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

——A Graphical Model——

*Sang-Hyung Ahn**

《Contents》

I. Introduction

This paper is sequal 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 ∞. Given a random variable Y, we write $Y \sim a$ almost surely if, for any constant $\varepsilon > 0$, the event $1-\varepsilon \leq \frac{Y}{a} \leq 1+\varepsilon$ 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-

* 서울大學校 經營大學 助教授

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 $Y_1, \cdots, Y_n$ are independent random variables and $0 \leq Y_j \leq 1$ for $j=1, \cdots, n$, then

$$\Pr(\bar{Y} \geq (1+\epsilon)\mu) \leq e^{-\epsilon^2 n\mu/3} \qquad \text{for } \epsilon=o(1)$$
$$\Pr(\bar{Y} \leq (1-\epsilon)\mu) \leq e^{-\epsilon^2 n\mu/2} \qquad \text{for } 0<\epsilon<1,$$

where $\bar{Y}=\frac{\sum_{i=1}^{n} Y_i}{n}$ and μ is the expected value of $\bar{Y}$.

LEMMA 2

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

$$p \text{ be such that } \sum_{h=j_1(i)}^{j_{p-1}(i)} y_h \leq 1 \leq \sum_{h=j_1(i)}^{j_p(i)} y_h.$$

Then

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

LEMMA 3

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

II. Probabilistic analysis

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

(this guarantees that $G_n(p)$ is almost surely connected), and

(ii) $kp^2 \geq \frac{\omega \log n}{n}$.

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 $(1+\epsilon)\log_b n \leq k \leq n$, where $\epsilon > 0$ is fixed.

Then $Z_{IP}=Z_{LP}$ almost surely.

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

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

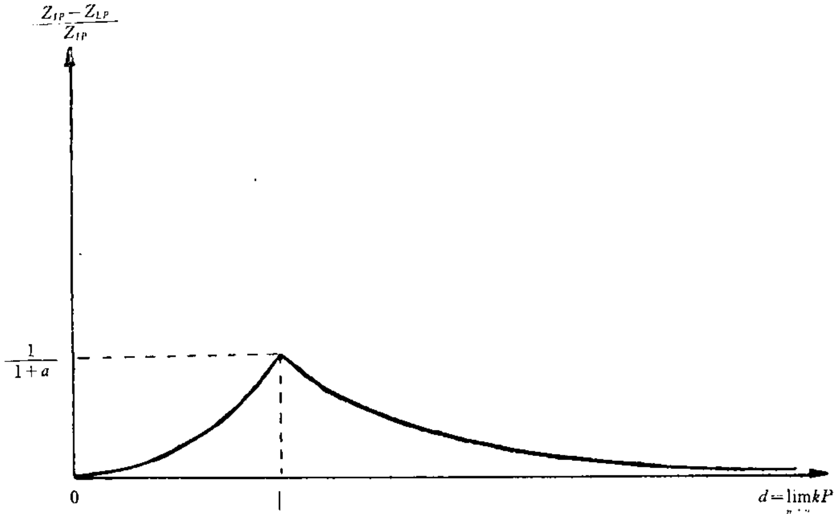
In addition, if we let $kP \to \alpha$, $0 \leq \alpha \leq \infty$, and $p \to \beta$, $0 \leq \beta < 1$, where a and b are fixed, then

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

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

The shape of the function $f(\alpha, \beta)=\frac{1-(1-\alpha)^{+}a^{\alpha}}{1+a^{\alpha}}$ 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 ∞ 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) $Z_{IP} \geq Z_{LP} \geq n-k.$

Theorem 4 (a) follows from (1) if we can show that $Z_{IP}=n-k$ almost surely. But $Z_{IP}=n-k$ 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

$$\begin{aligned}&\Pr\ (K \text{ is not a dominating set})\\ &\leq (n-k)\ \Pr\ (k+1 \text{ is not joined by an edge to } 1 \cdots, k)\\ &= (n-k)(1-p)^k \leq (n-k)\left(\frac{1}{n}\right)^{1+\epsilon} \leq n^{-\epsilon} \to 0.\end{aligned}$$

Thus Theorem 4 (a) is proved.

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

LEMMA 5

Consider $1 \leq k \leq \log_b n$.

Assume $p\min(1, kp) \geq \dfrac{\omega \log n}{n}$ where $\omega \to \infty$.

Then, $Z_{IP}=(1+o(1))(n-k)(1+q^k)$ almost surely.

Proof:

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

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

$$\begin{aligned}Z_{IP} &\leq \min_{|k|=k} \left(|N(K)|+2[n-k-|N(K)|]\right)\\ &= 2(n-k) - \max_{|k|=k} |N(K)|\end{aligned}$$

We prove the lemma by showing that

(2) $\max\limits_{|k|=k} |N(K)| = (1+(o))(n-k)(1-q^k)$ almost surely, and

(3) $Z_{IP}=(1+o(1))\min\limits_{|k|=k}(|N(K)|+2[n-k-|N(K)|])$ almost surely.

