Algorithms & Models of Computation
CS/ECE 374, Fall 2017

Algorithms for Minimum Spanning Trees
Lecture 20
Thursday, November 9, 2017

Part I
Algorithms for Minimum Spanning Tree

Minimum Spanning Tree

Input  Connected graph $G = (V, E)$ with edge costs
Goal  Find $T \subseteq E$ such that $(V, T)$ is connected and total cost of all edges in $T$ is smallest

- $T$ is the minimum spanning tree (MST) of $G$

Applications

- Network Design
  - Designing networks with minimum cost but maximum connectivity
- Approximation algorithms
  - Can be used to bound the optimality of algorithms to approximate Traveling Salesman Problem, Steiner Trees, etc.
- Cluster Analysis
Some basic properties of Spanning Trees

- A graph $G$ is connected iff it has a spanning tree
- Every spanning tree of a graph on $n$ nodes has $n - 1$ edges
- Let $T = (V, E_T)$ be a spanning tree of $G = (V, E)$. For every non-tree edge $e \in E \setminus E_T$ there is a unique cycle $C$ in $T + e$. For every edge $f \in C - \{e\}$, $T - f + e$ is another spanning tree of $G$.

Assumption

And for now ...

Assumption

Edge costs are distinct, that is no two edge costs are equal.

Cuts

Definition

Given a graph $G = (V, E)$, a cut is a partition of the vertices of the graph into two sets $(S, V \setminus S)$.

Edges having an endpoint on both sides are the edges of the cut.

A cut edge is crossing the cut.
Safe and Unsafe Edges

**Definition**

An edge \( e = (u, v) \) is a **safe** edge if there is some partition of \( V \) into \( S \) and \( V \setminus S \) and \( e \) is the unique minimum cost edge crossing \( S \) (one end in \( S \) and the other in \( V \setminus S \)).

**Definition**

An edge \( e = (u, v) \) is an **unsafe** edge if there is some cycle \( C \) such that \( e \) is the unique maximum cost edge in \( C \).

**Proposition**

If edge costs are distinct then every edge is either safe or unsafe.

**Proof.**

Exercise.

 unsafe edge

Example...

Every cut identifies one unsafe edge...

![Diagram of a graph with cuts and edges labeled](image1)

...the most expensive edge in the cycle.

Note: An edge \( e \) may be a safe edge for many cuts!
Example

![Graph with unique edge costs. Safe edges are red, rest are unsafe.]

Figure: Graph with unique edge costs. Safe edges are red, rest are unsafe.

And all safe edges are in the MST in this case...

Some key observations
Proofs later

Lemma
*If e is a safe edge then every minimum spanning tree contains e.*

Lemma
*If e is an unsafe edge then no MST of G contains e.*

Part III
The Algorithms

Greedy Template

Initially \( E \) is the set of all edges in \( G \).
\( T \) is empty (* \( T \) will store edges of a MST *)

while \( E \) is not empty do
  choose \( e \in E \)
  if \( e \) satisfies condition
    add \( e \) to \( T \)
return the set \( T \)

Main Task: In what order should edges be processed? When should we add edge to spanning tree?
Kruskal’s Algorithm
Process edges in the order of their costs (starting from the least) and add edges to \( T \) as long as they don’t form a cycle.

![Kruskal's Algorithm Diagram]

Prim’s Algorithm
\( T \) maintained by algorithm will be a tree. Start with a node in \( T \). In each iteration, pick edge with least attachment cost to \( T \).

![Prim's Algorithm Diagram]

Reverse Delete Algorithm
Initially \( E \) is the set of all edges in \( G \). \( T \) is \( E \) (* \( T \) will store edges of a MST *)

\[
\begin{align*}
&\text{while } E \text{ is not empty do} \\
&\quad \text{choose } e \in E \text{ of largest cost} \\
&\quad \text{if removing } e \text{ does not disconnect } T \text{ then} \\
&\quad \quad \text{return the set } T \\
&\quad \quad \text{remove } e \text{ from } T \\
&\end{align*}
\]

Returns a minimum spanning tree.

