Complexity

The STEINER TREE problem is defined as follows:

STEINER TREE
INSTANCE: A graph $G = (V, E)$, a subset $R \subseteq V$ of vertices, a number $k \in \mathbb{N}$
QUESTION: Is there a subtree of G that includes all the vertices of R and contains at most k edges ?

To start with an example, let us define $G = (V, E)$ with $V = \{1, 2, \dots, 12\}$ and

$$E = \{(1,2),(2,3),(3,4),(1,5),(5,6),(2,6),\\ (6,7),(3,7),(7,8),(4,8),(5,9),(9,10),(6,10),(10,11),(7,11),(11,12),(8,12)\}$$

G is indeed a "grid" with 3 rows and 4 columns.

Question 1 *Draw a Steiner tree for G if $R = V$ and $k = 13$. What if $R = \{1, 2, 5, 8, 11, 12\}$ and $k = 10$? What if $R = V$ and $k = 10$? What is the number of edges of a Steiner tree when $R = V$? What do you think of instances in which G is a connected graph and $k = |V|$?*

We want to show that STEINER TREE is NP-complete. Note that an instance of STEINER TREE is defined by three elements: G, R and k. We proceed by a reduction from the problem EXACT COVER BY 3-SETS. It is defined as follows:

EXACT COVER BY 3-SETS
INSTANCE: Numbers $n \in \mathbb{N}$ and $m \in \mathbb{N}$, a set $U = \{1, 2, \dots, 3m\}$, a family $F = \{S_1, S_2, \dots, S_n\}$ of subsets of U, such that $\forall i \in \{1, \dots, n\}, |S_i| = 3$
QUESTION: Is there a collection $C' \subseteq F$ such that every element of U appears in exactly one member of C' ?

Question 2 *Let $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and $F = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7\}$ with*

$$S_1 = \{6, 8, 9\}, S_2 = \{3, 4, 7\}, S_3 = \{2, 3, 5\},$$

$$S_4 = \{1, 2, 5\}, S_5 = \{3, 6, 7\} S_6 = \{1, 2, 3\}, S_7 = \{4, 5, 6\}$$

Show that this is a positive instance. What if $S_1 = \{7, 8, 9\}$? Given that every element of U appears in exactly one member of C', how many subsets **must** *be in C' ?*

Given an instance of EXACT COVER BY 3-SETS defined by $U = \{1, 2, \ldots, 3m\}$ and $F = \{S_1, \ldots, S_n\}$, we build an instance of STEINER TREE as follows:

- $V = \{r\} \cup \{c_1, c_2, \ldots, c_n\} \cup \{x_1, x_2, \ldots, x_{3m}\}$ (r is just a name for a vertex).
- $E = \{(r, c_1), (r, c_2), \ldots, (r, c_n)\} \cup \left(\bigcup_{j \in S_k} \{(c_k, x_j)\} \right)$ (there is an edge between vertex c_k and vertex x_j if the element $j \in U$ belongs to subset S_k)
- $R = \{r, x_1, x_2, \ldots, x_{3m}\}$
- $k = 4m$

Question 3 *With the instance of EXACT COVER BY 3-SETS defined in question 2, represent the corresponding instance of STEINER TREE (recall that 4 elements need to be defined). Prove that this instance of STEINER TREE is positive by drawing the corresponding tree. Prove that the construction of the instance of STEINER TREE is polynomial (Step 1).*

Coming back to general case (not the example), we now assume that we have a positive instance of EXACT COVER BY 3-SETS (Step 2.1). Then there is $C' \subseteq F$ wich satisfies the constraint of EXACT COVER BY 3-SETS.

Question 4 *(Here, your answer must general and not limited to an example). What is the size of C' ? Define a Steiner tree for the instance of STEINER TREE defined above. What is the number of edges in this tree T ? Conclude on the instance of STEINER TREE.*

We now assume that we have an instance of STEINER TREE which is positive (general case, Step 2.2). Then there is a tree T with at most $k = 4m$ edges which contains the vertices $R = \{r, x_1, x_2, \ldots, x_{3m}\}$.