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

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

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

Thus we have

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

$$(5)\quad \Pr[\max_{|K|=k}|N(K)|\geq(1+\epsilon)(n-k)(1-q^k)]\leq\binom{n}{k}e^{-\epsilon^2(n-k)(1-k)(1-q^k)/3}$$

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

We can use $\binom{n}{k}\leq\frac{ne}{k}$ 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., $\epsilon\to0$.

To prove $\epsilon\to0$, we consider two cases. Let $0<\alpha<1$ be a constant.

When $kp\leq\alpha$, $q^k=(1-p)^k=((1-p)^{1/p})^{kp}\leq\left(\frac{1}{e}\right)^{kp}\leq1-kp+\frac{(kp)^2}{2}$.

So, $\frac{\epsilon^2}{4}\leq\frac{k\log n}{(n-k)kp\left(1-\frac{a}{2}\right)}\to0$ since $\frac{\log n}{np}\to0$.

When $kp>\alpha$, $q^k=(1-p)^k=e^{k\log(1-p)}\leq e^{-kp}2\leq e^{-\alpha}<1$.

So $\frac{\epsilon^2}{4}\leq\frac{K\log\frac{n}{k}}{(n-k)(1-e^{-\alpha})}\to0$ since $\frac{\log x}{x}\to0$ when $x\to\infty$.

This completes the proof of (2).

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

$$(6)\quad R\equiv\frac{\log_b n}{(n-k)(1-q^k)}\to0,\ R\leq\frac{\log n}{p(n-k)(1-q^k)}.$$

Let us use the constant α introduced earlier.

When $kp > \alpha$, $q^k \leq e^{-\alpha} < 1$.

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

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

LEMMA 6

Consider $2 \leq k \leq \log_b n$.

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) $\min_i |N_1(i)| = (1-o(1))np$ almost surely,

(8) $\max_i |N_1(i)| = (1+o(1))np$ almost surely and,

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

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

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

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

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,

$$\Pr(\min_i |N_2(i)| \leq (1-\epsilon)n_2 p_2) \leq ne^{-\epsilon^2 n_2 p_2/2}$$

Set

$\epsilon = 2\left(\frac{\log n}{n_2 p_2}\right)^{1/2}$. We have to show $\epsilon < 1$.

Note that $n_2=(1-o(1))n$ and $p_2=1-(1-p)^{(1+o(1))np}\geq 1-e^{-(1+o(1))np^2}$.

If $np^2\geq\delta>0$ where δ is fixed, then

$$\frac{\epsilon^2}{4}\leq\frac{\log n}{(1+o(1))n(1-e^{-\delta})}\rightarrow 0.$$

If $np^2=o(1)$, then

$$\frac{\epsilon^2}{4}\sim\frac{\log n}{n^2p^2}=\frac{1}{\log n}\left(\frac{\log n}{np}\right)^2\rightarrow 0.$$

So we have just shown that, almost surely,

$$\min_i \ |N_2(i)|\geq(1-o(1))n_2p_2.$$

Next

we will use the fact that $kp^2\geq\frac{\omega}{n}$ to show $n_2p_2\geq\frac{n}{k}$.

If $np^2\geq\delta$, $0<\delta<1$ fixed, then

$n_2p_2\geq(1+o(1))n(1-e^{-\delta})\geq\frac{n}{k}$ for $k\geq 2$ and δ 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

$$n_2p_2\geq(1+o(1))n^2p^2\left(1-\frac{\delta}{2}\right)\geq(1+o(1))\frac{n\omega}{k}\left(1-\frac{\delta}{2}\right)\geq\frac{n}{k}.$$

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 $j=1,\cdots,n$ and x_{ij} defined in Lemma 2.

Let $\delta=\min_i|N_1(i)|$ be the minimum degree of $G({}_np)$.

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 $i=1,\cdots,n$.

On the other hand,

$$\text{if } \delta<\frac{n}{k}-1, \quad \text{then } c_i\leq\frac{k}{n}\delta+2\frac{k}{n}\left(\frac{n}{k}-1-\delta\right).$$

(x_{ij} 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

$$Z_{LP}\leq n\sum_{i=1}^{n} c_i\leq 2n-k\delta, \ \text{almost surely.}$$

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 $i=1, \cdots, n$ and let $\varDelta$ denote the maximum degree of $G_n(p)$. Then

$$Z_{LP}\geq n\left(2-\frac{1}{n}\right)-k\varDelta\left(1-\frac{1}{n}\right)=2n-(1+o(1))nkp \ \text{almost surely.}$$

This completes the proof of Lemma 6.

