The proof of Kruskal's algorithm

May 2, 2021

Kruskal's algorithm

Algorithm 1 Kruskal(G)

$F := \varnothing$
for each $v \in G.V$ **do**
  MAKE-SET(v)
end for
for each($u,\ v$) **in** $G.E$ ordered by weight($u,\ v$), increasing **do**
  if FIND-SET(u) $\neq$ FIND-SET(v) **then**
    $F := F \cup \{(u,\ v)\} \cup \{(v,\ u)\}$
    UNION(FIND-SET(u), FIND-SET(v))
  end if
end for
return F

Proof of correctness

0.1 Spanning tree

I show that the set of edges made by Kruskal's algorithm is a spanning tree.
Let $\boldsymbol{P}$ be a weighted connected graph and $\boldsymbol{K}$ be a subgraph of $\boldsymbol{P}$ created by Kruskal's algorithm. Kruskal's algorithm adds edges to connect different trees, so $\boldsymbol{K}$ has no cycle. Also, when $\boldsymbol{K}$ is generated by connecting two different trees, all the vertices are connected because the edges are selected from all the edges from $\boldsymbol{P}$. Therefore, $\boldsymbol{K}$ is the spanning tree of $\boldsymbol{P}$.

0.2 Minimality

Prove that the spanning tree $\boldsymbol{K}$ is the minimum spanning tree.
This proof uses reductio ad absurdum. Suppose $\boldsymbol{K}$ is the spanning tree of weighted connected graph $\boldsymbol{P}$ obtained by Kruskal's algorithm. And, let $\boldsymbol{Q_0}$ be the set of edges containing the largest number of edges of $\boldsymbol{K}$ among the minimum spanning tree of $\boldsymbol{P}$. By definiton, $\boldsymbol{K} \neq \boldsymbol{Q_0}$. As an example, consider $\boldsymbol{K}$ and $\boldsymbol{Q_0}$ as shown in the following figure.

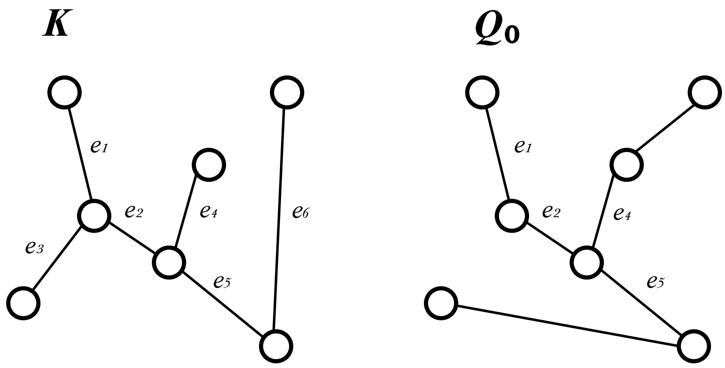

Figure1 $\{e_i\}_{i=1,\cdots,6}$: edge

Add to $\boldsymbol{Q_0}$ the edge f which is included in $\boldsymbol{K}$ but not in $\boldsymbol{Q_0}$. Then, an cycle is created in the set $\boldsymbol{Q_0} \cup \{f\}$ as shown in the figure below.

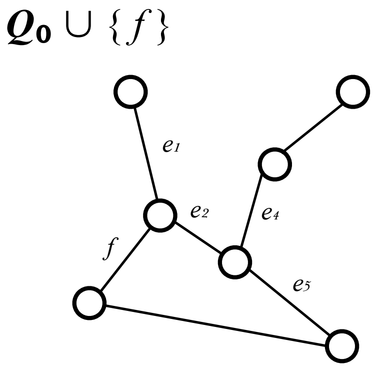

Figure2

$\boldsymbol{K}$ is a tree, and not all edges of this cycle are included in $\boldsymbol{K}$. In other words, this cycle has edges that are not included in $\boldsymbol{K}$. Let g be the edge that is arbitrarily selected from the edges that are not included $\boldsymbol{K}$, as shown in the figure below.

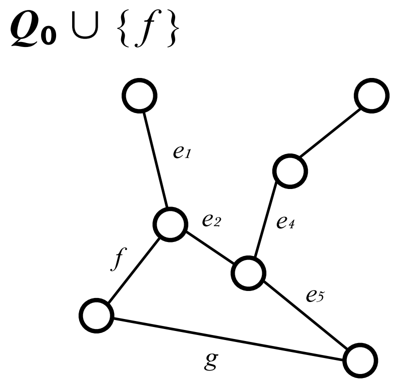

Figure3

Here, due to the characteristic of Kruskal's algorithm, f included in $\boldsymbol{K}$ is an edge added in ascending order of weight, so weight$(f) \leq$ weight(g) holds. Also, let $\boldsymbol{Q_1} = \boldsymbol{Q_0} \cup \{f\} \backslash \{g\}$ as shown in the figure below.

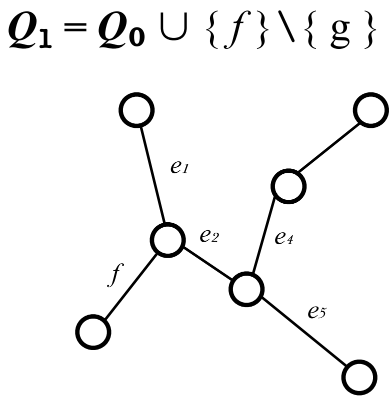

Figure4

Then, since $\boldsymbol{Q_1}$ contains the same number of edges as $\boldsymbol{Q_0}$ and weight$(f) \leq$ weight(g), $\boldsymbol{Q_1}$ is also minimum spanning tree. However, $\boldsymbol{Q_1}$ contains one more edge of $\boldsymbol{K}$ than $\boldsymbol{Q_0}$. Therefore, it contradicts the definiton $\boldsymbol{Q_0}$. Consequently, $\boldsymbol{K} = \boldsymbol{Q_0}$, and spanning tree obtained by Kruskal's algorithm is the minimum spanning tree.