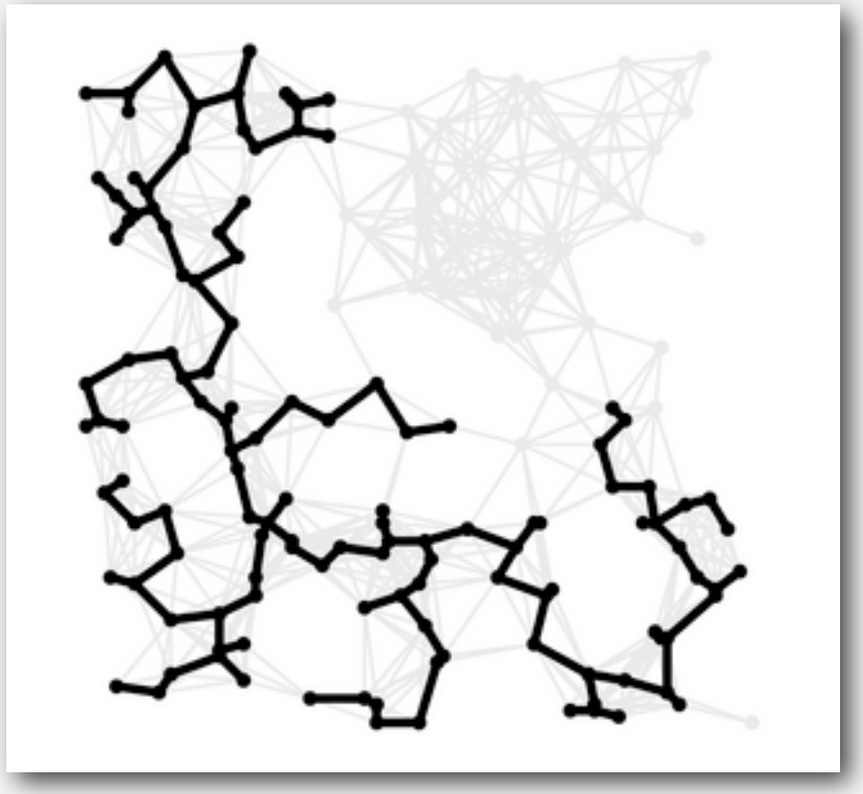
https://www.cs.purdue.edu/homes/cs251/slides/media/Prim500.mov

