Generalized k-Center Problems

*Shiva Chaudhuri**   *Naveen Garg**   *R. Ravi*†

Abstract

The k-center problem with triangle inequality is that of placing k center nodes in a weighted undirected graph in which the edge weights obey the triangle inequality, so that the maximum distance of any node to its nearest center is minimized. In this paper, we consider a generalization of this problem where, given a number p, we wish to place k centers so as to minimize the maximum distance of any node to its p^{th} closest center. We consider three different versions of this reliable k-center problem depending on which of the nodes can serve as centers and non-centers and derive best possible approximation algorithms for all three versions.

Keywords: Approximation algorithm, facility location, k-center problem, bottleneck problem.

*Max-Planck-Institut für Informatik, Im Stadtwald, 66123 Saarbrücken, Germany. Work supported by the EU ESPRIT LTR Project No. 20244 (ALCOM-IT). Email: `shiva,naveen@mpi-sb.mpg.de`

†GSIA, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213. Email: `ravi+@andrew.cmu.edu`. Work done while this author was visiting the Max-Planck-Institut.

1 Introduction

The k-center problem is a classical problem in facility location: given n cities and the distances between them, we wish to select k of these cities as centers so that the maximum distance of a city from its closest center is minimized. The problem is **NP**-hard and Hochbaum and Shmoys present a 2-approximation algorithm[1] for graphs with edge weights obeying triangle inequality [4]. Further they also show that no polynomial time algorithm for this problem can have a performance guarantee of $(2-\epsilon)$ for any $\epsilon > 0$, unless **P**=**NP**. In this paper we consider generalizations of the k-center problem with triangle inequality in which we require that each city has some number (say p) of centers 'close' to it. We extend the techniques of Hochbaum and Shmoys and provide similar best possible approximation algorithms.

Suppose that we wish to locate facilities at k out of n cities such that the maximum distance of a city to its p^{th}-closest facility is minimized. Considering 'p^{th} closest' (as against closest in the k-center problem) is important when the facilities concerned are subject to failure and we wish to ensure that even if up to $p-1$ facilities fail, every city has a functioning facility close to it. However, we can formulate such a generalization of the k-center problem in two different ways depending on whether the center nodes are required to be served by p centers or not. This leads to the first two formulations below. The third problem arises out of a generalization that partitions the nodes into suppliers and customers and requires that only customer nodes be serviced by p center nodes from among the suppliers.

The p-reliable k-center problem

Given a complete graph $G = (V, E)$ with edge weights $w : E \rightarrow \mathbf{R}^+$ that satisfy triangle inequality, the *p-reliable k-center problem* is to find a subset S, of at most k vertices which minimizes

$$\max_{v \in V} d_p(v, S)$$

where $d_p(v, S)$ is the distance of the p^{th}-closest vertex to v in S. Note that if $v \in S$ then the closest vertex to it is v itself.

The p-neighbor k-center problem

Another possible generalization of the k-center problem arises when we demarcate the roles of the centers and non-centers. Consider a setting where we wish to select k sites as warehouses in a distribution network with n sites. We only require that each site that is *not* a warehouse have at least p warehouses close to it while we have no such requirement on the warehouses. Then the problem, which we refer to as the *p-neighbor k-center problem* is to find a subset S of at most k vertices which minimizes

$$\max_{v \in V-S} d_p(v, S)$$

Note that setting $p = 1$ in both the above problems reduce them to the k-center problem.

[1] An α-approximation algorithm for a minimization problem runs in polynomial time and always outputs a solution of value no more than α times the optimal.

The p-neighbor k-supplier problem

In the k-supplier problem the vertices are partitioned as supplier and customer vertices and the problem is to choose k supplier vertices such that the maximum distance of any customer to its nearest chosen supplier is minimum. Hochbaum and Shmoys [5] provide a 3-approximation algorithm for this problem and a result due to Karloff [7] shows that computing a $(3-\epsilon)$-approximation for any $\epsilon > 0$ is **NP**-hard. The generalization of the k-supplier problem that we consider and refer to as the *p-neighbor k-supplier problem* is to choose a set of k suppliers such that the maximum distance of a customer to its p^{th}-nearest chosen supplier is minimized. Formally, if $\mathcal{S}$ denotes the supplier vertices and $\mathcal{C}$ the customer vertices, we wish to pick a set $S \subseteq \mathcal{S}$ of at most k vertices which minimizes