Question 5 *Show by induction that, for $n \geq 1$ a tree with at most n edges has at most $n+1$ vertices.*

Question 6 *What is the maximal number of vertices in T ? What is the maximal number of vertices $c_i, i \in \{1, \ldots, n\}$ in T (we will call it $\max_c$) ? What if there were $\max_c - 1$ vertices c_i in T ? What is the exact number of vertices c_i in T ? Show that if two of these vertices (say c_i and c_l) both share an edge with the same vertex x_p, tree T could not contain all the vertices of R ? Based on the tree T, what is the subset $C' \subset F$ we can define to show that the instance of EXACT COVER BY 3-SETS is positive.*

Question 7 *Show that STEINER TREE is in NP. Conclude*

Q1: Positive instance. Positive instance. If $R = V$ and $k = 10$, no enough edges for all vertices of V. Number of edges if $R = V$ is $|V| - 1$. If $R = V$ and $k = |V|$, there is always a Steiner tree.

Q2: $C' = \{S_1, S_2, S_4\}$. With $S_1 = \{7, 8, 9\}$, $C' = \{S_1, S_6, S_7\}$. As $|U| = 3m$ and $|S_i| = 3$, we must have $|C'| = m$.

Q3: Do not forget $R = \{v, x_1, x_2, \ldots, x_9\}$ and $k = 4m = 12$. Vertices $\{v, c_1, c_2, c_4, x_1, \ldots, x_9\}$ form a tree with 3+3*3=12 edges, thus these is a positive instance of STEINER TREE. The number of vertices of the instance we have built is $n + 3m + 1$ thus the construction is polynomial (indeed, the instance contains $n + 3n = 4n$ edges).

Q4: $|C'| = m$. In addition to the vertices in R, a Steiner tree T can be defined by choosing vertices c_i if $S_i \in C'$: T is composed of the set of vertices $V_T = \{v, x_1, x_2, \ldots, x_{3m}\} \cup \left(\bigcup \{c_i | S_i \in C'\} \right)$. Let us show this. Note first that vertex R shares a edge with all vertices c_i. If T is not a Steiner tree this must come from a vertex x_j which shares no edge with any $c_i \in V_T$. This would mean that element $j \in U$ does not belong to any $S_i \in C'$, which is false (QED). Number of vertices of T is $m + 3m = 4m$. Finally, the instance of STEINER TREE is positive.

Q5: For $n = 1$, an edge clearly has 1+1 vertices. Assume this is true for a tree of n edges: it contains at most $n + 1$ vertices. Consider a tree with at most n edges, we will add only one vertex, which total number is now at most $n + 1 + 1$.

Q6: As T has at most $4m$ edges, it contains at most $4m + 1$ vertices. There are $3m + 1$ vertices in R, thus T contains at most m vertices c_i: $\max_c = m$. With $m - 1$ vertices c_i in T, as they share 3 edges with vertices x_j, the maximal number of x_j in T would be $3m - 3 < 3m$. Thus we must have m vertices c_i in T. If the two vertices c_i and c_l both share an edge with a vertex x_j, then the total number of vertices x_p would be at most $3(m-2)+5 = 3m-1 < 3m$. Thus it's not possible. Consequently, we can define $C' = \{S_i | c_i \in T\}$

Q6: certificate in the case of STEINER TREE = a tree T in the graph G. We have to check that:

- it contains all the elements of R: $O(|V|)$
- it contains at most k edges: $O(|E|)$
- it contains no cycle. Note a cycle would have a length of at most $|V|$ (contain at most $|V|$ edges). Starting from an arbitrary vertex, we generate the list of its neighbors (at most $|V|$), then the list of their neighbors... We repeat this at most $|V|$ times. Finally it takes at most $O(|V|^2)$.