Prim's algorithm: visualization

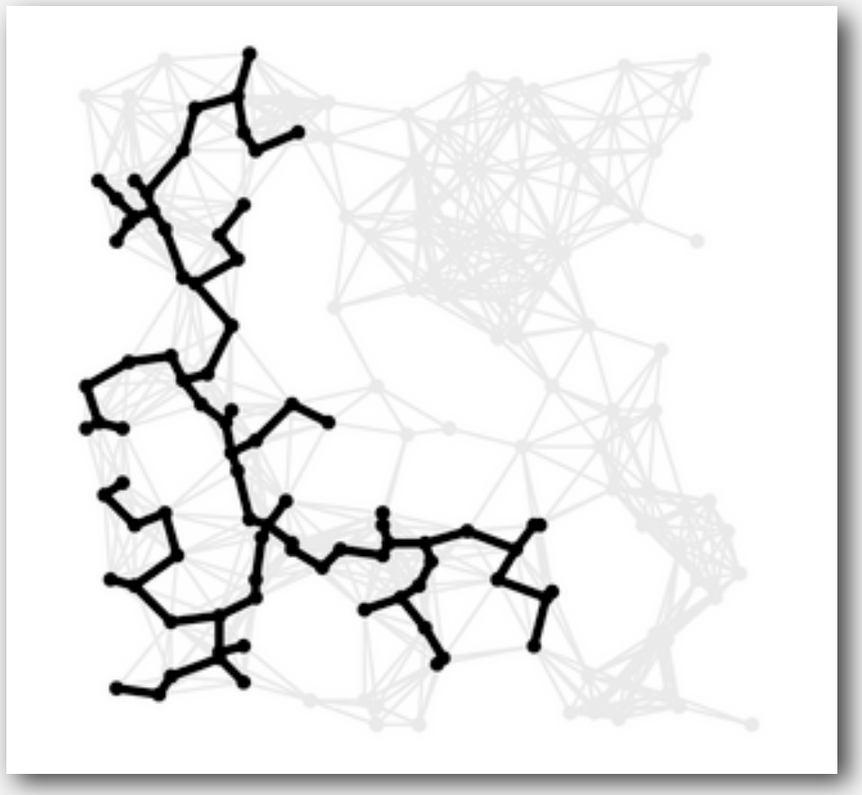
25%



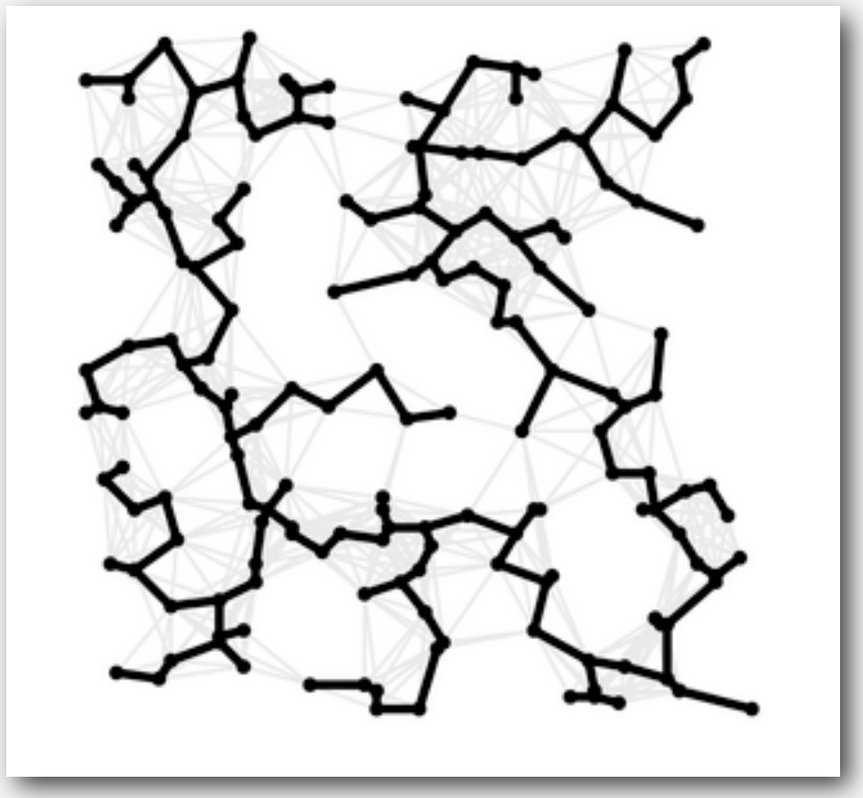
75%



50%



100%

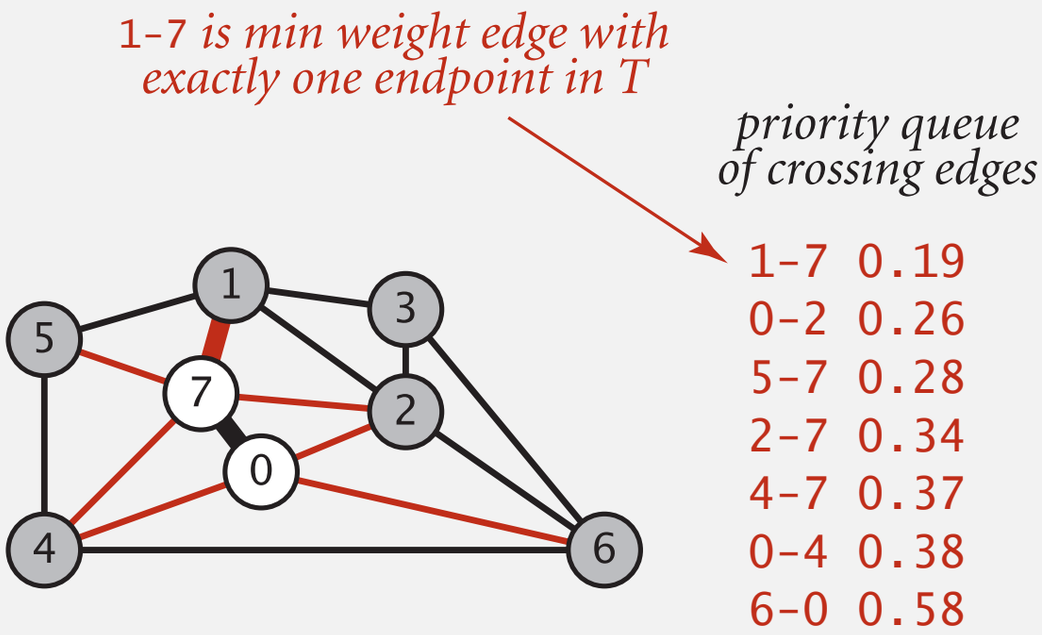


Prim's algorithm: implementation challenge

Challenge. Find the min weight edge with exactly one endpoint in T.

