Kruskals Minimum Spanning Tree algorithm

This note defines the MINIMUM-SPANNING-TREE problem, proposes a greedy solution, Kruskals algorithm, and proves that this algorithm is correct.

Remember that a Minimum Spanning Tree for an edge weighted graph $G = (V, E), w$ is a tree $(V, E^* \subseteq E)$ of minimal total weight. Let $\mathrm{MST}(G, w)$ be the set of minimum spanning trees for G, w. Remember that for a tree $T = (V, E)$, $w(T) = \sum_{e \in E} w(e)$.

Problem definition
MINIMUM-SPANNING-TREE:
INPUT: A connected graph and edge weight function $(G = (V, E), w : E \to \mathbb{N})$
OUTPUT: An member of $\mathrm{MST}(G, w)$. That is, find a minimum spanning tree for G, w

A greedy algorithm
Kruskal gave a very simple greedy algorithm to solve this problem:
Construct a minimum spanning tree by repeatedly adding in the smallest weighted edge from G which does not create a cycle in the solution constructed so far.

```
kruskalMST(graph G = (V, E), w : E → ℕ){
    C = E
    S = {}
    while(C ≠ {}){
        e=a minimum weight edge remaining in C
        C = C − {e}
        if(S ∪ {e} has no cycles)S = S ∪ {e}
    }
    return(S)
}
```

Proof of correctness
To establish correctness of kruskalMST, we need a couple of facts
Theorem: If (V, E) is a tree and $e \notin E$ then $(V, E \cup \{e\})$ has exactly 1 cycle and e is part of that cycle..
Theorem: The following are equivelent

$$(V, E) \text{ is a tree}$$

$$|V| = n, |E| = n - 1, G \text{ is connected}$$

$$|V| = n, |E| = n - 1, G \text{ has no cycles}$$

Theorem: kruskalMST is correct (ie it solves the MINIMUM-SPANNING-TREE problem).
Proof: We show that at each loop, the decision to add or not to add an edge to S is good. We need to make this idea more precise, so we introduce the following definitions.

- Let C_i be the state of C at the end of iteration i of the while loop
- Let S_i be the state of S at the end of iteration i of the while loop

- $\underline{S_i \text{ is promising}}$ if $\exists(V, S_i^*) \in \text{MST}(G, w)$ such that
  1. $S_i \subseteq S_i^*$ *We included only edges that we need*
  2. $S_i^* - S_i \subseteq C_i$ *We only left out edges that we don't need*

For what follows, we assume that $|V| = n$ and $|E| = m$.
The first thing to notice is that the kruskalMST eventually returns. To see this note that $|C_i| = m - i$ so C eventually becomes empty and control leaves the while loop.

We show by induction that $S_0, S_1, S_2, \ldots, S_m$ are all promising. S_m is promising $\Rightarrow \exists(V, S_m^*) \in \text{MST}(G, w)$ such that $S_m \subseteq S_m^*$ and $S_m^* - S_m = \{\}$ so $S_m^* = S_m$ and $(V, S_m) \in \text{MST}(G, w)$. That is, S_m consists of the edges in a minimum spanning tree for G, w.

Base case: $S_0 = \{\}$ is promising, to see this, let (V, S_0^*) be any member of $\text{MST}(G, w)$, then $S_0 \subseteq S_0^*$ and $S_0^* - S_0 \subseteq C_0 = E$.

Induction Hypothesis: Assume that S_i is promising, so we let $(V, S_i^*) \in \text{MST}(G, w)$ be such that $S_i \subseteq S_i^*$ and $S_i^* - S_i \subseteq C_i$.

Induction Step: Based on the Induction Hypothesis, we show that S_{i+1} is promising. Going through iteration $i + 1$ of the while loop, we have

$$e = \text{ an edge with minimum weight in} C_i$$

$$C_{i+1} = C_i - \{e\}$$

$$S_{i+1} = \begin{cases} S_i \cup \{e\} & \text{if } S_i \cup \{e\} \text{ has no cycle} \\ S_i & \text{otherwise} \end{cases}$$

(1) $S_{i+1} = S_i$: So $S_i \cup \{e\}$ contains a cycle, then $e \notin S_i^*$ (otherwise S_i^* would contain a cycle). Let $S_{i+1}^* = S_i^*$, then $S_i = S_{i+1} \subseteq S_{i+1}^* = S_i^*$ and $S_{i+1}^* - S_{i+1} \subseteq C_{i+1}$. So S_{i+1} is promising.

(2) $S_{i+1} \neq S_i$: So $S_i \cup \{e\}$ does not contain a cycle. So $S_{i+1} = S_i \cup \{e\}$. We have two subcases to consider

a. $e \in S_i^*$
Let $S_{i+1}^* = S_i^*$, then $S_{i+1} \subseteq S_{i+1}^*$ and $S_{i+1}^* - S_{i+1} = S_i^* - (S_i \cup \{e\}) \subseteq C_{i+1}$ and so S_{i+1} is promising.

b. $e \notin E^*$
(V, S_i^*) is a tree so $(V, S_i^* \cup \{e\})$ contains exactly one cycle (call it Z) and this cycle contains edge e. S_{i+1} contains no cycles so Z contains an edge $e' \in S_i^* - S_{i+1} \subseteq C_{i+1}$. Let $S_{i+1}^* = S_i^* \cup \{e\} - \{e'\}$. Then (V, S_{i+1}^*) is a tree (we added e to form a cycle and then removed e' to break the cycle). $w((V, S_{i+1}^*)) = w((V, S_i^*)) + w(e) - w(e') \leq w((V, S_i^*))$ since $w(e) \leq w(e')$. Now $(V, S_i^*) \in \text{MST}(G, w)$ so we must have $w((V, S_{i+1}^*)) = w((V, S_i^*))$. Summing up, $(V, S_{i+1}^*) \in \text{MST}(G, w)$ with the properties $S_{i+1} \subseteq S_{i+1}^*$ and $S_{i+1}^* - S_{i+1} \subseteq C_{i+1}$ and so S_{i+1} is promising.

By induction, $S_0, \ldots, S_m$ are all promising.