Applied Algorithms. Prof. Tami Tamir
Missing Proofs, lecture 5. July 5th.

Covering a tree

Theorem: The algorithm is optimal, that is, it uses the minimal possible number of centers.

Proof: Let k be the number of centers determined by the algorithm. We show that there exists a set H of k nodes such that for every two nodes v1,v2 in H it holds that d(v1,v2)> s_v1+s_v2. (*)

Given that such a set exists, at least k centers are required, since otherwise, by the pigeonhole principle, there exists a center that covers more than one node in the set, which is impossible.

The set H is defined as follows: For every center c set by the algorithm, add to H the node for which d(v,c)-s_v and is the first node processed by the algorithm which among the nodes covered by c.

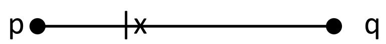
Proof of the property (*): W.l.o.g, the center covering v1 was set first. Since G is a tree, there is only one path between v1 and v2. By the way the algorithm proceeds, the center c1 covering v1 is located along this path (on the edge connecting v1 to the rest of the graph). Since it does not cover v2, it must be that that d(v1,v2)=d(v1,c1)+d(c1,v2)=s_v1+d(c1,v2) > s_v1+s_v2.

Local Center Property (slide 34) :

For the local center, x_e , on an edge (p,q),

$m(x_e) \geq \frac{m(p)+m(q)-c(p,q)}{2}$, where, c(p, q) denotes the length of (p,q).

Proof: Consider a point x on e with distance x from p.



For x=0, the point is p. For x=c(p,q) the point is q.

d(x,p)=x and d(x,p)=c(p,q)-x.

We have m(p) ≤ m(x)+x and m(q) ≤ m(x)+c(p,q)-x, since it is always possible to reach p by a path to x plus d(p,x), and it is always possible to reach q by a path to x plus d(x,q).

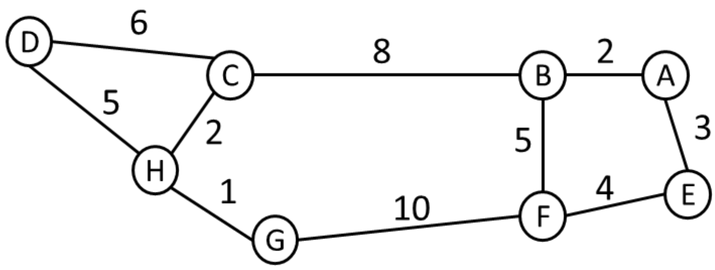
Summing up the above inequations, we get m(q)+m(p) ≤ 2m(x)+x –c(p,q)-x =2m(x)+c(p,q).

The above is valid for every x, in particular, for $x=x_e$.

By switching sides, we get $m(x_e) \geq \frac{m(p)+m(q)-c(p,q)}{2}$

Example of 2-approximation to k-center

Let k=3



Assume that A is selected as a first node.

The furthest node is D – since d(D,A)=16.   So D is added.

The next furthest node is F – since d(F,{A,B})=7

X_3 = {A,D,F}. The value of this solution is 6 – d(C,X_3)=d(G,X_3)=6.

A better solution is {B,F,H} – its value is 5.

For k=2, the algorithm halts with {A,D}, d(f,X2})=7.

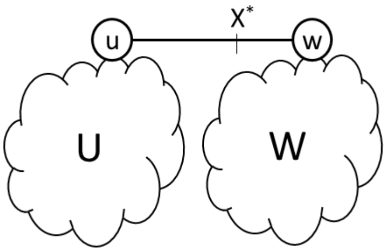
OPT(k=2)={B,H}, value =5 (achieved by E and F).

Hakimi's Theorem:

At least one optimal set of k-medians exist solely on the nodes of G.

Proof: Assume that k=1.

Let x* be the optimal 1-median. If x* is a node – we are done. Otherwise, x*is located on some edge (u,w). Split the graph's nodes into two disjoint sets $V=U\cup W$ such that $v\in U$ if and only if a shortest path from v to x* passes through v (otherwise, $v\in W$)



Compare $\sum_{v\in U} h(v)$ with $\sum_{v\in W} h(v)$. Assume w.l.o.g. that $\sum_{v\in U} h(v) \geq \sum_{v\in W} h(v)$. We show that x* can be replaced by the node u without hurting the objective function value.

$$J(x^*) = \sum_{v\in V} h(v)d(v,x^*) = \sum_{v\in U} h(v)[d(v,u)+d(u,x^*)] + \sum_{v\in W} h(v)d(v,x^*) =$$

$$\sum_{v\in U} h(v)d(v,u) + \sum_{v\in U} h(v)d(u,x^*) + \sum_{v\in W} h(v)d(v,x^*) \geq$$

assumption

$$\geq \sum_{v\in U} h(v)d(v,u) + \sum_{v\in W} h(v)d(u,x^*) + \sum_{v\in W} h(v)d(v,x^*) =$$

$$\geq \sum_{v\in U} h(v)d(v,u) + \sum_{v\in W} h(v)[d(u,x^*)+d(v,x^*)] \geq$$

Triangle inequality

$$\geq \sum_{v\in U} h(v)d(v,u) + \sum_{v\in W} h(v)d(u,v) = \sum_{v\in V} h(v)d(u,v) = J(u).$$

For k>1, a similar approach works for every facility located along an edge. Instead of splitting the whole graph, we split the set of nodes serviced by the facility into two sets according to the endpoint of the edge from which they reach the local center.