How difficult?

- $O(E)$ time.
- $O(V)$ time.
- $O(\log E)$ time.
- $O(\log^* E)$ time.
- Constant time.

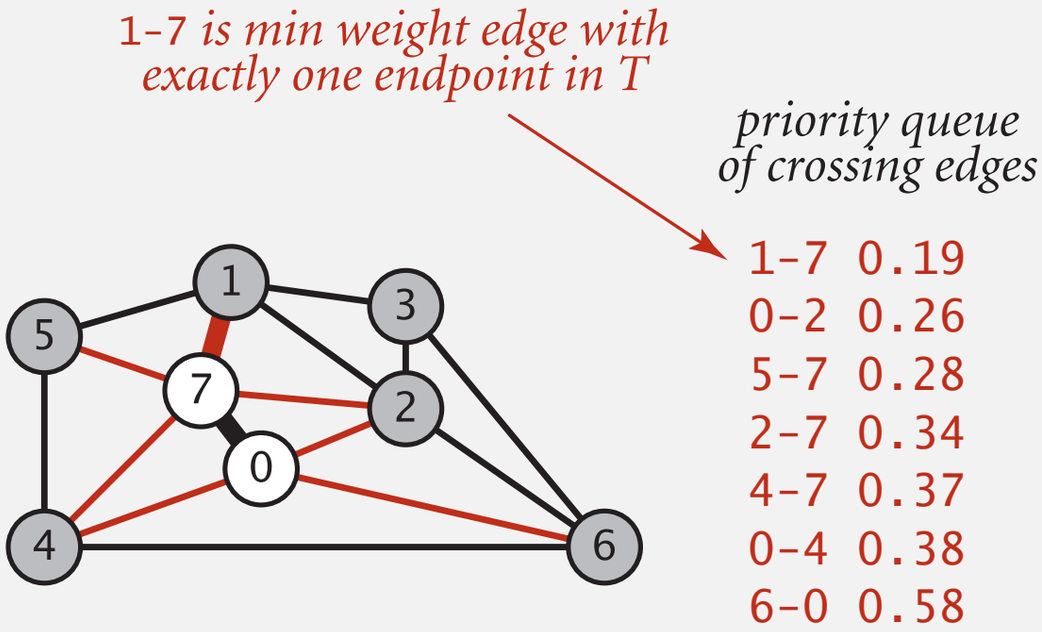


Prim's algorithm: implementation challenge

Challenge. Find the min weight edge with exactly one endpoint in T.

How difficult?

- $O(E)$ time. ← try all edges
- $O(V)$ time.
- $O(\log E)$ time.
- $O(\log^* E)$ time.
- Constant time.

