A Probabilistic Analysis of a relaxation for the $K$-median Problem

--- A Graphical Model ---

Sang-Hyung Ahn*

--- Contents ---
I. Introduction
II. Probabilistic analysis
III. Computational experiment
IV. Conclusion

I. Introduction

This paper is sequel to the paper[1], where we performed probabilistic analysis for the Euclidean $k$-median problem in the plane. In this paper, our analysis is concerned with the following graphical $k$-median problem.

Let $G_n(p)$ be a random graph with $n$ nodes, where each edge occurs independently with probability $p$, and $c_{ij}$ is the minimum number of edges on a path joining $i$ to $j$ for where the minimum is taken over all paths joining $i$ to $j$. Thus $c_{ij}$ is the shortest distance between $i$ and $j$, assuming that all edges have length one.

We analyze the ratio $\frac{Z_{IP} - Z_{LP}}{Z_{IP}}$ as $n$ goes to infinity, under some assumptions on the probability distribution of problem instances. We say that an event occurs almost surely if it occurs with a probability that goes to 1 as $n$ goes to $\infty$. Given a random variable $Y$, we write $Y \sim a$ almost surely if, for any constant $\epsilon > 0$, the event $1 - \epsilon \leq \frac{Y}{a} \leq 1 + \epsilon$ occurs almost surely.

In addition to the probabilistic analysis, we also report extensive computational experiments, based on the solution of thousands of medium-size problems in section 3. Some of the results predicted for very large problems by our proba-
bilistic analysis can already be observed on these test problems.

In the remainder of this section we state some useful results from the literature. Our proofs use the following lemmas.

**Lemma 1**

If \( Y_1, \ldots, Y_n \) are independent random variables and \( 0 \leq Y_j \leq 1 \) for \( j = 1, \ldots, n \), then

\[
\Pr(\bar{Y} \geq (1 + \epsilon) \mu) \leq e^{-\epsilon^2 n \mu / 3} \quad \text{for } \epsilon = o(1)
\]

\[
\Pr(\bar{Y} \leq (1 - \epsilon) \mu) \leq e^{-\epsilon^2 n \mu / 2} \quad \text{for } 0 < \epsilon < 1,
\]

\[
\sum_{i=1}^{n} Y_i
\]

where \( \bar{Y} = \frac{\sum Y_i}{n} \) and \( \mu \) is the expected value of \( Y_i \).

**Lemma 2**

An optimal solution \( x = (x_{ij} : i, j = 1, \ldots, n) \) of \( Z_{LP}(y) \) is obtained as follows. For each \( i \), sort the values \( c_{ij} \), \( j = 1, \ldots, n \), so that \( c_{ij_1(i)} \leq c_{ij_2(i)} \leq \cdots \leq c_{ij_n(i)} \) and let \( \hat{p} \) be such that

\[
\sum_{h = j_{\hat{p}}(i)}^{j_{\hat{p}+1}(i)} y_h \leq 1 \leq \sum_{h = j_{\hat{p}}(i)}^{j_{\hat{p}+1}(i)} y_h.
\]

Then

\[
x_{ij} = \begin{cases} 
  y_j & \text{for } j = j_1(i), \ldots, j_{\hat{p}-1}(i) \\
  1 - \sum_{h = j_{\hat{p}}(i)}^{j_{\hat{p}+1}(i)} y_h & \text{for } j = j_\hat{p}(i) \\
  0 & \text{for } j = j_{\hat{p}+1}(i), \ldots, j_n(i)
\end{cases}
\]

**Lemma 3**

\[
Z_{LP} \geq \sum_{i=1}^{n} V_i - k \max_{j=1, \ldots, n} \rho_j(V)
\]

**II. Probabilistic analysis**

We assume (i) \( \hat{p} \geq \frac{\omega \log n}{n} \) where \( \omega = \omega(n) \to \infty \).

(this guarantees that \( G_n(\hat{p}) \) is almost surely connected), and
(ii) \( kp^a \geq \frac{\omega \log n}{n} \).