Borůvka’s Algorithm
Simplest to implement. See notes.
Assume \( G \) is a connected graph.

\[
\begin{align*}
&\text{while } T \text{ is not spanning do} \\
&\quad \text{X} \leftarrow \emptyset \\
&\quad \text{for each connected component } S \text{ of } T \text{ do} \\
&\quad \quad \text{add to } X \text{ the cheapest edge between } S \text{ and } V \setminus S \\
&\quad \text{Add edges in } X \text{ to } T \\
&\end{align*}
\]

\[
\begin{align*}
&\text{return the set } T \\
&\end{align*}
\]
Correctness of MST Algorithms

- Many different MST algorithms
- All of them rely on some basic properties of MSTs, in particular the Cut Property to be seen shortly.

Key Observation: Cut Property

**Lemma**

*If e is a safe edge then every minimum spanning tree contains e.*

**Proof.**

- Suppose (for contradiction) e is not in MST $T$.
- Since e is safe there is an $S \subseteq V$ such that e is the unique min cost edge crossing S.
- Since $T$ is connected, there must be some edge $f$ with one end in $S$ and the other in $V \setminus S$.
- Since $c_f > c_e$, $T' = (T \setminus \{f\}) \cup \{e\}$ is a spanning tree of lower cost! Error: $T'$ may not be a spanning tree!!
Error in Proof: Example
Problematic example. $S = \{1, 2, 7\}$, $e = (7, 3)$, $f = (1, 6)$. $T - f + e$ is not a spanning tree.

(A) Consider adding the edge $f$.
(B) It is safe because it is the cheapest edge in the cut.
(C) Let’s throw out the edge $e$ currently in the spanning tree, which is more expensive than $f$ and is in the same cut. Put it $f$ instead…
(D) New graph of selected edges is not a tree anymore. BUG.

Proof of Cut Property

Proof.

Suppose $e \in (v, w)$ is not in $\text{MST}$ $T$ and $f$ is min weight edge in cut $(S \cup \{v\}, S)$. Assume $v \in S$.

Let $w'$ be the first vertex in $P$ belonging to $V \setminus S$; let $v'$ be the vertex just before it on $P$, and let $e' = (v', w')$.

$T' = (T \cup \{e'\}) \cup \{e\}$ is spanning tree of lower cost. (Why?)

Proof of Cut Property (contd)

Observation

$T' = (T \setminus \{e'\}) \cup \{e\}$ is a spanning tree.

Proof.

$T'$ is connected.

Removed $e' = (v', w')$ from $T$ but $v'$ and $w'$ are connected by the path $P - f + e$ in $T'$. Hence $T'$ is connected if $T$ is.

$T'$ is a tree.

$T'$ is connected and has $n - 1$ edges (since $T$ had $n - 1$ edges) and hence $T'$ is a tree.

Safe Edges form a Tree

Lemma

Let $G$ be a connected graph with distinct edge costs, then the set of safe edges form a connected graph.

Proof.

Suppose not. Let $S$ be a connected component in the graph induced by the safe edges.

Consider the edges crossing $S$, there must be a safe edge among them since edge costs are distinct and so we must have picked it.
Safe Edges form an MST

**Corollary**

Let \( G \) be a connected graph with distinct edge costs, then set of safe edges form the unique MST of \( G \).

**Consequence:** Every correct MST algorithm when \( G \) has unique edge costs includes exactly the safe edges.

---

Correctness of Prim’s Algorithm

**Prim’s Algorithm**

Pick edge with minimum attachment cost to current tree, and add to current tree.

**Proof of correctness.**

1. If \( e \) is added to tree, then \( e \) is safe and belongs to every MST.
   - Let \( S \) be the vertices connected by edges in \( T \) when \( e \) is added.
   - \( e \) is edge of lowest cost with one end in \( S \) and the other in \( V \setminus S \) and hence \( e \) is safe.
   - Set of edges output is a spanning tree
     - Set of edges output forms a connected graph: by induction, \( S \) is connected in each iteration and eventually \( S = V \).
     - Only safe edges added and they do not have a cycle.

