CSE 326: Data Structures
Spanning Trees

Brian Curless
Spring 2008

A Hidden Tree

Spanning Tree in a Graph

Vertex = router
Edge = link between routers

Spanning tree
- Connects all the vertices
- No cycles

Undirected Graph

\[ G = (V, E) \]
- \( V \) is a set of vertices (or nodes)
- \( E \) is a set of unordered pairs of vertices

\[ V = \{1, 2, 3, 4, 5, 6, 7\} \]
\[ E = \{(1, 2), (1, 6), (1, 5), (2, 7), (2, 3), (3, 4), (4, 7), (4, 5), (5, 6)\} \]

2 and 3 are adjacent
2 is incident to edge (2, 3)
Spanning Tree Problem

- Input: An undirected graph $G = (V,E)$. $G$ is connected.
- Output: $T$ contained in $E$ such that
  - $(V,T)$ is a connected graph
  - $(V,T)$ has no cycles

$\Rightarrow$ $(V,T)$ is a tree

Spanning Tree Algorithm

```
ST(Vertex i) {
  mark i;
  for each j adjacent to i {
    if (j is unmarked) {
      Add (i,j) to T;
      ST(j);
    }
  }
}
```

Main( ) {
  T = empty set;
  ST(2);
}

Example of Depth First Search

Example Step 2

ST(1)
Example Step 7

(1,2) (2,7) (7,5) (5,4) (4,3)

ST(1)
ST(2)
ST(7)
ST(5)
ST(3)

Example Step 8

(1,2) (2,7) (7,5) (5,4) (4,3)

ST(1)
ST(2)
ST(7)
ST(5)
ST(4)

Example Step 9

(1,2) (2,7) (7,5) (5,4) (4,3)

ST(1)
ST(2)
ST(7)
ST(5)
ST(4)

Example Step 10

(1,2) (2,7) (7,5) (5,4) (4,3)

ST(1)
ST(2)
ST(7)
ST(5)
How many edges in a spanning tree?

Before moving on, it will help us to know how many edges a spanning tree must have.

First, a couple of properties of trees...

A tree has one vertex of degree one

Property: A tree with $|V|>1$ has at least one vertex of degree one (i.e., with only one edge incident to it).

Proof by contradiction:
- Assume a “tree” with no such vertex.
- The graph has no endpoints and must contain a cycle.
- The graph is not a tree.
Removing a vertex of degree one gives a tree

**Property:** removing a vertex of degree one and the corresponding edge from a tree results in a tree.

**Proof:**
- Removing an edge cannot introduce cycles.
- Removing a vertex of degree one will not result in a disconnected graph.

So, the graph is acyclic and connected, i.e., a tree.

A spanning tree has $|V|-1$ edges

**Property:** a spanning tree over a graph with $|V|$ vertices has $|V|-1$ edges.

**Proof by induction:**
- Base case: 1 vertex $\rightarrow |V|-1 = 0$ edges.
- Inductive hypothesis: a spanning tree with $k-1$ vertices has $k-2$ edges.
- Inductive step: spanning tree with $k$ vertices has $k-1$ edges.

We must prove the inductive step:
- Spanning tree with $k$ vertices has at least one vertex of degree 1.
- Remove that vertex and its edge from the graph and the spanning tree.
- The result is a spanning tree of $k-1$ vertices, which must have $k-2$ edges (inductive hypothesis).
- Restoring that vertex and edge gives a tree with $k$ vertices and $k-1$ edges.

Best Spanning Tree

- Each edge has the probability that it won’t fail
- Find the spanning tree that is least likely to fail

Example of a Spanning Tree

Probability of success = \(0.85 \times 0.95 \times 0.89 \times 0.95 \times 1.0 \times 0.84 \)

\[= 0.5735\]
Minimum Spanning Trees

Given an undirected graph $G = (V,E)$, find a graph $G' = (V,E')$ such that:
- $E'$ is a subset of $E$
- $|E'| = |V| - 1$
- $G'$ is connected
- $\sum_{(u,v) \in E'} c_{uv}$ is minimal

$G'$ is a minimum spanning tree.

Applications: wiring a house, power grids, Internet connections

Minimum Spanning Tree Problem

- Input: Undirected Graph $G = (V,E)$ and a cost function $C$ from $E$ to the reals. $C(e)$ is the cost of edge $e$.
- Output: A spanning tree $T$ with minimum total cost. That is: $T$ that minimizes

$$C(T) = \sum_{e \in T} C(e)$$

Find the MST

Two Different Approaches

Prim’s Algorithm
Almost identical to Dijkstra’s

Kruskals’s Algorithm
Completely different!
Prim’s algorithm

Idea: Grow a tree by adding an edge from the “known” vertices to the “unknown” vertices. Pick the edge with the smallest weight.

