

# Computer Science 331

## Analysis of Prim's Algorithm: Proof of a Key Lemma

### 1 Introduction

Consider the algorithm to find a minimum-cost spanning tree that was presented and analyzed during the last full week of classes. Recall that if  $\mathbf{G} = (\mathbf{V}, \mathbf{E})$  is a weighted undirected graph that is given as input and  $\mathbf{G}$  is connected then, following an initialization step, this algorithm uses a priority queue to choose vertices and edges that should be added to a subgraph

$$\mathbf{G}_b = (\mathbf{V}_b, \mathbf{E}_b)$$

of the input graph  $\mathbf{G}$ . As documented in the lecture notes on the analysis of this algorithm, it is reasonably easy to establish that each of the following properties is satisfied after both the beginning and end of each execution of the body of the `while` loop that makes up the second part of this algorithm:

- $\mathbf{G}_b$  is a *tree*.
- $\mathbf{G}_b$  is a subgraph of some *spanning tree*  $\mathbf{T}$  of  $\mathbf{G}$ .

The objective of this handout is to provide a proof of the next major property (that is not quite so easy to establish).

- $\mathbf{G}_b$  is a subgraph of some *minimum-cost spanning tree*  $\mathbf{T}$  of  $\mathbf{G}$ .

The *proof* of this claim will proceed by induction on the number of executions of the body of the `while` loop that have taken place, so far.

### 2 Basis

One can see, by an examination of the code, that  $\mathbf{V}_b = \emptyset$  and  $\mathbf{E}_b = \emptyset$  immediately before the first execution of the body of the `while` loop, while  $\mathbf{V}_b = \{\mathbf{s}\}$  and  $\mathbf{E}_b = \emptyset$  immediately after this execution of the loop body for some vertex  $\mathbf{s} \in \mathbf{V}$ .

In each case, since  $G_b = (V_b, E_b)$  does not include any edges at all, it is clear that  $G_b$  is a subgraph of *every* spanning tree of  $G$ . In particular,  $G_b$  is a subgraph of every *minimum-cost spanning tree* of  $G$  and (since there are only finitely many spanning trees, so that one must have minimal cost), at least one minimum-cost spanning tree  $T$  of  $G$  does exist.

### 3 Inductive Step

#### 3.1 Notation

Suppose now that  $k \geq 1$  and that the subgraph  $G_b = (V_b, E_b)$  produced after  $k$  executions of the body of the `while` loop is a subgraph of some minimum-cost spanning tree  $T$  of  $G$ .

Suppose, as well, that the body of the loop is executed (at least) one more time. Then, of course,  $G_b$  is also the subgraph that has been produced at the *beginning* of the  $k + 1^{\text{st}}$  execution of the loop body. Let

$$G'_b = (V'_b, E'_b)$$

be the graph that is produced *after* the  $k + 1^{\text{st}}$  execution of the loop body.

#### 3.2 Proof that the New Graph is a Subgraph

In this case, one can see by inspection of the code that

$$V'_b = V_b \cup \{v\} \tag{1}$$

for some vertex  $v$  such that  $v \notin V_b$ , and

$$E'_b = E_b \cup \{(u, v)\} \tag{2}$$

for some vertex  $u \in V$ . In particular,  $u = \pi(v)$ . As noted during the discussion of this algorithm in class, one can show (without too much trouble) that  $(\pi(v), v) \in E$  and  $\pi(v)$  is **black** at this point in the computation, so that  $\pi(v) \in V_b$  and  $G'_b$  is a subgraph of  $G$ .

#### 3.3 Proof that the New Graph is a Subgraph of a Minimum-Cost Spanning Tree

In order to complete this proof it is necessary and sufficient to show that there exists a minimum-cost spanning tree  $T'$  of  $G$  such that  $G'_b$  is a subgraph of  $T'$ .

*Two cases* can be identified and should be considered: Either the edge  $(u, v)$  is an edge in  $T$  or it is not.

### 3.3.1 Case: The New Edge is Included in the Current Spanning Tree

Suppose that the edge  $(u, v)$  is an edge in  $T$ . Then, since every vertex in  $V$  is included in  $T$  and every edge in  $E_b$  is included in  $T$  (because  $G_b$  is a subgraph of  $T$ ), it follows by the definition of  $E'_b$  (given at line (2), above) that  $G'_b$  is a subgraph of  $T$  in this case.

It therefore suffices to set  $T' = T$  in this case in order to complete the proof.

