Lecture 7: k-center Problem

1 Introduction

The k-center problem is a facility location problem using complete graphs with weighted edges.

Definition 1.0.1 Let \( G = (V, E) \) be a complete undirected graph with edge costs satisfying the triangle inequality. Let \( C_{ij} \) be the shortest path distance between the nodes \( i \) and \( j \). The problem is to find a subset of the nodes \( S \subseteq V \), with \( |S| = k \) such that, the longest distance, over all nodes in the graph, of the distance of a node from its nearest node in \( S \) is minimized.

In math terms the problem can be expressed as \( S \subseteq V; |S| = k \) such that the \( \text{cost}(S) = \max_{i \in V} \min_{j \in S} C_{i,j} \) is minimized.

Reorder edges \( e_1, e_2, \ldots, e_m \) so that \( \text{cost}(e_1) \leq \text{cost}(e_2) \leq \text{cost}(e_3) \leq \ldots \leq \text{cost}(e_m) \). Consider the subgraph of ”cheap” edges.

Definition 1.0.2 \( G_1 = (V, E_1) \) where \( E_1 \) is the edge (or edges) of lowest cost. If \( G_i \) contains edges \( \{e_1, e_2, \ldots, e_r\} \), we define \( G_{i+1} = (V, E_i + 1) \) to contain \( E_i \cup \{e_{r+1}\} \) plus any additional edges that cost the same as \( e_{r+1} \).

Definition 1.0.3 A Dominating Set of \( G \) is a subset \( S \subseteq V \) such that every node in \( V - S \) is adjacent to a node in \( S \).

The optimal solution to k-center is a dominating set in \( G \). For example, each node in a complete graph is by itself a dominating set.

Let the cost of the optimal solution to k-center be \( c^* \). Let \( e_{c1} \) be the first edge of cost \( c^* \) in \( \{e_1, \ldots, e_m\} \) and let \( e_{c2} \) be the last edge of cost \( c^* \) in \( \{e_1, \ldots, e_m\} \).

\(^1\)These notes were revised from scribe notes from Jeff Livingston in 2002.
Claim 1.0.4 There is a dominating set in $G_{c2}$ of size $k$ or less.

Claim 1.0.5 There is no dominating set in $G_{c2-1}$ of size $k$ or less. If there was, this dominating set would be a solution to the $k$-center problem of cost at most $e_{c2-1} < e_{c2}$.

From the two claims, we can conclude that the $k$-center problem with triangle inequality is equivalent to finding the smallest index $i$ such that $G_i$ has a dominating set of size $k$. This is only true when the triangle inequality is in effect. There is no known polynomial time algorithm to solve the $k$-center problem since $k$-center is NP-Hard. Our goal is to lower bound the size of a dominating set in $G_i$.

Definition 1.0.6 The square of a graph $H = (V, E)$, denote $H^2 = (V, E^2)$ has an edge between $i$ and $j$ iff there is a path of length 1 or 2 hops between $i$ and $j$.

Lemma 1.0.7 Given $H$, let $I$ be an independent set in $H^2$ then $|I| \leq \text{dom}(H)$ where $\text{dom}(H)$ denotes the size of the minimum cardinality dominating set in $H$.

Proof Let $D$ be a minimum cardinality dominating set in $H$. For each $d \in D$, its neighborhood is a complete subgraph in $H^2$. A complete subgraph is also called a clique. $H^2$ contains $|D|$ cliques spanning all nodes. Any independent set can pick at most 1 node per clique. So $|I| \leq |D|$.

2 Algorithm

1. Construct $G^2_1, \ldots, G^2_m$.
2. Compute a maximal independent set (MIS) $M_i$ in each graph $G^2_i$.
3. Find the smallest index $i$ such that the $|M_i| \leq k$, say $M_j$.
4. Return $M_j$. 
Lemma 2.0.8 For j as defined in the algorithm, cost(e_j) ≤ c^*

Proof For every i < j, we have |M_i| > k otherwise we would have output j. Since dom(G_i) ≥ |M_i| by lemma 1.0.7. This implies that the size of the dom(G_i) > k. The first index for which the k-center solution forms a dominating set > j. So c^* ≥ cost(e_j).

Theorem 2.0.9 The algorithm returns a solution of cost at most 2*OPT.

Proof First observe the MIS S in H_2 is also a dominating set in H_2, and has size exactly k. (If S is not a dominating set, i.e. some vertex v is not in S and it has no neighbor in S. But this implies S∪{v} is an IS! There is a contradiction. S is a MIS.) Thus if we have a MIS equal to the dominating set G^2_i called D, then every node is on a path of length at most 2 from some node in D in G_i. Since each edge in G^2_i has cost < c^*, each node is on a path of length at most 2c^* from some center, and the result follows.