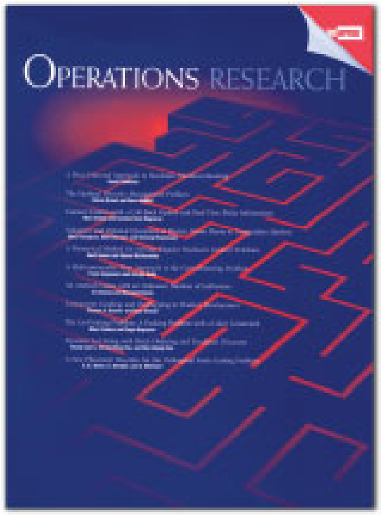

Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Worst-Case and Probabilistic Analysis of Algorithms for a Location Problem

Gerard Cornuejols, George L. Nemhauser, Laurence A. Wolsey,

Worst-Case and Probabilistic Analysis of Algorithms for a Location Problem

GERARD CORNUEJOLS

Carnegie-Mellon University, Pittsburgh, Pennsylvania

GEORGE L. NEMHAUSER

Cornell University, Ithaca, New York

LAURENCE A. WOLSEY

University of Louvain, Louvain-La-Neuve, Belgium

(Received April 1978; accepted April 1979)

We consider a location problem whose mathematical formulation is $\max_S\{z(S): S \subseteq N, |S| = K\}$, where $z(S) = \sum_{i\in I} \max_{j\in S} c_{ij}$ and $C = (c_{ij})$ is any non-negative $m \times n$ matrix with row index set I and column index set N. We show that any procedure which uses matrix C only to calculate values of the function $z(S)$ cannot, with a number of values polynomial in n, guarantee to find an optimal solution. However when C is the edge-vertex incidence matrix of a graph, we show that if n is suitably large and K is fixed or does not grow too rapidly with n, the K vertices of largest degree nearly always constitute an optimal solution

In [2] Cornuejols, Fisher and Nemhauser analyzed the worst-case performance of some approximations for a location problem closely related to the K-median and uncapacitated location problems. Given a non-negative, $m \times n$ matrix $C = (c_{ij})$ with column index set $N = \{1, \cdots, n\}$ and row index set $I = \{1, \cdots, m\}$, for each nonempty $S \subseteq N$ define

$$z(S) = \textstyle\sum_{i\in I}\max_{j\in S}c_{ij} \quad \text{and} \quad z(\phi) = 0. \tag{1}$$

The problem considered in [2] is to find an S of cardinality equal to a specified integer $K (K \le n)$ such that $z(S)$ is maximum; i.e.,

$$\max_{S\subseteq N}\{z(S): |S| = K, z(S) \text{ given by } (1)\}. \tag{2}$$

Unless we specify otherwise, problem (2) can have as its input any positive integers m, n, and $K \le n$, and any non-negative $m \times n$ matrix C.

This paper contains two new results for (2). The first is a negative result on worst-case performance given in Section 1. It is that procedures which use matrix C only to calculate function values cannot, with a number of values of the function $z(S)$ polynomial in n, guarantee to find an optimal solution to (2). The second is a positive result on probabilistic behavior given in Section 2. It is that when C is the edge-vertex incidence

Operations Research
Vol. 28, No 4, July-August 1980

0030-364X/80/2804-0847 $01 25

matrix of a graph, n is suitably large and K is fixed or does not grow too rapidly with n, the K vertices of largest degree nearly always constitute an optimal solution.

Although it is not possible to infer firm practical conclusions from these results, some speculations of practical significance can be drawn. The negative worst-case result is limited, both theoretically and practically, by the fact that the family of matrices C used in its proof have a number of rows that is an exponential function of n. Nevertheless, this result suggests that if an efficient method for (2) exists it should use the matrix C explicitly, rather than just to calculate function values. Many procedures in the literature for solving or obtaining approximate solutions to (2) use only function values (see Section 1). The practical value of the result on probabilistic behavior is limited by its asymptotic nature and by the class of matrices to which it applies. Nevertheless, it suggests that large random problems are easy to solve, which may be a partial explanation for the very good computational experience of many algorithms for problem (2) and for the uncapacitated plant location problem.