---

Correctness of Kruskal’s Algorithm

**Kruskal’s Algorithm**

Pick edge of lowest cost and add if it does not form a cycle with existing edges.

**Proof of correctness.**

1. If \( e = (u, v) \) is added to tree, then \( e \) is safe
   - When algorithm adds \( e \) let \( S \) and \( S' \) be the connected components containing \( u \) and \( v \) respectively
   - \( e \) is the lowest cost edge crossing \( S \) (and also \( S' \)).
   - If there is an edge \( e' \) crossing \( S \) and has lower cost than \( e \), then \( e' \) would come before \( e \) in the sorted order and would be added by the algorithm to \( T \)
   - Set of edges output is a spanning tree: exercise
Correctness of Borůvka’s Algorithm

**Proof of correctness.**
Argue that only safe edges are added.

---

Correctness of Reverse Delete Algorithm

**Reverse Delete Algorithm**
Consider edges in decreasing cost and remove an edge if it does not disconnect the graph

**Proof of correctness.**
Argue that only unsafe edges are removed.

---

When edge costs are not distinct

**Heuristic argument:** Make edge costs distinct by adding a small tiny and different cost to each edge

**Formal argument:** Order edges lexicographically to break ties
1. \( e_i \prec e_j \) if either \( c(e_i) < c(e_j) \) or \( c(e_i) = c(e_j) \) and \( i < j \)
2. Lexicographic ordering extends to sets of edges. If \( A, B \subseteq E \), \( A \neq B \) then \( A \prec B \) if either \( c(A) < c(B) \) or \( c(A) = c(B) \) and \( A \setminus B \) has a lower indexed edge than \( B \setminus A \)
3. Can order all spanning trees according to lexicographic order of their edge sets. Hence there is a unique MST.

Prim’s, Kruskal, and Reverse Delete Algorithms are optimal with respect to lexicographic ordering.

---

Edge Costs: Positive and Negative

1. Algorithms and proofs don’t assume that edge costs are non-negative! MST algorithms work for arbitrary edge costs.
2. Another way to see this: make edge costs non-negative by adding to each edge a large enough positive number. Why does this work for MSTs but not for shortest paths?
3. Can compute *maximum* weight spanning tree by negating edge costs and then computing an MST.
   **Question:** Why does this not work for shortest paths?
Implementing Borůvka’s Algorithm

No complex data structure needed.

```
T is ∅ (* T will store edges of a MST *)
while T is not spanning do
    X ← ∅
    for each connected component S of T do
        add to X the cheapest edge between S and V \ S
        Add edges in X to T
    return the set T
```

- \( O(\log n) \) iterations of while loop. Why? Number of connected components shrink by at least half since each component merges with one or more other components.
- Each iteration can be implemented in \( O(m) \) time.

**Running time**: \( O(m \log n) \) time.

Implementing Prim’s Algorithm

**Implementing Prim’s Algorithm**

```
Prim_ComputeMST
    E is the set of all edges in G
    S = {1}
    T is empty (* T will store edges of a MST *)
    while S ≠ V do
        pick e = (v, w) ∈ E such that v ∈ S and w ∈ V − S
        e has minimum cost
        T = T ∪ e
        S = S ∪ w
    return the set T
```

**Analysis**

- Number of iterations = \( O(n) \), where \( n \) is number of vertices
- Picking \( e \) is \( O(m) \) where \( m \) is the number of edges
- Total time \( O(n m) \)

More Efficient Implementation

```
Prim_ComputeMST
    E is the set of all edges in G
    S = {1}
    T is empty (* T will store edges of a MST *)
    for v ∈ S, a(v) = minw∈S c(w, v)
    for v ∈ S, e(v) = w such that w ∈ S and c(w, v) is minimum
    while S ≠ V do
        pick v with minimum a(v)
        T = T ∪ \{(e(v), v)\}
        S = S ∪ {v}
        update arrays a and e
    return the set T
```

Maintain vertices in \( V \setminus S \) in a priority queue with key \( a(v) \).
Priority Queues