Let \( e \) be the base of natural logarithms, and \( b = \frac{1}{1 - p} \).

The main results of this section is the following theorem.

**Theorem 4**

(a) Consider \((1+\epsilon)\log n \leq k \leq n\), where \(\epsilon > 0\) is fixed.

Then \( Z_LP = Z_LP \) almost surely.

(b) Consider \(2 \leq k \leq \log n\), \( p \min(1, kp) \leq \frac{\omega \log n}{n} \), where \(\omega \to \infty\).

Then \( \frac{Z_{IP} - Z_{LP}}{Z_{IP}} \leq \frac{1}{1+e} \) almost surely.

In addition, if we let \( kP \to \alpha \), \( 0 \leq \alpha \leq \infty \), and \( p \to \beta \), \( 0 \leq \beta < 1 \), where \( a \) and \( b \) are fixed, then

\[
\frac{Z_{IP} - Z_{LP}}{Z_{IP}} \sim \frac{1 - (1 - \alpha)^a + a^a}{1 + a^a} \text{ almost surely,}
\]

where \( a = e \) if \( b = 0 \) and \((1 - \beta)^{-1/b}\) if \( \beta > 0 \).

The shape of the function \( f(\alpha, \beta) = \frac{1 - (1 - \alpha)^a + a^a}{1 + a^a} \) can be seen in figure 1.

The maximum of this function is \( \frac{1}{e+1} \) attained when \( \alpha = 1 \) and \( \beta = 0 \). When \( \alpha = 0 \) or \( \infty \) the function takes the value 0.

(Fig. 1) Relative Gap as a function of \( kp \) when \( 2 \leq k \leq \log n \).
Proof of Theorem 4 (a):

This part of the theorem is a rephrasing of a known result and is easy to prove. As \(c_{ij} \geq 1\) for \(i \neq j\), we must have

\[
(1) \quad Z_{lp} \geq Z_{lp} \geq n - k.
\]

Theorem 4 (a) follows from (1) if we can show that \(Z_{lp} = n - k\) almost surely. But \(Z_{lp} = n - k\) if and only if there is a subset \(K\) of \(I\), \(|K| = k\), such that, for any \(j \in I - K\), there exists \(i \in K\) such that \(i\) and \(j\) are joined by an edge of \(G_n(p)\), i.e., \(K\) is a dominating set.

Let \(K = \{1, \cdots, k\}\). Then

\[
\Pr (K \text{ is not a dominating set}) 
\leq (n - k) \Pr (k + 1 \text{ is not joined by an edge to } 1\cdots, k) 
= (n - k)(1 - p)^k \leq (n - k) \left(\frac{1}{n}\right)^{k^*} \leq n^{-k^*} \rightarrow 0.
\]

Thus Theorem 4 (a) is proved.

Our proof of Theorem 4 (b) will use the next two lemmas.

**Lemma 5**

Consider \(1 \leq k \leq \log n\).

Assume \(p \min (1, kp) \geq \omega \log n\) where \(\omega \rightarrow \infty\).

Then, \(Z_{lp} = (1 + o(1)) (n - k) (1 - q^t)\) almost surely.

**Proof:**

For subset \(K\) of \(I\), let \(N(K)\) be the neighbor set of \(K\), i.e.

\(N(K) = \{j \in X - K : \text{there exists an edge joining } j \text{ to a node of } K\}\). We have

\[
Z_{lp} \leq \min_{|K| = k} (|N(K)| + 2(n - k - |N(K)|))
= 2(n - k) - \max_{|K| = k} |N(K)|
\]

We prove the lemma by showing that

\[
(2) \quad \max_{|K| = k} |N(K)| = (1 + o(1))(n - k)(1 - q^t) \text{ almost surely, and}
\]

\[
(3) \quad Z_{lp} = (1 + o(1)) \min_{|K| = k} (|N(K)| + 2(n - k - |N(K)|)) \text{ almost surely.}
\]