### 3.3.2 Case: The New Edge is Not Included in the Current Spanning Tree

Suppose, instead, that the edge  $(u, v)$  is *not* an edge in  $T$ . Recall again that, since  $T$  is a spanning tree of  $G$ ,  $T$  includes all the vertices that  $G$  does; let

$$T = (V, E_T) \tag{3}$$

and consider the graph

$$U = (V, E_U) \quad \text{where } E_U = E_T \cup \{(u, v)\}. \tag{4}$$

**Claim 1.**  *$U$  is a subgraph of  $G$  that contains a cycle.*

*Proof.* First, note that  $U$  includes the same set of vertices as  $G$ . Since  $E_T \subseteq E$  (because  $T$  is a subgraph of  $G$ ) and since  $(u, v) \in E$ ,  $E_U \subseteq E$ , so that  $U$  is a subgraph of  $G$ .

It follows by Corollary 5 in the lecture notes on Trees, Spanning Trees, and Subgraphs that, since  $T$  is a spanning tree of  $G$ ,  $|E_T| = |V| - 1$ .

Now, since  $(u, v) \notin E_T$  it follows by the definition of  $U$ , given at line (4) above, that  $U$  includes  $|V|$  vertices and  $(|V| - 1) + 1 = |V|$  edges. It follows by Lemma 4 in the above set of notes that  $U$  contains a cycle, as claimed.  $\square$

**Claim 2.**  *$U$  contains a cycle that includes the edge  $(u, v)$ .*

*Proof.* It follows by Claim 1, above, that  $U$  includes some cycle

$$(u_1, u_2), (u_2, u_3), \dots, (u_{k-1}, u_k), (u_k, u_1).$$

Suppose that the edge  $(u, v)$  is *not* one of these edges; then, since

$$E_U = E_T \cup \{(u, v)\}$$

it follows that each of the edges  $(u_i, u_{i+1})$  (for  $1 \leq i \leq k - 1$ ) and  $(u_k, u_1)$  belongs to the set  $E_T$ . Thus the above cycle is also a cycle in  $T$  — which is impossible, since  $T$  is a *tree*.

Thus the edge  $(u, v)$  is included the above cycle, so that  $U$  has a cycle including the edge  $(u, v)$ , as claimed.  $\square$

Notice that, when listing the sequence of vertices in any cycle, we may begin with any one of the vertices that are to be listed. Reversing the *order* of the listing of vertices (which is also possible for an undirected graph), if necessary, we may now assume, by an application of Claim 2, that  $U$  includes a cycle

$$(\mathbf{u}_1, \mathbf{u}_2), (\mathbf{u}_2, \mathbf{u}_3), \dots, (\mathbf{u}_{k-1}, \mathbf{u}_k), (\mathbf{u}_k, \mathbf{u}_1) \quad (5)$$

such that  $\mathbf{v} = \mathbf{u}_1$  and  $\mathbf{u} = \mathbf{u}_2$ . Recall that, since this is a cycle,  $k \geq 3$ . We may also assume (by considering the *shortest* cycle that begins with  $\mathbf{v}$  and  $\mathbf{u}$ ) that the vertices  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$  are *distinct* — that is, this is a “simple cycle.”

Consider the colours of the vertices in this cycle at the *beginning* of the  $k + 1^{\text{st}}$  execution of the body of the loop:  $\mathbf{u}_1 = \mathbf{v}$  is **grey** at this point because  $\mathbf{v}$  is stored on the priority queue and, as discussed in the *analysis of the algorithm*,  $\mathbf{u}_2 = \mathbf{u} = \pi(\mathbf{v})$  is a **black** node at this point.

Recall as well that the neighbours of all **black** nodes are either **black** or **grey**.

Considering the colours of each of the vertices  $\mathbf{u}_3, \mathbf{u}_4, \dots$  in turn, until a **grey** vertex is found, we see that exactly one of the following two cases must apply. Vertices  $\mathbf{w}$  and  $\mathbf{x}$  will be defined in each case.

1. The vertex  $\mathbf{u}_i$  is **black** for  $2 \leq i \leq k$ , so that  $\mathbf{v} = \mathbf{u}_1$  is the only **grey** node in this cycle. Let  $\mathbf{w} = \mathbf{u}_k$  and let  $\mathbf{x} = \mathbf{u}_1 = \mathbf{v}$  in this case.
2. There exists an integer  $\ell$  such that  $3 \leq \ell \leq k$ ,  $\mathbf{u}_j$  is **black** for  $2 \leq j \leq \ell - 1$ , and such that  $\mathbf{u}_\ell$  is **grey**. Let  $\mathbf{w} = \mathbf{u}_{\ell-1}$  and let  $\mathbf{x} = \mathbf{u}_\ell$  in this case.