1. WORST-CASE PERFORMANCE OF BLACK-BOX ALGORITHMS

A real-valued function z whose domain is all of the subsets of N is said to be *submodular* if $z(S) + z(T) \geq z(S \cup T) + z(S \cap T)$, $\forall\, S, T \subseteq N$ and *nondecreasing* if $z(S) \leq z(T)$, $\forall\, S \subset T \subseteq N$.

The function z given by (1) is submodular and nondecreasing. Nemhauser, Wolsey and Fisher [9], generalized the results of [2] to the problem

$$\max_{S \subseteq N}\{z(S) : |S| = K,\ z \text{ submodular and nondecreasing},\ z(\phi) = 0\}. \quad (3)$$

Problem (3) can have as its input any positive integers n and $K \leq n$ and any submodular and nondecreasing function defined on the subsets of N with $z(\phi) = 0$.

To describe all submodular functions in a consistent manner requires that each function be represented by a list of length 2^n. Nemhauser and Wolsey [7] assumed that submodular functions were described by lists of function values and analyzed the performance of approximation algorithms for problem (3).

For a maximization problem with a positive optimum value we say that an algorithm has a *performance measure* of θ if θ is the greatest lower bound on the fraction of the optimal value that is achieved by the algorithm over all instances (possible inputs) of the problem.

It was shown in [7] that if submodular functions were described by lists of function values then any algorithm for (3) requiring a number of function values polynomial in n cannot have a performance measure exceeding $(e - 1)/e$, where e is the base of the natural logarithm.

The submodular functions given by (1) are naturally described by $m \times n$ matrices. However many procedures in the literature for solving or obtaining approximate solutions to (2) use only function values. Therefore it is meaningful to assume that the class of submodular functions given by (1) is described by lists of function values or, equivalently, to consider the class of algorithms for (2) that treat the function (1) as if it were described by a list of function values. We call members of this class *black-box algorithms*. We imagine that when a black-box algorithm is applied to (2) it does not have access to matrix C, which is hidden in a subroutine. However, for any $S \subseteq N$ the algorithm can obtain $z(S)$ from the subroutine.

We now give three examples of black-box algorithms that have been proposed for (2). The *greedy algorithm* begins with $S^0 = \phi$; given that S^k, $k < K$, it sets $S^{k+1} = S^k \cup \{i^k\}$, where i^k is any element that satisfies $z(S^k \cup \{i^k\}) \geq z(S^k \cup \{j\})$, $\forall\, j \in N - S^k$. It terminates with S^K. The greedy algorithm requires $0(Kn)$ function values. It is shown in [2] that the greedy algorithm for (2) has performance measure equal to $(e - 1)/e$. Some variations of the greedy algorithm that require more function values also have performance measure equal to $(e - 1)/e$.

The pair (i, j), $i \in S$ and $j \in N - S$, is said to be an improving pair with respect to S if $z(S \cup \{j\} - \{i\}) > z(S)$. The *interchange algorithm* begins with an arbitrary set S^0 of cardinality K. Given S^k, it terminates if S^k does not have an improving pair; otherwise it chooses an arbitrary improving pair (i^k, j^k) and sets $S^{k+1} = S^k \cup \{j^k\} - \{i^k\}$. It is shown in [2] that regardless of how an improving pair is chosen, the performance measure of the interchange algorithm equals one-half. Furthermore, a poor choice of improving pairs can yield a number of interchanges that is exponential in n, see [9].

