Prof. Lenore Cowen
Scribe: Yang Wang

Lecture 5: k-center Problem

1 Introduction

The k-center problem with triangle inequality is to place k centers so that no one has to go too far to get to their closest center.

k-center Problem: Let $G = (V, E)$ be a complete undirected graph whose edges are shortest paths between each pair of nodes. Let C_{ij} denote the path distance between i and j. The problem is to output a subset of nodes $S \subseteq V$ with $|S| = k$, such that the longest distance of a node to its nearest node in S is minimal. That is to say we want to find $S \subseteq V$, $|S| = k$ such that the value $cost(S) = max_j min_{i \in S} C_{ij}$ is minimized.

We assume that it satisfies the triangle inequality, then reorder the edge e_1, e_2..., e_m, such that $cost(e_1) \leq cost(e_2) \leq ... \leq cost(e_m)$. Let $G_i = (V, E_i)$, where $E_i = \{e_1, e_2, ..., e_i\}$. Note that our orginial graph G is now G_{e_m}.

Definition 1.0.1 *A dominating set of G is a subset $S \subseteq V$ such that every node in $V - S$ is adjacent to a vertex in S.*

Claim 1.0.2 *Since any set of vertices of a complete graph is a dominating set, the optimal solution to a k-center problem is a dominating set in G.*

Claim 1.0.3 *The optimal solution to a k-center problem is a dominating set in G_i for $i \geq i_0$.*

Let C^* be the cost of an optimal solution. Let e_c be the last edge of cost c^* in ordering $e_1, e_2, ..., e_m$.

Claim 1.0.4 *There is dominating set in G_c of size k or less.*

Claim 1.0.5 *There is no dominating set in G_{c-1} of size k or less, if there was, it would be a set of k-center of cost at most $e_{c-1} < e_c$*

According to the two claims above, 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. However, finding dominating set like k-center is NP-hard. We now lower-bound the size of the dominating set in G_i.

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

Definition 1.0.7 *An independent set in a graph $H = (V, E)$ is a set $S \subseteq V$, such that $\forall i \in S, \forall j \in S$, $(i, j) \notin E$.*

Definition 1.0.8 *Maximum independent set: S is independent set, and $|S|$ is of the largest size.*

Definition 1.0.9 *Maximal independent set: S is independent set, and $\forall V \notin S$, $S \cup V$ is not an independent set.*

Finding a maximum independent set of a graph is NP-hard, but finding a maximal independent set of a graph can be solved in polynomial time using a simple greedy algorithm.

Lemma 1: Given H, let I be an independent set in H^2, then $|I| \leq dom(H)$, where dom(H) denotes size of a minimum cardinality dominating set in H.

Proof Let D be a minimum cardinality dominating set in H. For each $d \in D$, its neighborhood forms a clique in H^2. H^2 contains $|D|$ cliques spanning all the vertices which implies any independent set in H^2 can pick at most 1 vertex per clique. So $|I| \leq |D|$.

2 2-Approximation Algorithm for k-center

Algorithm:

1. Construct G_1^2, G_2^2,... ,G_m^2.
2. Compute a maximal independent set M_i in each graph G_i^2.
3. Find the smallest index i, such that $|M_i| \leq k$, say M_j
4. Return M_j

Lemma 2: For j as defined in the algorithm above, $cost(e_j) \leq c^*$.

Proof For every $i < j$, we have $|M_i| > k$ since $dom(G_i) \geq |M_i|$ by Lemma 1. That implies $dom(G_i) > k$. So the first index for which the k-center problem forms a dominating set $> i$, so $c^* > cost(e_i)$

Theorem 2.0.10 *The algorithm above returns a solution of cost at most* $2 * OPT$*.*

Proof Observe a maximal independent set in H^2 is also a dominating set in H^2. Thus, if we have a maximal independent set equal to the dominating set in G_i^2, call it D, then every vertex is on a path of length at most 2 from D in G_i. Since $i < c^*$ by lemma 2, then each edge $e \in G_i$, $cost(e) < cost(c^*)$, the path of length 2 in G has edges of cost less than $cost(c^*)$. Thus, by triangle inequality, $cost \leq 2c^*$, $\forall$ vertex to their closest vertex in D.