Consider a fixed subset \(K\) of \(I\), \(|K| = k\). The quantity \(|N(K)|\) is distributed
as $B(n-k, 1-q^t)$. Thus, by Lemma 1, for any small $\epsilon>0$

$$\Pr[|N(K)| \leq (1-\epsilon) (n-k) (1-q^t)] \leq e^{-\epsilon^2 (n-k) (1-k) (1-q^t) / 2} \text{ and }$$

$$\Pr[|N(K)| \geq (1+\epsilon) (n-k) (1-q^t)] \leq e^{-\epsilon^2 (n-k) (1-k) (1-q^t) / 3}$$

Thus we have

(4) $\Pr[\max_{i \leq k} |N(K)| \leq (1-\epsilon) (n-k) (1-q^t)] \leq e^{-\epsilon^2 (n-k) (1-k) (1-q^t) / 2}$

(5) $\Pr[\max_{i \leq k} |N(K)| \geq (1+\epsilon) (n-k) (1-q^t)] \leq \left(\frac{n}{k}\right) e^{-\epsilon^2 (n-k) (1-k) (1-q^t) / 3}$

To obtain (2) we put $\epsilon = 2\left(\frac{k \log \frac{n}{k}}{n-k} (1-q^t)\right)^{1/2}$.

We can use $\left(\frac{n}{k}\right) \leq \frac{ne}{k}$ in (5).

Then the right hand sides in (4) and (5) are both $o(1)$. Thus (2) is proved, provided that the assumption of Lemma 1 holds, i.e., $\epsilon \to 0$.

To prove $\epsilon \to 0$, we consider two cases. Let $0 < \alpha < 1$ be a constant.

When $kp \leq \alpha$, $q^t = (1-p)^t = ((1-p)^{1/p})^k \leq \left(\frac{1}{e}\right)^{k} \leq 1-kp + \frac{(kp)^2}{2}$.

So, $\frac{\epsilon^2}{4} \leq \frac{k \log n}{(n-k) kp (1-\frac{\alpha}{2})} \to 0$ since $\log n \to 0$.

When $kp > \alpha$, $q^t = (1-p)^t = e^{k \log (1-p)} \leq e^{-kp} \leq e^{-\alpha} < 1$.

So $\frac{\epsilon^2}{4} \leq \frac{K \log \frac{n}{k}}{(n-k) (1-e^{-n})} \to 0$ since $\frac{\log x}{x} \to 0$ when $x \to \infty$.

This completes the proof of (2).

To prove (3) it suffices to show that, almost surely, every node in $\bar{I} \cup N(K)$ is joined to at least one node of $N(K)$, i.e. $N(K)$ is a dominating set. We have just shown that $|N(K)| = (1+o(1)) (n-k) (1-q^t)$ almost surely. In addition, we have shown in Theorem 4 (a) that, if a set of nodes has cardinality $\log_2 n$ or more, then it is almost surely a dominating set. So (3) holds if we can show

(6) $R = \frac{\log_2 n}{(n-k) (1-q^t)} \to 0$, $R \leq \frac{\log n}{p (n-k) (1-q^t)}$.

Let us use the constant $\alpha$ introduced earlier.
When \( kp > 0, \, q^t e^{-n} < 1 \).

So \( R \leq \frac{\log n}{p(n-k)p k \left(1 - \frac{a}{2}\right)} \to 0 \) since \( kp^2 \geq \omega \log n \).

This proves (6) and therefore (3) and the lemma.

**Lemma 6**

Consider \( 2 \leq k \leq \log n \).

Assume \( p \geq \frac{\omega \log n}{n} \) and \( kp^2 \geq \frac{\omega}{n} \) where \( \omega \to \infty \).

Then \( Z_{Lp} = \max(n-k, 2n-nkp(1+o(1))) \) almost surely.