$$\max_{v \in \mathcal{C}} d_p(v, \mathcal{S})$$

1.1 Main results

In this paper we present polynomial-time algorithms achieving an approximation ratio of 2 for the p-reliable k-center problem and the p-neighbor k-center problem and an approximation ratio of 3 for the p-neighbor k-supplier problem. Since these problems are generalizations of the k-center and k-supplier problems, these approximation ratios are the best possible. The techniques used in this paper are mainly graph-theoretic; we relate the size of certain types of dominating sets in a graph to the size of certain types of independent sets in its square.

1.2 Related work

Location problems including several versions of the k-center problem are surveyed in [3]. Kariv and Hakimi [6] describe exact solution methods for the k-center problem.

Turning to approximation algorithms, other than the work of Hochbaum and Shmoys mentioned above, Gonzalez [2] as well as Feder and Greene [1] also describe 2-approximation algorithms for the k-center problem. A generalization with vertex weights is addressed by Hochbaum and Shmoys in [5] which also describes a general paradigm for approximating bottleneck problems. In [9], Wang and Cheng considered a generalization of the k-center problem where the distance to the center is multiplied by a vertex priority in the objective; they developed a 2-approximation algorithm for this problem.

The p-neighbor k-center problem was considered previously by Krumke [8]; he provided a 4-approximation algorithm for this problem. We use ideas from his work for deriving a lower bound for this problem but provide a different algorithm to achieve an approximation ratio of 2.

In the next section, we describe a basic paradigm of Hochbaum and Shmoys for approximating bottleneck problems using the instance of the k-center problem. In the following three sections, we apply the paradigm and generalize the techniques for the three different problems mentioned above.

2 The Basic Paradigm

The problems mentioned in the introduction fall into a general class of problems known in the literature as *bottleneck problems*. Roughly speaking a bottleneck problem is one in which we are trying to optimize a bottleneck, *i.e.* minimizing the maximum or maximizing the minimum value of some quantity. Thus for the k-center problem we wish to find from among all dominating sets of size k, the one in which the longest covering edge is minimum.

Hochbaum and Shmoys [5] developed a general paradigm for approximating **NP**-hard bottleneck problems; we illustrate this paradigm with the k-center problem. Let $w_1, w_2, w_3 \dots$ be the edge weights in increasing order and let G_i be the subgraph induced by edges of weight at most w_i. First observe that the optimum value for the k-center problem is equal to one of the edge weights; in particular it is the minimum edge weight w_i such that G_i has a dominating set of size at most k. While it is easy to generate the subgraphs $G_1, G_2, G_3 \dots$, the problem of checking if these subgraphs have a dominating set of size at most k is **NP**-complete. However, suppose that in the subgraph G_i we can find an independent set I of size more than k such that no vertex in G_i is adjacent to two vertices of I. Then any dominating set in G_i has a unique vertex dominating each vertex of I and therefore cannot be of size k or less.

Given a graph $G = (V, E)$ the x^{th} *power of* G, denoted by $G^x = (V, E^x)$ is a graph with the same vertex set as G and an edge between two vertices if they are connected by a path of at most x edges in G. Then I is an independent set of vertices in G_i^2. Thus to argue that G_i has no dominating set of size at most k, it suffices to find an independent set in G_i^2 of size larger than k. What if the largest independent set we can find in G_i^2 is of size no more than k? While we cannot say anything for sure about the size of a dominating set in G_i, we claim that G_i^2 has a dominating set of size at most k.

To prove this claim we only need to assume that the independent set in G_i^2 that we find (say I) is maximal, *i.e.* the addition of any other vertex to I yields a set which is not independent. But this implies that every vertex not in I has a neighbor in I which means that I is a dominating set in G_i^2.

Let G_j be the first subgraph in the sequence $G_1, G_2, G_3, \dots$ such that the maximal independent set found in G_j^2 is of size no more than k. Since G_{j-1}^2 has an independent set of size larger than k, every dominating set in G_{j-1} is of size more than k and hence the optimum value is at least w_j. Further, G_j^2 has a dominating set (the maximal independent set found) of size at most k. Since the edge weights satisfy triangle inequality, the longest edge in G_j^2 has weight at most $2w_j$. Thus we have a k-center in which the distance of any vertex to its closest center is at most twice the optimum.