Data structure to store a set $S$ of $n$ elements where each element $v \in S$ has an associated real/integer key $k(v)$ such that the following operations

- **makeQ**: create an empty queue
- **findMin**: find the minimum key in $S$
- **extractMin**: Remove $v \in S$ with smallest key and return it
- **add$(v, k(v))$**: Add new element $v$ with key $k(v)$ to $S$
- **Delete$(v)$**: Remove element $v$ from $S$
- **decreaseKey $(v, k'(v))$**: decrease key of $v$ from $k(v)$ (current key) to $k'(v)$ (new key). Assumption: $k'(v) \leq k(v)$
- **meld**: merge two separate priority queues into one

Prim’s using priority queues

$E$ is the set of all edges in $G$
$S = \{ 1 \}$
$T$ is empty (∗ $T$ will store edges of a MST ∗)
for $v \not\in S$, $a(v) = \min_{w \in S} c(w, v)$
for $v \not\in S$, $e(v) = w$ such that $w \in S$ and $c(w, v)$ is minimum
while $S \neq V$ do
    pick $v$ with minimum $a(v)$
    $T = T \cup \{(e(v), v)\}$
    $S = S \cup \{v\}$
    update arrays $a$ and $e$
return the set $T$

Running time of Prim’s Algorithm

$O(n)$ extractMin operations and $O(m)$ decreaseKey operations

- Using standard Heaps, extractMin and decreaseKey take $O(\log n)$ time. Total: $O((m+n) \log n)$
- Using Fibonacci Heaps, $O(\log n)$ for extractMin and $O(1)$ (amortized) for decreaseKey. Total: $O(n \log n + m)$.
- Prim’s algorithm and Dijkstra’s algorithms are similar. Where is the difference?
- Prim’s algorithm = Dijkstra where length of a path $\pi$ is the weight of the heaviest edge in $\pi$. (Bottleneck shortest path.)

Kruskal’s Algorithm

Kruskal_ComputeMST
Initially $E$ is the set of all edges in $G$
$T$ is empty (∗ $T$ will store edges of a MST ∗)
while $E$ is not empty do
    choose $e \in E$ of minimum cost
    if $(T \cup \{e\}$ does not have cycles)
        add $e$ to $T$
return the set $T$

- Presort edges based on cost. Choosing minimum can be done in $O(1)$ time
- Do BFS/DFS on $T \cup \{e\}$. Takes $O(n)$ time
- Total time $O(m \log m) + O(mn) = O(mn)$
Implementing Kruskal’s Algorithm Efficiently

Kruskal’s algorithm finds the minimum spanning tree (MST) of a connected, undirected graph. It is a greedy algorithm that finds a subset of the edges that forms a tree that includes every vertex, where the total weight of all the edges in the tree is minimized.

**Kruskal_ComputeMST**
- Sort edges in $E$ based on cost
- $T$ is empty (i.e., $T$ will store edges of a MST)
- each vertex $u$ is placed in a set by itself
  - while $E$ is not empty do
    - pick $e = (u, v) \in E$ of minimum cost
    - if $u$ and $v$ belong to different sets
      - add $e$ to $T$
      - merge the sets containing $u$ and $v$
  - return the set $T$

Need a data structure to check if two elements belong to same set and to merge two sets.
Using **Union-Find** data structure can implement Kruskal’s algorithm in $O((m + n) \log m)$ time.

**Best Known Asymptotic Running Times for MST**

Prim’s algorithm using Fibonacci heaps: $O(n \log n + m)$. If $m$ is $O(n)$ then running time is $\Omega(n \log n)$.

**Question**
Is there a linear time ($O(m + n)$ time) algorithm for MST?

1. $O(m \log^* m)$ time [Fredman, Tarjan 1987]
2. $O(m + n)$ time using bit operations in RAM model [Fredman, Willard 1994]
3. $O(m + n)$ expected time (randomized algorithm) [Karger, Klein, Tarjan 1995]
4. $O((n + m) \alpha(m, n))$ time Chazelle 2000
5. Still open: Is there an $O(n + m)$ time deterministic algorithm in the comparison model?