Nemhauser and Wolsey [8] give an integer programming formulation of (3) that, of course, applies to (2). For a given matrix C and integer K let $Z(C, K)$ be the maximum value of (2). Then

$$Z(C, K) = \max \eta$$

$$\begin{aligned} \eta \leq z(S) &+ \textstyle\sum_{j\in N-S} [z(S \cup \{j\}) - z(S)]y_j \\ &- \textstyle\sum_{j\in S} [z(N) - z(N - \{j\})](1 - y_j),\ \forall\, S \subseteq N \\ &\textstyle\sum_{j\in N} y_j = K, \\ &y_j \in \{0, 1\}, j \in N. \end{aligned} \tag{4}$$

$S^* = \{j \in N: y_j^* = 1\}$ is an optimal solution to (2) if and only if $(y_1^*, \cdots, y_n^*)$ is an optimal solution to (4). We might solve (4) by a constraint generation algorithm that begins with an arbitrary subset of the inequalities and generates a violated inequality at each iteration. However, any algorithm applied to the formulation (4) is a black-box algorithm for (2), since the objective function and inequalities depend only on function values.

We now give a family of matrices that will be used to demonstrate that all black-box algorithms for problem (2), using a number of function values polynomial in n, have a performance measure less than one. Unfortunately, these matrices have a number of rows that is an exponential function of n. We have not been able to establish the worst-case performance of black-box algorithms over the family of matrices with m a polynomial function of n. Also we do not know whether a polynomial (in n) algorithm exists for problem (2) with performance measure larger than $(e-1)/e$, even if the black-box assumption is removed.

Let r^T be the characteristic vector of $T \subseteq N$; i.e., $r_j^T = 1$ if $j \in T$ and $r_j^T = 0$ if $j \in N - T$. Let $S^* = \{1, \cdots, k\}$ and C_n^K, $2 \leq K \leq n$, be an $m_n^K \times n$ matrix consisting of zeroes and ones with the following properties: r^T is a row of C_n^K if and only if $n - K \leq |T| \leq n - 1$ and (a) $N - S^* \not\subseteq T$ and $\alpha = |T| - (n - K)$ is even, or (b) $N - S^* \subseteq T$ and α is odd.

Figure 1 gives matrices C_3^2, C_4^2 and C_6^3. Let $z_n^K(S)$ be the function (1) for the matrix C_n^K. From the definition of C_n^K it follows that $z_n^K(S)$ depends only on n, K, $|S|$ and $|S \cap (N - S^*)|$. For each $S \subseteq N$ define $i = |S \cap S^*|$, $j = |S|$ and $v_n^K(i, j) = z_n^K(S)$. Thus $v_n^K(i, j)$ is defined for all $j = 0, \cdots, n$ and all $i = \max(0, j - (n - K)), \cdots, \min(j, K)$. Figure 1 also gives the functions v_3^2, v_4^2 and v_6^3.

LEMMA 1. *For all i, i', j, K and n for which v_n^K is defined*:
(a) *If $j \neq K$ then $v_n^K(i, j) = v_n^K(i', j)$*,
(b) $v_n^K(K, K) = v_n^K(i, K) + 1$, $i < K$.

Proof. (a) Case 1 ($j > K$). $|S| > K$. Thus if T is any subset of N with $|T| \geq n - K$, $S \cap T \neq \phi$, which implies $z_n^K(S) = v_n^K(i, j) = m_n^K$ for all $j > K$ and all i.

Case 2 ($0 < j < K$, $i < j$). We consider an S with $|S| = j$ and $|S \cap S^*| = i$. For α odd there are $\binom{K}{\alpha}$ rows of C_n^K with $n - K + \alpha$ ones since there is one such row r^T for each T of size $n - K + \alpha$ that contains $N - S^*$. Since $j - i > 0$, $S \cap (N - S^*) \neq \phi$ and therefore S intersects every such T.