**Proof:**

Given a node \( i \), let \( N_1(i) = \{j : c_{ij} = 1\} \) and \( N_2(i) = \{j : c_{ij} = 2\} \). First we give probabilistic estimates of \( |N_1(i)| \) and \( |N_2(i)| \). We will show

\[(7) \min_i |N_1(i)| = (1-o(1))n p \text{ almost surely,} \]

\[(8) \max_i |N_1(i)| = (1+o(1))n p \text{ almost surely and,} \]

\[(9) \min_i |N_2(i)| \geq \min\left(\frac{n}{k}, (1-o(1))np\right) \text{ almost surely.} \]

Note that \( |N_1(i)| \) is distributed as \( B(n-1, P) \). So, by Lemma 1,

\[
\Pr(\min_i |N_1(i)| \leq (1-\epsilon)(n-1)p) \leq ne^{-\frac{r^2}{n-1}p/2} \]

\[
\Pr(\max_i |N_1(i)| \geq (1+\epsilon)(n-1)p) \leq ne^{-\frac{r^2}{n-1}p/3} \]

Putting \( \epsilon = 2\left(\frac{\log n}{(n-1)p}\right)^{1/2} \) yields (7) and (8).

Now consider \( |N_2(i)| \). We will assume \( p \to 0 \) (otherwise \( N_1(i) \) is a dominating set by Theorem 4 (a), and (9) follows). Conditional on \( |N_1(i)| \), the quantity \( |N_2(i)| \) is distributed as \( B(n_2, p_2) \), where \( n_2 = n - |N_1(i)| - 1 \) and \( p_2 = 1 - (1-p)^{|N_1(i)|} \).

By Lemma 1,

\[
\Pr(\min_i |N_2(i)| \leq (1-\epsilon)n_2p_2) \leq ne^{-\frac{r^2}{n_2p_2}} \]

Set

\[
\epsilon = 2\left(\frac{\log n}{n_2p_2}\right)^{1/2}. \text{ We have to show } \epsilon < 1. \]
Note that \( n_2 = (1 - o(1)) n \) and \( p_2 = 1 - (1 - p)^{(1+o(1))n} \geq 1 - e^{-(1+o(1))np} \).

If \( np^2 \delta > 0 \) where \( \delta \) is fixed, then
\[
\frac{\epsilon^2}{4} \leq \frac{\log n}{(1+o(1))n(1-e^{-\delta})} \to 0.
\]

If \( np^2 = o(1) \), then
\[
\frac{\epsilon^2}{4} \leq \frac{\log n}{n^2 p^2} = \frac{1}{\log n} \left( \frac{\log n}{np} \right)^2 \to 0.
\]

So we have just shown, almost surely,
\[
\min_i |N_2(i)| \geq (1 - o(1)) n_2 p_2.
\]

Next

we will use the fact that \( kp^2 \geq \frac{\omega}{n} \) to show \( n_2 p_2 \geq \frac{n}{k} \).

If \( np^2 \delta \), \( 0 < \delta < 1 \) fixed, then
\[
n_2 p_2 \geq (1 + o(1)) n(1-e^{-\delta}) \geq \frac{n}{k} \text{ for } k \geq 2 \text{ and } \delta \text{ close enough to } 1.
\]

If \( np^2 \leq \delta < 1 \), then
\[
1 - e^{-(1+o(1))np^2} \geq np^2 \left( 1 - \frac{np^2}{2} \right).
\]

So
\[
n_2 p_2 \geq (1 + o(1)) n^2 p^2 \left( 1 - \frac{\delta}{2} \right) \geq (1 + o(1)) \frac{\omega}{k} \left( 1 - \frac{\delta}{2} \right) \geq \frac{n}{k}.
\]

This complete the proof of (9).

Now we are ready to get a probabilistic estimate of \( Z_{LP} \). First we obtain an upper bound by considering the solution

(10) \( y_j = \frac{k}{n} \) for \( j = 1, \ldots, n \) and \( x_{ij} \) defined in Lemma 2.

Let \( \delta = \min_i |N_1(i)| \) be the minimum degree of \( G(\epsilon, p) \).