We have shown a 2-approximation to the k-center problem: can we do any better? The answer is no, as we show in the following section.

3 Hardness of Approximation

Theorem 3.0.11 *The problem of approximating the k-center problem with triangle inequality within a factor of* $2 - \epsilon$ *is NP-hard for any* $\epsilon > 0$*.*

Proof By reduction from dominating set. Given a graph $G = (V, E)$, we construct an instance of k-center satisfying triangle inequality, such that if

G has a dominating set of size $\leq k$,then the optimal cost of the k-center is 1, otherwise the optimal cost of the k-center is 2. We put the following weights on the complete graph (note they satisfy the triangle inequality)

$$w(e) = \begin{cases} 1 & \text{if } e \in E; \\ 2 & \text{if } e \notin E. \end{cases}$$

Thus, the the approximation algorithm will output a solution of cost 1 if there is a dominating set in G and output a solution of cost 2 otherwise. So we can use the approximation algorithm to decide whether there is a dominating set in G. Therefore, the approximation algorithm is also NP-hard.

4 Weighted k-center Problem

It is actually the same problem as before, but we are given a weight function on vertices, $w : v \rightarrow R^+$, and a bound $w \in R^+$, the problem is to pick $S \subseteq V$, the total cost is at most W, such that $cost(S) = max_{v \in V} min_{s \in S} cost(e_{vs})$ is minimized. Note this reduces to the ordinary k-center problem when all vertex weights are 1, and $W = k$.

Let $wdom(G)$ denote the weight of the min-weight dominating set in G, then with respect to the graph G (containing the ith lightest edge), we need to find the smallest index i, such that $wdom(G_i) \leq W$. Notice that if i^* is the index, then the cost of the optimal solution is $OPT(cost(e_{i^*}))$

Given a vertex weighted graph H, let I be an independent set in H^2. For each $u \in I$, let $S(u)$ denote the least costly vertex that is a neighbor of u in H (note we allow u be its own neighbor). Let $S = \{s(u) | u \in I\}$.

Lemma 3: $w(S) \leq wdom(H)$

Proof Let D be a min-weight dominating set of H, I claim I can construct a matching between verties in I and vertices in D who are neighbors that saturates I, because

1) every 2 vertices in I cannot have a common neighbor in D

2) each vertex in I has a neighbor in D because D is dominating.

Remains to show: each vertex in S weighs less than its corresponding vertex to I's matched vertex in D. That's because we picked the neighbor or I of min-weight (weight at most its neighbor in D).

3-approximation algorithm of weighted k-center:

1. Construct $G_1^2, G_2^2, ..., G_m^2$.
2. Compute a MIS M_i in each graph G_i^2.
3. Compute $S_i = \{s_i(u) | u \in M_i\}$.
4. Find minimum index i such that $w(S_i) \leq W$, say j.
5. Return S_j.

Theorem 4.0.12 *The algorithm above is* $3 * OPT$ *for weighted k-center.*

Proof Let S_i be the solution by the algorithm. Let OPT be the cost ot the optimal solution. It holds that $w(S_i) > W$ for every $i < j$. This and the Lemma 3 imply that $wdom(G_i) > W$ for every $i < j$. Hence the optimal cost of k-center is at least the cost of the most expensive edge in G_j. Hence, $cost(e) \leq OPT$ for every edge in G_j.

Now, notice that M_j is a dominating set in G_i^2. Hence, every vertex in G is on a path of length at most 1 from some vertex in M_j in G_j^2. Hence, every vertex is on a path of length at most 2 from some vertex in M_j. Since every edge in G_j has weight at most OPT, the triangle inequality guarantee that the edges connecting any vertex in G to some vertex in S_j weigh no more than $3 * OPT$.