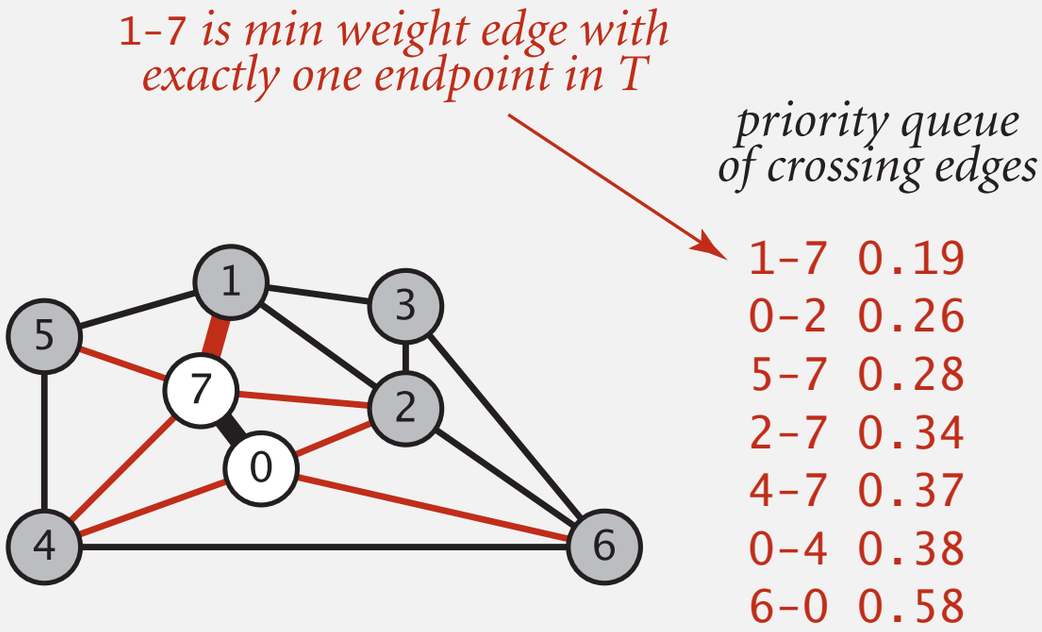


Prim's algorithm: implementation challenge

Challenge. Find the min weight edge with exactly one endpoint in T.

How difficult?

- $O(E)$ time. ← try all edges
- $O(V)$ time.
- $O(\log E)$ time. ← use a priority queue !
- $O(\log^* E)$ time.
- Constant time.

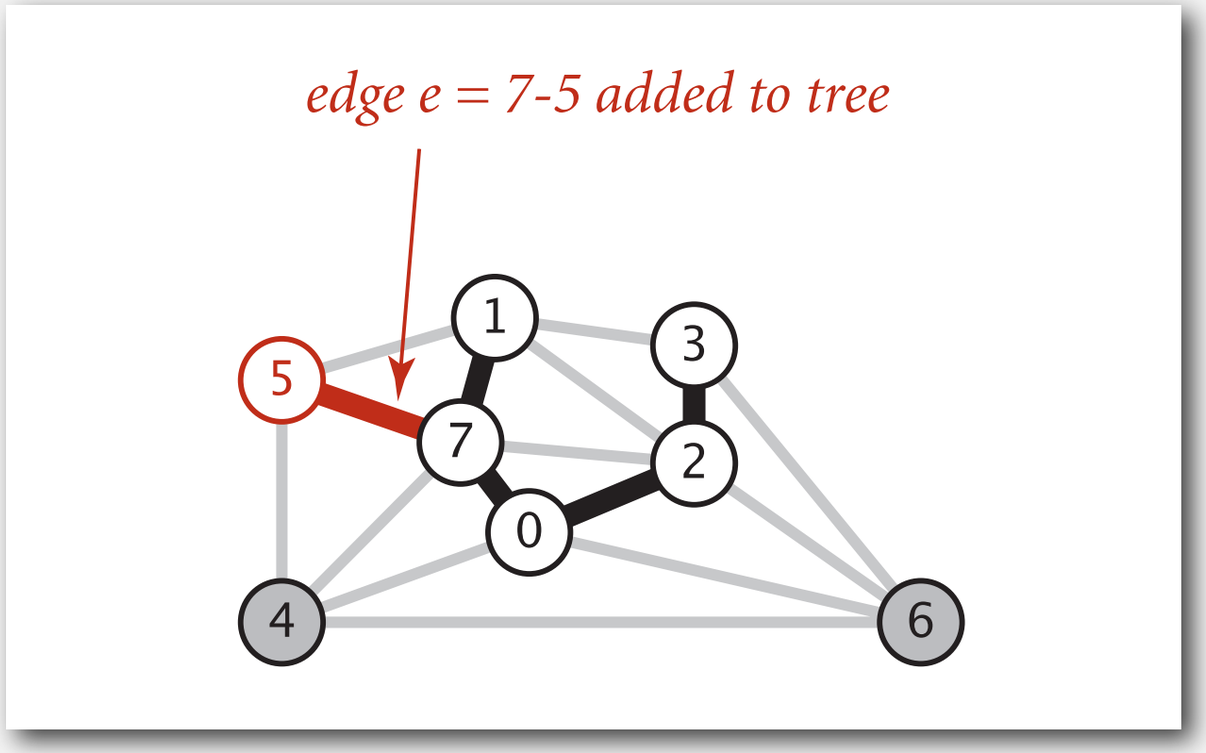


Prim's algorithm: proof of correctness

Proposition. Prim's algorithm computes the MST.

Pf. Prim's algorithm is a special case of the greedy MST algorithm.