Note that,

if \( \delta \geq \frac{n}{k} - 1 \), then \( Z_{LP} = n - k \) because, using the solution (10),

we have \( c_i = \sum_{j=1}^{n} c_{ij} x_{ij} = 1 - \frac{k}{n} \) for \( i = 1, \ldots, n \).
On the other hand,

\[ c_i \leq \frac{k}{n} \delta + 2 \frac{k}{n} \left( \frac{n}{k} - 1 - \delta \right). \]

\((x_i, j)\) only takes positive values for points \( j \) at distance one or two of \( i \) since, by (9), the number of points at distance 2 is at least \( \min\left( \frac{n}{k}, (1+o(1)) \right) \) which is more than the \( \frac{n}{k} - 1 - \delta \) points needed.

Therefore

\[ Z_{LP} \leq n \sum_{i=1}^{n} c_i \leq 2n - kn, \text{ almost surely.} \]

To obtain a probabilistic lower bound for \( Z_{LP} \) we consider the dual bound given by Lemma 3. We put \( V_i = 2 - \frac{1}{n} \) for \( i = 1, \ldots, n \) and let \( \Delta \) denote the maximum degree of \( G_n(\phi) \). Then

\[ Z_{LP} \geq n \left( 2 - \frac{1}{n} \right) - kn \Delta \left( 1 - \frac{1}{n} \right) = 2n - (1+o(1))nk \phi \text{ almost surely.} \]

This completes the proof of Lemma 6.

Proof of Theorem 4 (b)

It follows from Lemmas 5 and 6 that

\[ \frac{Z_{LP} - Z_{LP}}{Z_{LP}^+} \sim (1+q^+ - \max(1, 2 - kp)) \frac{1+q^+}{(1+q^+)^+} \text{ almost surely} \]

\[ = \frac{q^+ - (1-kp)^+}{1+q^+}. \]

Setting \( a = (1-p)^{-1/p} \) and \( kp = \alpha \), we get

\[ \frac{Z_{LP} - Z_{LP}}{Z_{LP}^+} \sim \frac{1 - (1-\alpha)^+\alpha}{1+\alpha} \text{ almost surely.} \]

It is easy to check that the maximum of this function is achieved when \( p \to 0 \) and \( \alpha = 1 \).

Then its value is \( \frac{1}{1+e} \). \\

An interesting range of parameters which is not considered in Theorem 4 is the case \( 2 \leq k \leq \log n \) and \( p \geq \frac{\omega \log n}{n} \geq kp^2 \) where \( \omega \to \infty \). In this range, the expressions for \( Z_{LP} \) and \( Z_{LP}^+ \) are more complicated than those found in Lemmas
5 and 6. However we conjecture that \( \frac{Z_{IP} - Z_{LP}}{Z_{IP}} \to 0 \) almost surely.

In the range covered by Theorem 4, it is easy to identify conditions under which the ratio \( \frac{Z_{IP} - Z_{LP}}{Z_{IP}} \) is almost surely bounded away from 0.

For example, consider

(11) \( \epsilon \leq kp \leq 1/\epsilon, \ k \geq 2 \) and

(12) \( \omega \sqrt{\log n/n} \leq \rho \leq 1 - \epsilon \), where \( \omega \to \infty \) and \( 0 < \epsilon < 1 \) is fixed.

Then

\[
-k \log b = -kp \left(1 + \frac{p}{2} + \frac{p^2}{3} + \cdots\right) \geq \frac{-kp}{1 - \rho} \geq \frac{-1}{\epsilon^2}
\]

So \( k \leq \log n \) for \( n \) large enough and, by Theorem 4 (b), there is a fixed value \( f(\epsilon) > 0 \) such that

(13) \( \frac{Z_{IP} - Z_{LP}}{Z_{IP}} \geq f(\epsilon) \) almost surely.

In addition, we can show that, under these conditions, a branch and bound algorithm based on the LP bound \( Z_{LP} \) almost surely requires a search tree which is exponential in \( k \). Actually, almost complete enumeration is required.

**Theorem 7**

Assume (11) and (12). A branch and bound procedure that branches by fixing a variable \( y_j \) to 0 or 1 at each node of the search tree which is not pruned, and uses the LP bound to prune the search tree, will almost surely expand at least \( n^{(1 - o(1))k} \) nodes.