**Claim 3.** *The vertices  $\mathbf{w}$  and  $\mathbf{x}$  satisfy each of the following properties.*

- (a) *The vertex  $\mathbf{w}$  is **black** at the beginning of the  $k + 1^{\text{st}}$  execution of the body of the *while* loop.*
- (b) *The vertex  $\mathbf{x}$  is **grey** at the beginning of the  $k + 1^{\text{st}}$  execution of the body of the *while* loop.*
- (c) *The edge  $(\mathbf{w}, \mathbf{x})$  is not an edge in  $G'_b$ . That is,  $(\mathbf{w}, \mathbf{x}) \notin E'_b$ .*

*Proof.* Properties (a) and (b) are easily established by a consideration of the definition of  $\mathbf{w}$  and  $\mathbf{x}$  in each of the two cases that are listed above.

Suppose that the first case is applicable, so that  $\mathbf{x} = \mathbf{u}_1 = \mathbf{v}$  and  $\mathbf{w} = \mathbf{u}_k$ . Since  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$  are distinct,  $k \geq 3$ ,  $\mathbf{u} = \mathbf{u}_2$  and  $\mathbf{w} = \mathbf{u}_k$ ,  $\mathbf{w} \neq \mathbf{u}$ .

Now notice that, since  $\mathbf{v}$  is being added to the set of vertices to be included in the tree  $G'_b$  during this execution of the loop body (at the same time as the edge  $(\mathbf{u}, \mathbf{v})$  is being added),

the degree of  $v$  in  $G'_b$  is *one* and  $u$  is the *only* neighbour of  $v$  in  $G'_b$ . Since  $w \neq u$   $w$  cannot be a neighbour of  $v$  in  $G'_b$  as well, so the edge  $(w, v)$  (which is the same as the edge  $(w, x)$ ) *does not* belong to  $G'_b$ . That is, property (c) holds.

Suppose, instead, that the second case is applicable. In this case  $x = u_\ell$  is a **grey** node that is different from  $v$ . Since  $x$  is grey at this point  $x \notin V_b$ , and  $x \neq v$ , so  $x \notin V'_b$ . It is clearly impossible for  $(w, x)$  to belong to  $G'_b$  in this case. Thus property (c) also holds in this case as well.  $\square$

We are now ready to define the minimum-cost spanning tree  $T'$  that will be used to complete the proof for this case: Let

$$T' = (V, E'_T) \quad \text{where } E'_T = (E_T \setminus \{(w, x)\}) \cup \{(u, v)\} \quad (6)$$

That is, the edges in  $T'$  are all of the edges in  $T$  *except for*  $(w, x)$ , together with the new edge  $(u, v)$ .

**Claim 4.**  *$T'$  is a subgraph of  $G$  with  $|V|$  vertices and  $|V| - 1$  edges.*

*Proof.* It is clear by inspection of the definitions of  $T'$  and  $U$  (at lines (6) and (4), respectively) that  $T'$  is a subgraph of  $U$ . Since  $U$  is a subgraph of  $G$  it is easy to show that  $T'$  is a subgraph of  $G$  as well.

Since the set of vertices in  $T'$  is  $V$  it is clear that there  $|V|$  edges in  $T'$ .

To see that the number of edges in  $T'$  is  $|V| - 1$ , note first that, since  $(u, v)$  is an edge in  $G'_b$  while  $(w, x)$  is not an edge in this graph, the edges  $(u, w)$  and  $(w, x)$  are clearly distinct. It follows that  $(w, x)$  is an edge in  $T$ , since this is an edge in  $U$  (by the choice of  $w$  and  $x$ ) and since  $(u, v)$  is the only edge in  $U$  that is *not* also an edge in  $T$ . That is,  $(w, x) \in E_T$ .

On the other hand,  $(u, v) \notin E_T$  under the case that we are now considering.

It now follows by the definition of  $T'$  at line (6) that

$$|E'_T| = (|E_T| - 1) + 1 = |E_T|,$$

and  $|E_T| = |V| - 1$  since  $T$  is a spanning tree of  $G$ .  $\square$

**Claim 5.**  *$T'$  is a connected graph.*

*Proof.* It is necessary and sufficient to prove that there is a path from  $y$  to  $z$  *consisting of edges in  $T'$*  for each pair of distinct vertices  $y, z \in V$ .