Prim’s Algorithm for MST

A node-based greedy algorithm
Builds MST by greedily adding nodes

1. Select a node to be the “root”  
   • mark it as known  
   • Update cost of all its neighbors  
2. While there are unknown nodes left in the graph
   a. Select an unknown node \( b \) with the smallest cost from some known node \( a \)
   b. Mark \( b \) as known
   c. Add \( (a, b) \) to MST
   d. Update cost of all nodes adjacent to \( b \)

Prim’s Algorithm Analysis

Running time:
Same as Dijkstra’s: \( O(|E| \log |V| + |V| \log |V|) \)

Can we reduce this? \( 1 \leq |V| - 1 \) \( O(|E| \log |V|) \)

Correctness:
Proof is similar to Dijkstra’s
Kruskal’s MST Algorithm

Idea: Grow a forest out of edges that do not create a cycle. Pick an edge with the smallest weight.

\[ G = (V, E) \]

Kruskal’s Algorithm for MST

An edge-based greedy algorithm
Builds MST by greedily adding edges

1. Initialize with
   - empty MST
   - all vertices marked unconnected
   - all edges unmarked
2. While there are still unmarked edges
   a. Pick the lowest cost edge \((u, v)\) and mark it
   b. If \(u\) and \(v\) are not already connected, add \((u, v)\) to the MST and mark \(u\) and \(v\) as connected to each other

Doesn’t it sound familiar?

Example of Kruskal 1

<table>
<thead>
<tr>
<th>Edge</th>
<th>Weight</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,4)</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>(2,4)</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>(7,5)</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>(5,6)</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>(5,9)</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>(1,6)</td>
<td>3</td>
<td>R</td>
</tr>
<tr>
<td>(2,3)</td>
<td>3</td>
<td>R</td>
</tr>
<tr>
<td>(7,4)</td>
<td>3</td>
<td>R</td>
</tr>
</tbody>
</table>

Example of Kruskal 2

1  2  3  4  5  6  7
1  3  4  7  0  3  3

\((7,4) (2,1) (7,5) (5,6) (5,4) (1,6) (2,7) (2,3) (3,4) (1,5)\)

0  1  2  3  3  3  4
Data Structures for Kruskal

- Sorted edge list
  
  \[(7,4) (2,1) (7,5) (5,6) (5,4) (1,6) (2,7) (2,3) (3,4) (1,5)\]
  
  0 1 1 2 2 3 3 3 3 4

- Disjoint Union / Find
  - Union(a, b) - union the disjoint sets named by a and b
  - Find(a) returns the name of the set containing a

Example of DU/F 1

Example of DU/F 2

Example of DU/F 3
Kruskal’s Algorithm with DU / F

Sort the edges by increasing cost; \( |E| \log |E| \)
Initialize A to be empty;
for each edge \((i,j)\) chosen in increasing order do
\hspace{1cm} u := \text{Find}(i);
\hspace{1cm} v := \text{Find}(j);
\hspace{1cm} if not(u = v) then
\hspace{1cm} add \((i,j)\) to A;
\hspace{1cm} Union(u,v);
\hspace{1cm} \text{\textcopyright It}aks

This algorithm will work, but it goes through all the edges.
Is this always necessary?

Kruskal’s Algorithm: Correctness

It clearly generates a spanning tree. Call it \( T_K \).

Suppose \( T_K \) is not minimum:

- Pick another spanning tree \( T_{\text{min}} \) with lower cost than \( T_K \).
- Pick the smallest edge \( e_i=(u,v) \) in \( T_K \) that is not in \( T_{\text{min}} \).
- \( T_{\text{min}} \) already has a path \( p \) in \( T_{\text{min}} \) from \( u \) to \( v \)
  \hspace{1cm} \Rightarrow \text{Adding } e_i \text{ to } T_{\text{min}} \text{ will create a cycle in } T_{\text{min}}.
- Pick an edge \( e_2 \) in \( p \) that Kruskal’s algorithm considered
  \hspace{1cm} \text{after adding } e_1 \text{ (must exist: } u \text{ and } v \text{ unconnected when } e_i \text{ considered)}
  \hspace{1cm} \Rightarrow \text{cost}(e_2) \geq \text{cost}(e_i)
  \hspace{1cm} \Rightarrow \text{can replace } e_2 \text{ with } e_i \text{ in } T_{\text{min}} \text{ without increasing cost!}
- Keep doing this until \( T_{\text{min}} \) is identical to \( T_K \)
  \hspace{1cm} \Rightarrow T_K \text{ must also be minimal – contradiction!}

Your Turn

Find MST using Prim’s and Kruskal’s

```c
void Graph::kruskal() {
    int edgesAccepted = 0;
    DisjSet s(NUM_VERTICES);
    buildHeap(edge set);
    while (edgesAccepted < NUM_VERTICES - 1) {
        e = smallest weight edge not deleted yet;
        // edge e = (u, v)
       uset = s.find(u);
        vset = s.find(v);
        if (uset != vset) {
            edgesAccepted++;
            s.unionSets(uset, vset);
        }
    }
}
```

Total Cost: \( O(|E| + |E| \log |E| + |E| \log |V| + |V|) \)
Reducing Best to Minimum

Let $P(e)$ be the probability that an edge doesn’t fail. Define:

$$C(e) = -\log_{10}(P(e))$$

Minimizing $\sum_{e \in \mathcal{E}} C(e)$ is equivalent to maximizing $\prod_{e \in \mathcal{E}} P(e)$ because

$$\prod_{e \in \mathcal{E}} P(e) = \prod_{e \in \mathcal{E}} 10^{-C(e)} = 10^{-\sum_{e \in \mathcal{E}} C(e)}$$