**Proof:**

We first note that, under the above assumption, \( -\epsilon \geq -k \log b \geq -1/\epsilon^2 \) and therefore

(14) \( e^{-1/\epsilon^2} \leq q^k \leq e^{-\epsilon} \).

In addition, the assumptions of Lemma 4-4 hold and \( k = o(\sqrt{n}) \) so that

(15) \( Z_{IP} \geq (1 - o(1))n(1 + q^k) \) almost surely.

Let \( Z_{LP}(J_0, J_1) \) be the LP value of the subproblem where \( J_0 = \{ j : y_j \) is fixed to 0} and \( J_1 = \{ j : y_j \) is fixed to 1}. Let \( \alpha < 1 \) and \( \beta > 0 \) be fixed. \( \beta > 0 \) We prove the theorem by showing that,
for any subsets $J_0, J_1$ of $\{1, \ldots, n\}$ such that

$$J_0 \cap J_1 = \emptyset, \quad |J_1| \leq \lceil a_k \rceil \quad \text{and} \quad |J_0| \leq \lceil \beta n \rceil$$

(17) $Z_{LP}(J_0, J_1) < (1 + o(1)) n (1 + q^a)$$

Comparing (15) and (17), we see that for any $\alpha$, by choosing $\beta$ small enough, and using (14) that

(18) $Z_{LP}(J_0, J_1) < Z_{LP}$ for all $J_0, J_1$ satisfying (16).

We shall see that this implies that the algorithm must explore at least

(19) \[ \left( \left\lceil \frac{\beta n}{\alpha k} \right\rceil, \left\lceil \frac{\alpha k}{\alpha k} \right\rceil \right) \geq \left( \frac{\beta n}{\alpha k} \right)^a \quad \text{n Nodes.} \]

Since $\alpha$ was arbitrary we have an almost surely lower bound of $n^{(1-o(1))}$ on the number of nodes explored. On the other hand no branch and bound tree has more than $n^{(1-o(1))}$ nodes.

To verify (19), imagine that setting $y_j=0$ means branching to the left and setting $y_j=1$ means branching to the right. (16)$\sim$(18) imply that any tree contains all possible paths which make $\lceil \alpha k \rceil$ right branches and $\lceil \beta n \rceil$ left branches. The number of such paths is precisely the left-hand side of (19).

We now turn to the proof of (17). As increasing $J_1$ or $J_0$ only serves to increase $Z_{LP}$ we can restrict our attention to $|J_0| = \lceil \beta n \rceil$ and $|J_1| = \lceil \alpha k \rceil$. Using Lemma 1 we can easily prove that the following holds almost surely for $G_n(p)$.

(20) $J$ being a subset of $\{1, \ldots, n\}$ and $|J| = \lceil \alpha k \rceil$ imply

$$|N(J)| \geq (1 - o(1)) n (1 - q^a)$$

see (4).

Furthermore it is easy to see that

(21) $\text{diam}(G_n(p)) = 2$ almost surely.

where $\text{diam}$ refers to the diameter of $G_n(p)$.

Indeed

$$\Pr[\exists i, j \in \{1, \ldots, n\} \text{ such that } i, j \text{ are not joined be a path of length } 2]$$

$$\leq \left( \frac{n}{2} \right) (1 - p^2)^{n-2}$$

$$\leq n^2 e^{-\alpha \ln(n^2) / n} \to 0.$$
Thus (21) is proved. That is, $\Pr\{\text{diam}(G_n(p))=1\} = p^k \to 0$ where $k = \binom{n}{2}$. To obtain an upper bound on $Z_{LP}(J_0, J_1)$ let

$$
y_j = \begin{cases} 
0 & \text{if } j \in J_0 \\
1 & \text{if } j \in J_1 \\
\gamma & \text{if } j \not\in J_0 \cup J_1 
\end{cases}
$$

where $\gamma = \frac{k - \lfloor ak \rfloor}{n - \lfloor \beta n \rfloor - \lfloor ak \rfloor}$.