Note that, since  $T$  is a spanning tree of  $G$ ,  $T$  is connected and includes the same set of vertices as  $T'$ , so there is a path

$$y = w_1, w_2, \dots, w_h = z \quad (7)$$

from  $y$  to  $z$  in  $T$ . That is,  $(w_i, w_{i+1}) \in E_T$  for  $1 \leq i \leq h - 1$ . We may also assume that this is a “simple path,” that is,  $w_1, w_2, \dots, w_h$  are distinct.

Now, either the edge  $(w, x)$  is included in the above path, or it is not. These cases will be considered separately.

*Case:  $(w, x)$  is included in the above path:* Recall that  $(w, x)$  is one of the edges included in the cycle

$$u_1, u_2, \dots, u_k$$

in the graph  $U$  that is introduced at line (5), above, and that all the edges in this cycle are distinct. Since  $(w, x)$  is the only edge included in  $U$  that is not also included in  $T'$ , *all of the other edges* in this cycle can be used to form a path between  $w$  and  $x$  in  $T'$ .

A path from  $y$  to  $z$  in  $T'$  can now be formed from the path shown at line (7) by replacing each occurrence of the edge  $(w, x)$  with the path between  $w$  and  $x$  in  $T'$  (in the appropriate direction) that has just been described.

*Case:  $(w, x)$  is not included in the above path:* In this case, each edge in this path is an edge included in  $T'$ , so that the path shown at line (7) is a path from  $y$  to  $z$  in  $T'$ .

Since  $y$  and  $z$  were arbitrarily chosen from  $V$  it follows that  $T'$  is connected, as claimed.  $\square$

**Claim 6.**  $T'$  is a tree.

*Proof.* This follows directly from Claims 4, 5, and from Lemma 7 in the online lecture notes on Trees, Spanning Trees, and Subgraphs.  $\square$

**Claim 7.**  $T'$  is a spanning tree of  $G$ .

*Proof.*  $T'$  is a subgraph of  $G$ , by Claim 4. This subgraph is a tree, by Claim 6, and it includes all the vertices in  $G$ . It is therefore a spanning tree of  $G$ .  $\square$

**Claim 8.**  $w((u, x)) \geq w((u, v))$ .

*Proof.* Recall that  $w$  and  $x$  were defined in one of two ways, depending on the location of  $w$  in the cycle

$$u_1, u_2, \dots, u_k$$

in  $U$  introduced at line (5) above. Each of these will be considered separately.

*Case:  $x = u_1 = v$  and  $w = u_k \neq u_2 = u$ :* Recall, from the analysis of this algorithm that  $u = \pi[v]$ ,

$$w((u, v)) = d[v] = \min_{\substack{y \in V \\ \text{colour}[y] = \text{black} \\ (y, v) \in E}} w((y, v)).$$

It follows by the choice of  $\mathbf{w}$  (as described above) that the colour of  $\mathbf{w}$  is **black** at the beginning of the  $k+1^{\text{st}}$  execution of the body of the **while** loop and that  $(\mathbf{w}, \mathbf{v}) = (\mathbf{w}, \mathbf{x}) \in \mathbf{E}$ . It therefore follows by the above relationship that

$$\mathbf{w}((\mathbf{u}, \mathbf{v})) = \mathbf{d}[\mathbf{v}] = \mathbf{d}[\mathbf{w}] \leq \mathbf{w}((\mathbf{w}, \mathbf{x})),$$

as desired.

*Case:*  $\mathbf{w} = \mathbf{u}_{\ell-1}$  and  $\mathbf{x} = \mathbf{u}_{\ell}$  where  $3 \leq \ell \leq k$ . Recall that  $\mathbf{w}$  is a **black** node and that  $\mathbf{x}$  is a **grey** node that is different from  $\mathbf{v}$ , at the beginning of the  $k+1^{\text{st}}$  execution of the body of the **while** loop, in this case.

Note that  $\mathbf{w}((\mathbf{u}, \mathbf{v})) = \mathbf{d}[\mathbf{v}]$  in this case (and, for the same reason) once again. Since the element  $\mathbf{v}$  is removed from the priority queue at the beginning of this execution of the body of the loop (with priority  $\mathbf{d}[\mathbf{v}]$  at this point), and since  $\mathbf{x}$  is also stored on the priority queue with priority  $\mathbf{d}[\mathbf{x}]$ , but *not* removed at this point, it is clear that  $\mathbf{d}[\mathbf{v}] \leq \mathbf{d}[\mathbf{x}]$ .