- Suppose edge e = min weight edge connecting a vertex on the tree to a vertex not on the tree.
- Cut = set of vertices connected on tree.
- No crossing edge is black.
- No crossing edge has lower weight.



Prim's algorithm: lazy implementation

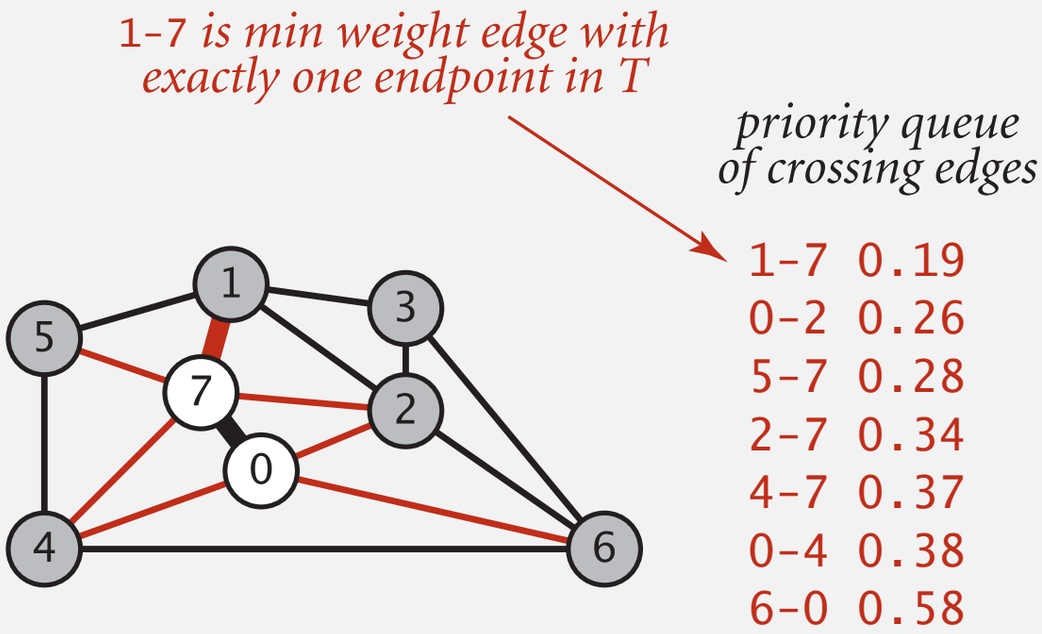
Challenge. Find the min weight edge with exactly one endpoint in T.

Lazy solution. Maintain a PQ of edges with (at least) one endpoint in T.

- Delete min to determine next edge $e = v$–w to add to T.
- Disregard if both endpoints v and w are in T.
- Otherwise, let v be vertex not in T:

  - add to PQ any edge incident to v (assuming other endpoint not in T)

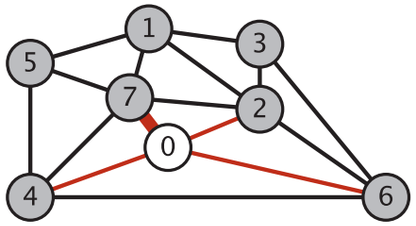
  - add v to T



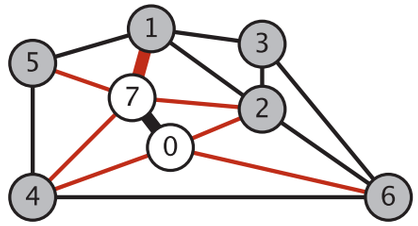
Prim's algorithm example: lazy implementation

Use `MinPQ`: key = edge, prioritized by weight.

(lazy version leaves some obsolete edges on the PQ)

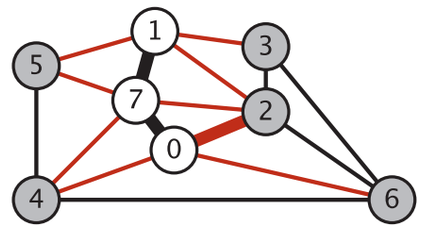


```
* 0-7 0.16
* 0-2 0.26
* 0-4 0.38
* 6-0 0.58
```