For α even there are $\binom{n}{n-K+\alpha} - \binom{K}{\alpha}$ rows of C_n^K with $n - K + \alpha$ ones, i.e., the number of different subsets of size $n - K + \alpha$ minus the number that contain $N - S^*$. The number of these rows r^T with $S \cap T = \phi$ is $\binom{n-j}{n-K+\alpha}$ for $\alpha \leq K - j$ and 0 otherwise. (Note that $S \cap T \neq \phi$ for each of the $\binom{K}{\alpha}$ missing rows.) Hence the number with $S \cap T \neq \phi$

is $\binom{n}{n-K+\alpha} - \binom{K}{\alpha} - \binom{n-j}{n-K+\alpha}$ for $\alpha \le K - j$ and $\binom{n}{n-K+\alpha} - \binom{K}{\alpha}$ otherwise. Thus $z(S)$ does not depend on i.

$$C_3^2 = \begin{pmatrix} 1 & & \\ & 1 & \\ 1 & & 1 \\ & 1 & 1 \end{pmatrix} \qquad C_4^2 = \begin{pmatrix} 1 & 1 & & \\ 1 & & 1 & \\ 1 & & & 1 \\ & 1 & 1 & \\ & 1 & & 1 \\ 1 & & 1 & 1 \\ & 1 & 1 & 1 \end{pmatrix}$$

$$C_6^3 = \begin{pmatrix} 1 & 1 & 1 & & & \\ 1 & 1 & & 1 & & \\ 1 & 1 & & & 1 & \\ 1 & 1 & & & & 1 \\ 1 & & 1 & 1 & & \\ 1 & & 1 & & 1 & \\ 1 & & 1 & & & 1 \\ 1 & & & 1 & 1 & \\ 1 & & & 1 & & 1 \\ 1 & & & & 1 & 1 \\ & 1 & 1 & 1 & & \\ & 1 & 1 & & 1 & \\ & 1 & 1 & & & 1 \\ & 1 & & 1 & 1 & \\ & 1 & & 1 & & 1 \\ & 1 & & & 1 & 1 \\ & & 1 & 1 & 1 & \\ & & 1 & 1 & & 1 \\ & & 1 & & 1 & 1 \\ 1 & & & 1 & 1 & 1 \\ & 1 & & 1 & 1 & 1 \\ & & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & \\ 1 & 1 & 1 & 1 & & 1 \\ 1 & 1 & 1 & & 1 & 1 \end{pmatrix}$$

$j \backslash i$	0	1	2
0	0		
1	2	2	
2		3	4
3			4

v_3^2

$j \backslash i$	0	1	2
0	0		
1	4	4	
2	6	6	7
3		7	7
4			7

v_4^2

$j \backslash i$	0	1	2	3
0	0			
1	14	14		
2	21	21	21	
3	24	24	24	25
4		25	25	25
5			25	25
6				25

v_6^3

Figure 1

Case 3. $(0 < j < K,\ i = j)$. We consider an $S \subset S^*$, i.e., $|S| = j = i$. For α odd let r^T be one of the $\binom{K}{\alpha}$ rows of matrix C_n^K with $n - K + \alpha$ ones.

There are $\binom{K}{\alpha} - \binom{K-j}{\alpha}$ of these rows with $S \cap T \neq \phi$ for $\alpha \leq K - j$ and $\binom{K}{\alpha}$ for $\alpha > K - j$. For α even let r^T be one of the $\binom{n}{n-K+\alpha} - \binom{K}{\alpha}$ rows of matrix C_n^K with $n - K + \alpha$ ones. The number of these rows r^T with $S \cap T = \phi$ *is* $\binom{n-j}{n-K+\alpha} - \binom{K-j}{\alpha}$ for $\alpha \leq K - j$ and 0 otherwise. Hence the number with $S \cap T \neq \phi$ is $\binom{n}{n-K+\alpha} - \binom{K}{\alpha} - \binom{n-j}{n-K+\alpha} + \binom{K-j}{\alpha}$ if $\alpha \leq K - j$ and $\binom{n}{n-K+\alpha} - \binom{K}{\alpha}$ otherwise.

Comparing the results of cases 2 and 3, we see that

$$v_n^K(j, j) = v_n^K(i, j) - \sum_{\alpha \text{ odd}} \binom{K-j}{\alpha} + \sum_{\alpha \text{ even}} \binom{K-j}{\alpha} = v_n^K(i, j)$$

since $\sum_{r \text{ odd}} \binom{x}{r} = \sum_{r \text{ even}} \binom{x}{r}$.