Summarizing, we have the following two key ingredients in this 2-approximation for the k-center problem.

1. If G has a dominating set of size at most k then no independent set in G^2 is of size larger than k.
2. A maximal independent set is also a dominating set.

The first observation is useful in establishing a lower bound on the optimum value while the second gives a solution of value at most twice the lower bound.

3 The p-neighbor k-center problem

Note that in this version of the problem, only the noncenter nodes need to be considered in computing the bottleneck objective value. For this problem, we first generalize the notion of independent and dominating sets following Krumke [8] and sketch his proof of a lower bound relating these sets. However, to obtain the upper bound we describe a different algorithm motivated by proving a stronger graph-theoretic lemma about these sets.

Definition 3.1 *A set of vertices $S \subseteq V$ is* p-dominating *if every vertex not in the set has at least p neighbors in it,* i.e. $\forall v \in V - S : \deg_S(v) \geq p$*. Thus, a 1-dominating set is the same as a* dominating set*.*

Definition 3.2 *A set of vertices $S \subseteq V$ is* p-independent *if every vertex in the set has at most $p - 1$ neighbors in it,* i.e. $\forall v \in S : \deg_S(v) \leq p - 1$*. Thus, a 1-independent set is the same as an* independent set*.*

The following lemmas relate the size of a p-dominating set in a graph G to the size of a p-independent set in G and G^2. These can be viewed as extending the relationship between dominating sets and maximal independent sets. The first lemma appears in [8] as Proposition 5. We sketch the proof here for completeness.

Lemma 3.1 *[8] If G has a p-dominating set of size k then no p-independent set in G^2 has size more than k.*

Proof: Let D be a p-dominating set in G ($|D| = k$) and I a p-independent set in G^2 and let v be a vertex in $I - D$. Let S_1 be the vertices in D that are neighbors of v and S_2 the vertices in $V - D$ that are neighbors of the vertices in S_1. Further, let $S = S_1 \cup S_2$. Since each vertex in S is a neighbor of v in G^2, the set I contains at most p vertices from S. The set D on the other hand contains at least p vertices from S (the subset S_1). Since $D - S$ is a p-dominating set and $I - S$ a p-independent set in the graph obtained by deleting the set S we can repeat the above argument in this residual graph $G[V - S]$. Continuing in this manner we will eventually reach a situation when there is no vertex in the residual graph that belongs to the p-independent set but not to the p-dominating set. Since at each step the number of vertices deleted from I was at most the number deleted from D, we have that $|I| \leq |D| = k$. ■

While Krumke showed that a maximal p-independent set in G is p-dominating in G^2, we show below that there is a p-independent set in G that is also a p-dominating set in G (rather than G^2). This reduces the performance ratio of the resulting algorithm from 4 to 2.

Lemma 3.2 *Given a graph $G = (V, E)$ and $1 \leq p \leq n$, there exists a p-independent set $S \subseteq V$ that is also p-dominating.*

Proof: Let S be a p-independent set that is not p-dominating. In particular let $v \in V - S$ be such that $\deg_S(v) = q < p$. Let U be the neighbors of v in S that have exactly $p-1$ neighbors in S and let $G[U]$ be the subgraph induced by U in G. Let I be a maximal independent set (and hence also a dominating set) in $G[U]$. Therefore the set $S - I \cup \{v\}$ is also p-independent.

Define the *potential* of a p-independent set, S, as $\psi(S) = p \cdot |S| - |E(G[S])|$, where $E(G[S])$ denotes the edge set of the subgraph induced by S in G. Since

$$\begin{aligned} |S| - |S - I \cup \{v\}| &= |I| - 1 \\ |E(G[S])| - |E(G[S - I \cup \{v\}])| &= (p-1)|I| - (q - |I|) \end{aligned}$$

we have

$$\psi(S) - \psi(S - I \cup \{v\}) = q - p < 0$$

Given any p-independent set that is not p-dominating we can obtain another p-independent set that has strictly larger potential. Therefore the p-independent set with maximum potential is also p-dominating. ■

The proof of the above lemma also yields a polynomial time procedure for computing a p-independent set that is also p-dominating. We start with some p-independent set and if this is not p-dominating we find a vertex that has less than p neighbors in the set. Then as in the proof we delete and add vertices to the set to obtain another p-independent set with strictly larger potential. Since the potential of a p-independent set is at least zero and at most $p|V|$, we will obtain a p-independent set that is also p-dominating in at most $p|V|$ steps.