The values for $x_{ij}$ are then chosen as follows:

- for $i \in J_1$, $x_{ii} = 1$ and $x_{ij} = 0$ for $j \neq i$
- for $i \in N(J_1)$, $x_{ii} = 1$ and $x_{ij} = 0$ for $j \neq i$ where $t$ is a node of $J_1 \cap N_i(t)$.
- for $i \not\in J_1 \cup N(J_1)$, the values are defined in Lemma 2.

With this solution we find, using (21) that

$$
c_\gamma = \begin{cases} 
0 & \text{if } i \in J_1 \\
1 & \text{if } i \in N(J_1) \\
\leq 2 & \text{if } i \not\in J_1 \cup N(J_1) 
\end{cases}
$$

Hence $Z_{LP}(J_0, J_1) \leq |N(J_1)| + 2(n - |N(J_1)|)$ and (17) follows on using (20).

In [3], a different graphical model is associated with the variation of the $k$-median problem known as the $k$-plant location problem. The $k$-plant location problem is defined using two sets $I = \{1, \cdots, n\}$ and $J = \{1, \cdots, m\}$. The quantity $c_{ij}$ is defined for each $1 \leq i \leq m$ and $1 \leq j \leq n$. The problem consists of finding a subset $S$ of $J$, $|S| = k$, that minimizes $\sum_{i=1}^n c_{ij}$.

A $k$-plant location problem arises from a graph $G$ by defining $J$ as its node set, $I$ as its edge set and $c_{ij} = 0$ if $j$ is incident with $i$, 1 otherwise. (The problem is to find $k$ nodes that cover the maximum number of $G$.) It is shown that in [3] that

$$Z_{LP} = Z_{LP}$$

when $G = G_n(p)$ is a random graph with $0 \leq \epsilon \leq p \leq 1 - \epsilon$; $\epsilon$ fixed, and $k \leq n^\alpha$, $\alpha < 1/6$ fixed.
III. Computational experiment

In this section, we report our computational experience with medium-size \( k \)-median problem for a graphical model.

For each problem we computed \( Z_{IP} \) and \( Z_{LP} \). The value of \( Z_{LP} \) was obtained by solving a Lagrangian dual by subgradient optimization as explained in [2]. In the process of computing \( Z_{LP} \), this algorithm generates a feasible solution at each subgradient iteration. Of course, if it happens that the value of the best feasible solution generated equals \( Z_{LP} \), the algorithm terminates since, then, \( Z_{IP} = Z_{LP} \). For most of the test problems with no gap \( Z_{IP} - Z_{LP} \), the algorithm terminated in less than 100 subgradient iterations, due to the above stopping criterion. If, after 100 subgradient iterations, there was still a gap between the best feasible solution (an upper bound on \( Z_{IP} \)) and the best Lagrangian relaxation (a lower bound on \( Z_{LP} \)), we resorted to branch and bound to find \( Z_{IP} \). When the subgradient algorithm clearly converged to a value different from \( Z_{IP} \), we accepted it as showing that \( Z_{IP} \neq Z_{LP} \). In the cases where the subgradient algorithm converged to a value close to \( Z_{IP} \) we used the simplex algorithm to compute \( Z_{LP} \). This allowed us to settle cases where was a very small but positive gap \( Z_{IP} - Z_{LP} \).

Our expriments involves two cases of random graphs. One is graphical model with unit edge cost and the other is that with random edge cost.

1) Unit-edge cost case

First we report the results when the edge lengths are equal to 1. Starting from a random tree on 50 nodes, we generated a sequence of graphs, adding 50 random edges at a time to the previous graph. Table 1 contains the value of \( Z_{IP} \) and \( Z_{LP} \) for each \( 2 \leq k \leq 10 \). Only one figure means that \( Z_{IP} = Z_{LP} \). Note that when \( Z_{IP} = Z_{LP} = n - k \) for same graph, it contains a dominating set and therefore every subsequent graph in the sequence also does.