Proof of Theorem 4 (b)

It follows from Lemmas 5 and 6 that

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

$$=\frac{q^k-(1-kp)^+}{1+q^k}.$$

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

$$\frac{Z_{IP}-Z_{LP}}{Z_{IP}}\sim\frac{1-(1-\alpha)^+a^\alpha}{1+a^\alpha} \ \text{almost surely.}$$

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 $2\leq k\leq\log_b n$ and $p\geq\frac{\omega\log n}{n}\geq kp^2$ where $\omega\to\infty$. In this range, the expressions for Z_{IP} 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) $\omega\sqrt{\log n/n} \leq p \leq 1-\epsilon$, where $\omega \to \infty$ and $0<\epsilon<1$ is fixed.

Then

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

So $k \leq \log_b n$ 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 sub problem 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,

(16) for any subsets J_0, J_1 of $\{1, \cdots, n\}$ such that

$$J_0 \cap J_1 = \phi, \quad |J_1| \leq \lceil \alpha k \rceil \text{ and } |J_0| \leq \lceil \beta n \rceil$$

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

Comparing (15) and (17), we see that for any α, by choosing β small enough, and using (14) that

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

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

(19) $\binom{\lceil \beta n \rceil + \lceil \alpha k \rceil}{\lceil \alpha k \rceil} \geq \left(\frac{\beta n}{\alpha k}\right)^{\alpha k} = n^{(1-o(1))\alpha k}$ nodes.

Since α 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 $\binom{n}{k} = n^{(1-o(1))}$ 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)～(18) imply that any tree contains all possible paths which make $\lceil ak \rceil$ 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 $\{1, \cdots, n\}$ and $|J| = \lceil \alpha k \rceil$ imply

$$|N(J)| \geq (1-o(1))n(1-q^{\alpha k}) \quad \text{see (4).}$$

Furthermore it is easy to see that

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

where diam refers to the diameter of $G_n(p)$.

Indeed

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

$$\leq \binom{n}{2}(1-p^2)^{n-2}$$

$$\leq n^2 e^{-(n-2)p^2}$$

$$\leq n^2 e^{-\omega \log n (n-2)/n} \rightarrow 0.$$

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

$$y_j=\begin{cases}0 & \text{if } j\in J_0\\ 1 & \text{if } j\in J_1\\ \gamma & \text{if } j\notin J_0\cup J_1\end{cases}$$

$$\text{where } \gamma=\frac{k-\lceil ak\rceil}{n-\lceil \beta n\rceil-\lceil ak\rceil}$$

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 $i\in N(J_1)$: $x_{it}=1$ and $x_{ij}=0$ for $j\neq t$ where t is a node of $J_1\cap N_1(i)$.

for $i\notin J_1\cup N(J_1)$: the values are defined in Lemma 2.

With this solution we find, using (21) that

$$c_i\begin{cases}=0 & \text{if } i\in J_1\\ =1 & \text{if } i\in N(J_1)\\ \leq 2 & \text{if } i\notin J_1\cup N(J_1)\end{cases}$$

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

$$Z_{IP}=Z_{LP} \text{ almost surely}$$

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

III. Computational expriment

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 expreiments 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

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

* 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

# of edges	# of open facilities									total out of go
	2	3	4	5	6	7	8	9	10	
49	0	1	1	0	0	0	0	0	1	3
99	1	1	1	2	3	1	0	1	1	11
149	2	1	2	2	1	0	0	0	0	8
199	1	2	1	0	1	0	2	1	2	10
249	1	2	2	1	1	0	1	3	1	12
299	2	1	2	2	1	2	1	1	1	13
349	2	2	4	1	5	1	0	3	2	20
399	1	3	2	0	2	1	0	1	1	11
449	3	2	2	1	1	2	2	1	0	14
499	0	1	1	2	1	1	1	1	0	8
549	1	1	4	0	0	1	2	1	1	11
599	1	1	0	2	2	2	2	0	2	12
649	2	0	1	2	0	0	3	1	1	10
699	0	2	2	1	0	2	2	0	1	10
749	0	1	1	0	1	2	1	1	1	8
799	1	1	0	1	1	1	0	2	3	10
849	0	0	2	0	2	1	0	2	3	10
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 anf k. For this model, the fraction of instances with a gap was about 12%. The average $\frac{Z_{IP}-Z_{LP}}{Z_{IP}}$ 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 $(1+\epsilon)\log_b n \leq k \leq n$, where $\epsilon > 0$ is fixed.

Then $Z_{IP}=Z_{LP}$ almost surely.

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

Then $\frac{Z_{IP}-Z_{LP}}{Z_{IP}} \leq \frac{1}{1+e}$ 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

〔1〕 Ahn, S., "Probabilistic Analysis of a Relaxation for k-Median Problem—the Euclidean Model in the Plane." 경영논집 XX 제 3 호, 1986년 9월, 46-62.

〔2〕 Cornuejols, G., Fisher, M.L. and Nemhauser, G.L. "Location of bank accounts to optimize float: an analytic study of exact and approximate algorithms, Management Sci." *Management Sci.* 23 (1977), 789-810.

〔3〕 Cornuejols, G., Nemhauser, G.L. and Wolsey, L.A. "Worst-case and probabilistic analysis of algorithms for a location problem." *Operations Res.* 28 (1980), 847-858.