Let G_i be the first subgraph in the sequence $G_1, G_2, G_3 \ldots$ for which the p-independent set found in G_i^2 by using the above procedure is of cardinality at most k. By triangle inequality it follows that the longest edge in G_i^2 is of length at most $2w_i$ and hence we have a p-neighbor k-center of value $2w_i$. Since in G_{i-1}^2 we found a p-independent set of cardinality more than k, G_{i-1} does not have a p-dominating set of size k or less by Lemma 3.1. Hence the optimum value is strictly larger than w_{i-1} (*i.e.* at least w_i) and this gives a 2-approximation algorithm for this problem.

4 The p-reliable k-center problem

Recall that this version of the problem assumes that all nodes including the center nodes are considered in evaluating the bottleneck objective. Therefore we modify the definition of p-dominating sets appropriately.

Definition 4.1 *A vertex v* dominates *a vertex u if u is adjacent to v or $u = v$,* i.e. *a vertex* dominates *its neighbors and itself. A set of vertices $S \subseteq V$ is p*-dominating *if every vertex in the graph is dominated by at least p vertices in the set.*

This time we relate the size of a p-dominating set in the graph G to that of an independent set (rather than a p-independent set) in G and G^2.

Lemma 4.1 *If G has a p-dominating set of size k then no independent set in G^2 has size more than k/p.*

Proof: Let I be an independent set in G^2. A vertex in G can dominate at most one vertex of I. Therefore any p-dominating set in G is of size at least $p|I|$. Since G has a p-dominating set of size k, $|I| \leq k/p$. ∎

Lemma 4.2 *If no independent set in G is of size more than k/p then G has a p-dominating set of size k.*

Proof: We first argue that there exists a p-dominating set, S, such that in the subgraph induced over S, any vertex with degree more than $p-1$ has a neighbor with degree exactly $p-1$. We then show that S has an independent set of size $|S|/p$.

Claim 4.1 *There exists a p-dominating set S such that any vertex, $v \in S$, with $\deg_S(v) > p-1$ has a neighbor $u \in S$ with $\deg_S(u) = p-1$.*

Proof: Let S be a p-dominating set and v a vertex in it violating the claim. Then $\deg_S(v) = q > p-1$ and $\deg_S(u) \geq p$ for every $u \in S$ that is a neighbor of v. Let U be the neighbors of v in $V - S$ that have exactly p neighbors in S and let I be a maximal independent set in $G[U]$. Since I is also a dominating set in $G[U]$, the set $S \cup I - \{v\}$ is p-dominating.

Define the *potential* of a p-dominating set, S, as $\psi(S) = (p-1)|S| - |E(G[S])|$. Since

$$\begin{aligned} |S \cup I - \{v\}| - |S| &= |I| - 1 \\ |E(G[S \cup I - \{v\}])| - |E(G[S])| &= (p-1)|I| - q \end{aligned}$$

we have

$$\psi(S \cup I - \{v\}) - \psi(S) = q - (p-1) > 0$$

Thus, given a p-dominating set violating the claim we can obtain another with strictly larger potential. Hence the p-dominating set with maximum potential has the property claimed. ∎

Claim 4.2 *Any p-dominating set S satisfying the previous claim contains an independent set of size at least $|S|/p$.*

Proof: Pick arbitrarily from S a vertex of degree at most $p-1$ and include it in the independent set. Delete the vertex and its neighbors. Continue in this manner till there are no more vertices of degree at most $p-1$ in S. Since every vertex in S of degree more than $p-1$ has a neighbor in S of degree exactly $p-1$, we stop only when we have deleted all vertices in S. Every time we include a vertex in the independent set we delete no more than p vertices from S. Thus the size of the independent set is at least $|S|/p$. ∎