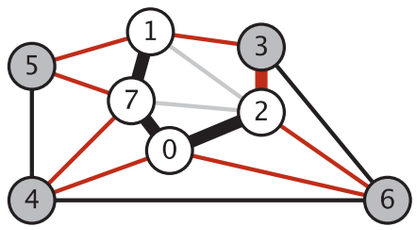


** marks new priority queue entry*

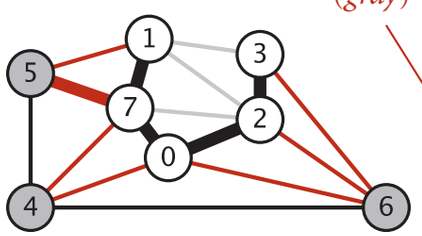
```
* 1-7 0.19
  0-2 0.26
* 5-7 0.28
* 2-7 0.34
* 4-7 0.37
  0-4 0.38
  6-0 0.58
```



```
  0-2 0.26
  5-7 0.28
* 1-3 0.29
* 1-5 0.32
  2-7 0.34
* 1-2 0.36
  4-7 0.37
  0-4 0.38
  0-6 0.58
```

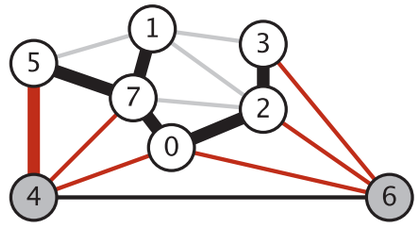


```
* 2-3 0.17
  5-7 0.28
  1-3 0.29
  1-5 0.32
  2-7 0.34
  1-2 0.36
  4-7 0.37
  0-4 0.38
* 6-2 0.40
  6-0 0.58
```

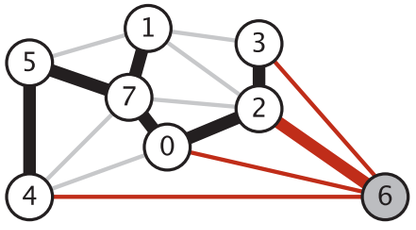


obsolete edges (gray)

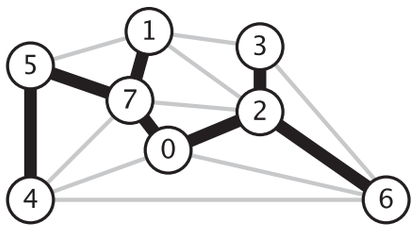
```
  5-7 0.28
  1-3 0.29
  1-5 0.32
  2-7 0.34
  1-2 0.36
  4-7 0.37
  0-4 0.38
  6-2 0.40
* 3-6 0.52
  6-0 0.58
```



```
  1-3 0.29
  1-5 0.32
  2-7 0.34
* 4-5 0.35
  1-2 0.36
  4-7 0.37
  0-4 0.38
  6-2 0.40
  3-6 0.52
  6-0 0.58
```



```
  1-2 0.36
  4-7 0.37
  0-4 0.38
  6-2 0.40
  3-6 0.52
  6-0 0.58
* 6-4 0.93
```



```
  3-6 0.52
  6-0 0.58
  6-4 0.93
```

Prim's algorithm: lazy implementation

```
public class LazyPrimMST
{
   private boolean[] marked;    // MST vertices
   private Queue<Edge> mst;     // MST edges
   private MinPQ<Edge> pq;      // PQ of edges

    public LazyPrimMST(WeightedGraph G)
    {
        pq = new MinPQ<Edge>();
        mst = new Queue<Edge>();
        marked = new boolean[G.V()];
        visit(G, 0);

        while (!pq.isEmpty())
        {
           Edge e = pq.delMin();
           int v = e.either(), w = e.other(v);
           if (marked[v] && marked[w]) continue;
           mst.enqueue(e);
           if (!marked[v]) visit(G, v);
           if (!marked[w]) visit(G, w);
        }
   }
}
```

assume G is connected

repeatedly delete the min weight edge e = v–w from PQ

ignore if both endpoints in T

add edge e to tree

add v or w to tree

Prim's algorithm: lazy implementation

```
private void visit(WeightedGraph G, int v)
{
   marked[v] = true;
   for (Edge e : G.adj(v))
      if (!marked[e.other(v)])
         pq.insert(e);
}

public Iterable<Edge> mst()
{  return mst;  }
```

add v to T

for each edge e = v–w, add to PQ if w not already in T

Prim's algorithm: running time

Proposition. Lazy Prim's algorithm computes the MST in time proportional to $E \log E$ in the worst case.

Pf.

operation	frequency	binary heap
delete min	E	log E
insert	E	log E