The problems in region A of Table 1 are relatively easy problems in the
(Table 1) A Graphical Model with unit edge lengths: \( n = 50 \)

<table>
<thead>
<tr>
<th># of edges</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>139</td>
<td>114</td>
<td>98</td>
<td>88</td>
<td>80</td>
<td>72</td>
<td>65</td>
<td>59</td>
<td>54</td>
</tr>
<tr>
<td>99</td>
<td>89</td>
<td>77</td>
<td>68.5/69</td>
<td>62</td>
<td>57</td>
<td>52</td>
<td>48</td>
<td>44.5/46</td>
<td>42</td>
</tr>
<tr>
<td>149</td>
<td>77</td>
<td>69</td>
<td>62</td>
<td>55.5/57</td>
<td>50</td>
<td>46</td>
<td>43</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>199</td>
<td>72</td>
<td>63</td>
<td>44</td>
<td>48</td>
<td>45</td>
<td>43</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>249</td>
<td>72</td>
<td>61</td>
<td>52</td>
<td>46</td>
<td>44/45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>299</td>
<td>69</td>
<td>56</td>
<td>48</td>
<td>46</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>349</td>
<td>65</td>
<td>52.5/54</td>
<td>48</td>
<td>45/46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>399</td>
<td>62</td>
<td>50</td>
<td>47/48</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>449</td>
<td>61</td>
<td>49</td>
<td>46/47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>499</td>
<td>58</td>
<td>47.5/52</td>
<td>46/47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>56</td>
<td>48</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>599</td>
<td>54</td>
<td>47/48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>649</td>
<td>52</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>699</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>749</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>799</td>
<td>48.5/49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>849</td>
<td>48/49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>899</td>
<td>48/49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1199</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* figures are the values of \( Z_{LP} \) and \( Z_{IP} \)
* only one figure means \( Z_{IP} = Z_{LP} \)

sense that \( Z_{LP} = Z_{IP} \) due to the existence of a dominating set in the graphs. Among the instances where a dominating set did not exist, about 28% had a gap.

2) Random edge cost case

Next we turn to the graphical model with non-unit edge lengths. We started from 10 random trees on 50 nodes. We then added random edges, 50 at a time, until the graphs contained 849 edges. The edge lengths were computed using the same scheme as earlier. Namely, the nodes were assigned random integer coordinates in a square of size 10×10 and the length of an edge was the Euclidean distance between its two endpoints, rounded to the closest integer. The distance between two nodes of the graph was taken to be the length of the shortest path joining them in the graph. Table 2 reports the number of instances
Table 2. A Graphical Model with Random edge lengths, N=50

<table>
<thead>
<tr>
<th># of edges</th>
<th># of open facilities</th>
<th>total out of go</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>99</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>149</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>199</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>249</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>299</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>349</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>399</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>449</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>499</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>549</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>599</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>649</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>699</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>749</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>799</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>849</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

out of 170 18 22 28 17 22 17 17 19 21 181 out of 1530

* each figure represent # of problems with duality gaps out of 10 problems.
* # of cases with duality gap: 181/1530=11.9%.

with a gap (out of 10), as a function of the number of edges in the graph and k. For this model, the fraction of instances with a gap was about 12%. The average \( \frac{Z_{1P} - Z_{LP}}{Z_{1P}} \) taken over the instances with a gap was less than 1%.

IV. Conclusion

We provided a probabilistic analysis of the strong linear programming relaxation of the k-median problem. We perform our analysis under various probabilistic assumptions. The model we considered was a graphical model. For this model, we showed

1. Consider \((1+\epsilon)\log n \leq k \leq n\), where \(\epsilon > 0\) is fixed.
   Then \(Z_{1P} - Z_{LP}\) almost surely.

2. Consider \(2 \leq k \leq \log n\), \(p \min(1, k^p) \leq \frac{\omega \log n}{n}\), where \(\omega \to \infty\).
Then \( \frac{Z_{lp} - Z_{lp'}}{Z_{lp}} \leq \frac{1}{1+\varepsilon} \) almost surely.

Then we reported our computational experience with medium-size \( k \)-median problem. For a graphical model with random edge lengths the fraction of instances with a gap was about 12\% and a graphical model with unit edge lengths the fraction was 28\% among the instances where a dominating set did not exist. In either case the fraction is higher than reported one in the literature.

References