Finally, it follows by the notes on the analysis of this algorithm that, since  $\mathbf{w}$  is **black** and  $(\mathbf{w}, \mathbf{x}) \in \mathbf{E}$ ,

$$\mathbf{w}((\mathbf{w}, \mathbf{x})) \geq \mathbf{d}[\mathbf{x}] = \min_{\substack{y \in V \\ \text{colour}[y] = \text{black} \\ (y, \mathbf{x}) \in \mathbf{E}}} \mathbf{w}((y, \mathbf{x})).$$

Thus

$$\mathbf{w}(\mathbf{u}, \mathbf{v}) = \mathbf{d}[\mathbf{v}] \leq \mathbf{d}[\mathbf{x}] \leq \mathbf{w}(\mathbf{w}, \mathbf{x})$$

in this case as well. □

**Claim 9.**  $T'$  is a minimum-cost spanning tree of  $G$ .

*Proof.* Since  $T'$  is a spanning tree of  $G$ , by Claim 7, and since  $T$  is a minimum-cost spanning tree of  $G$ , by assumption,  $\mathbf{w}(T) \leq \mathbf{w}(T')$ . It is necessary and sufficient to show that  $\mathbf{w}(T') \leq \mathbf{w}(T)$  as well — for then  $\mathbf{w}(T') = \mathbf{w}(T)$  and *both* of these spanning trees are minimum-cost spanning trees of  $G$ .

Recall that the set  $\mathbf{E}_T$  of edges in  $T$  and the set  $\mathbf{E}'_T$  of edges in  $T'$  are related as follows (see, in particular, line (6) above):

$$\mathbf{E}'_T = (\mathbf{E}_T \setminus \{(\mathbf{w}, \mathbf{x})\}) \cup \{(\mathbf{u}, \mathbf{v})\}.$$

Furthermore,  $(\mathbf{w}, \mathbf{x}) \in \mathbf{E}_T$ .

Therefore

$$\begin{aligned}
w(\mathbf{T}') &= \sum_{(y,z) \in \mathbf{E}'_{\mathbf{T}}} w((y,z)) \\
&= \sum_{(y,z) \in \mathbf{E}_{\mathbf{T}}} w((y,z)) - w((\mathbf{w}, \mathbf{x})) + w((\mathbf{u}, \mathbf{v})) \quad (\text{since } \mathbf{E}'_{\mathbf{T}} = (\mathbf{E}_{\mathbf{T}} \setminus \{(\mathbf{w}, \mathbf{x})\}) \cup \{(\mathbf{u}, \mathbf{v})\}) \\
&= w(\mathbf{T}) - (w((\mathbf{w}, \mathbf{x})) - w((\mathbf{u}, \mathbf{v}))) \\
&\leq w(\mathbf{T}) \quad (\text{since } w((\mathbf{w}, \mathbf{x})) \geq w((\mathbf{u}, \mathbf{v})) \text{ by Claim 8, above}). \quad \square
\end{aligned}$$

**Claim 10.**  $\mathbf{G}'_{\mathbf{b}}$  is a subgraph of  $\mathbf{T}'$ .

*Proof.* Since the set of vertices in  $\mathbf{G}'_{\mathbf{b}}$  is a subset of  $\mathbf{V}$  and  $\mathbf{T}'$  includes all the vertices in  $\mathbf{V}$ , it is necessary and sufficient to show that every edge in  $\mathbf{G}'_{\mathbf{b}}$  is an edge in  $\mathbf{T}'$ .

The set of edges in  $\mathbf{G}'$  is

$$\begin{aligned}
\mathbf{E}'_{\mathbf{b}} &= \mathbf{E}_{\mathbf{b}} \cup \{(\mathbf{u}, \mathbf{v})\} && (\text{as shown at line (2)}) \\
&\subseteq \mathbf{E}_{\mathbf{T}} \cup \{(\mathbf{u}, \mathbf{v})\} && (\text{since } \mathbf{G}_{\mathbf{b}} \text{ is a subgraph of } \mathbf{T}, \text{ so that } \mathbf{E}_{\mathbf{b}} \subseteq \mathbf{E}_{\mathbf{T}}).
\end{aligned}$$

Since  $\mathbf{E}'_{\mathbf{T}} = (\mathbf{E}_{\mathbf{T}} \cup \{(\mathbf{u}, \mathbf{v})\}) \setminus \{(\mathbf{w}, \mathbf{x})\}$ , as shown at line (6), it is sufficient to establish that  $(\mathbf{w}, \mathbf{x}) \notin \mathbf{E}'_{\mathbf{b}}$  in order to complete the proof. This has already been established (and is included as part of Claim 3, above).  $\square$

It follows by Claims 9 and 10 that  $\mathbf{G}'_{\mathbf{b}}$  is a subgraph of minimum-cost spanning tree of  $\mathbf{G}$ , as required to complete the proof in this case.