Since G has no independent set of size larger than $\frac{k}{p}$, $|S| \leq k$. This proves the lemma. ∎

The proof of the above lemma also yields a polynomial time algorithm for obtaining a p-dominating set of size at most k. We start with a p-dominating set which perhaps does not satisfy Claim 4.1 and add/delete vertices to it (as in the proof) till it satisfies the property in the claim. Since each modification strictly increases the potential and the potential of a p-dominating set is at least $-|E|$ and at most $(p-1)|V|$, the maximum number of steps needed is at most $(p-1)|V| + |E|$. The

p-dominating set so obtained may not be of size k. But then, using the procedure in the proof of Claim 4.2 we can find an independent set of size at least k/p.

As in the previous section, let G_i be the first subgraph in the sequence $G_1, G_2 \ldots$ for which the independent set found using the above procedure on G_i^2 is of size less than k/p. Then, we have a p-dominating set in G_i^2 of size at most k and hence a solution of value $2w_i$ for the p-reliable k-center problem. Since we found an independent set of size at least k/p in G_{i-1}^2, there is no p-dominating set in G_{i-1} of size at most k. Hence, the optimum value is at least w_i and we have a 2-approximation algorithm for the p-reliable k-center problem.

5 The p-neighbor k-supplier problem

The vertex set of the graph is now partitioned into sets $\mathcal{S}$ and $\mathcal{C}$, the set of suppliers and customers respectively. We wish to find the minimum i such that each customer is dominated by at least p suppliers from a subset of k suppliers in the subgraph G_i.

We use the same definition of a p-dominating set as in the previous section. However, now we are only interested in suppliers dominating customers and hence a p-dominating set now refers to a set of suppliers which dominates every customer at least p times. Furthermore, an independent set is now an *independent set of customers*. The proof of the following lemma is along the lines of that of Lemma 4.1 and is hence omitted.

Lemma 5.1 *If G has a p-dominating set of size k then no independent set in G^2 has size more than k/p.*

Let G_i be the first subgraph such that the maximal independent set found in G_i^2 is of size at most k/p. Since G_{i-1}^2 has an independent set of size more than k/p, G_{i-1} does not have a p-dominating set of size at most k. Hence the optimum value is at least w_i.

The maximal independent set (of customers) in G_i^2 dominates all the other customers by maximality. For each customer in the maximal independent set we pick p suppliers adjacent to it in G_i (if a customer does not have p suppliers adjacent to it in G_i then the optimum value is strictly larger than w_i and we move onto the next subgraph G_{i+1}). Thus the total number of suppliers picked is at most k. Furthermore, this set of suppliers is a p-dominating set in G_i^3. Since the longest edge in G_i^3 is of length at most $3w_i$, we have a solution to the p-neighbor k-supplier problem of value at most thrice the optimum.

References

[1] T. Feder and D. H. Greene, "Optimal algorithms for approximate clusters," *ACM Symp. on Theory of Comp.* 434-444 (1988).

[2] T. F. González, "Clustering to minimize the maximum intercluster distance," *Theoretical Computer Science* **38** 293-306 (1985).

[3] G. Y. Handler and P. B. Mirchandani, *Location on Networks: Theory and Algorithms*, MIT Press, Cambridge MA, (1979).

[4] D. S. Hochbaum and D. B. Shmoys, "A best possible heuristic for the k-center problem," *Math. of Oper. res.*, Vol. 10, No. 2, 180-184 (1985).

[5] D. S. Hochbaum and D. B. Shmoys, "A unified approach to approximation algorithms for bottleneck problems," *Journal of the ACM* **33** 533-550 (1986).

[6] O. Kariv and S. L. Hakimi, "An algorithmic approach to network location problems. Part I: The p-centers problem," *SIAM J. of Appl. Math.* **37** 513-538 (1979).

[7] H. Karloff, cited as private communication in [5].

[8] S. O. Krumke, "On a generalization of the p-center problem," *Inf. Proc. Lett.*, 56, 67-71 (1995).

[9] Q. Wang and K. H. Cheng, "A heuristic algorithm for the k-center problem with vertex weight," *Proc. SIGAL '90*, LNCS 450, 388-396 (1990).