(b) We consider an S with $|S| = K$. Every T with $|T| \geq n - K$ satisfies $S \cap T \neq \phi$ except $T = N - S$. But by (a) of the definition of C_n^K, r^{N-S} is a row unless $S = S^*$. Thus $z_n^K(S^*) = v_n^K(K, K) = m_n^K$ and $v_n^K(i, K) = m_n^K - 1$ for $i < K$.

THEOREM 1. *A black-box algorithm that finds an optimal solution to all problems of the form* (2) *requires, in the worst-case, at least* $\binom{n}{K} - 1$ *function values.*

Proof. We consider the family of problems generated by the matrices $\{C_n^K\}$. To solve a problem in this family we must identify the set $S^* = \{1, \cdots, K\}$. Since all sets of a given cardinality not equal to K have the same value, no information about S^* is gained by looking at these sets. But all sets of cardinality K have the same value except S^*, which implies that the identification of S^* requires exhaustive search of the sets of cardinality K. (To be precise, on the family $\{C_n^K\}$ having evaluated $\binom{n}{K} - 2$ sets of cardinality K without finding S^*, we still would not know which of the remaining two sets to choose.)

The poor behavior of black-box algorithms on the matrices of Lemma 1 is clearly limited to black-box algorithms. It is simple, we leave the details to the reader, to solve problem (2) by a polynomial (in n) procedure that uses information from the rows of the matrices $\{C_n^K\}$.

2. COVERING EDGES BY VERTICES IN RANDOM GRAPHS

We consider the case of (2) in which (1) is obtained from the family of edge-vertex incidence matrices of undirected graphs $G = (V, E)$ without multiple edges or loops. Thus the elements of matrix C are zeroes and ones with exactly two ones in each row; no rows are repeated. The problem is to choose a set S of K vertices to maximize the number of edges incident to (covered by) S.

We note that with respect to polynomial computation, black-box algorithms are not a limitation for the edge covering problem. With $(n - 1)n/2 + n$ function values we can determine C, since for vertices j and k

$$z(\{j\}) + z(\{k\}) = \begin{cases} z(\{j, k\}) + 1 & \text{if there is an edge } (j, k) \\ z(\{j, k\}), & \text{otherwise.} \end{cases}$$

For graph G let $V(e)$ be the pair of vertices joined by edge $e \in E$. An integer programming formulation of the edge covering problem is

$$\begin{aligned} Z = \max \textstyle\sum_{e \in E} \sum_{v \in V(e)} x_{ev} & \\ \textstyle\sum_{v \in V(e)} x_{ev} \leq 1, & \quad \forall\, e \in E \\ x_{ev} - y_v \leq 0, & \quad \forall\, e \in E, \quad v \in V(e) \\ \textstyle\sum_{v \in V} y_v = K & \\ x_{ev} \geq 0, & \quad \forall\, e \in E, \quad v \in V(e) \\ y_v \in \{0, 1\}, & \quad \forall v \in V. \end{aligned} \tag{5}$$

where $y_v = 1$ if vertex v is part of the cover and $y_v = 0$ otherwise. Note that if $e = (u, v)$ then $x_{eu} + x_{ev} = 1$ in an optimal solution if and only if $y_u + y_v \geq 1$.

By omitting the integrality conditions in (5) we obtain a linear programming relaxation of the edge covering problem

$$\begin{aligned} Z^{LP} = \max \textstyle\sum_{e \in E} \sum_{v \in V(e)} x_{ev} & \\ \textstyle\sum_{v \in V(e)} x_{ev} \leq 1, & \quad \forall\, e \in E \\ x_{ev} - y_v \leq 0, & \quad \forall\, e \in E, \quad v \in V(e) \\ \textstyle\sum_{v \in V} y_v = K & \\ x_{ev} \geq 0, & \quad \forall\, e \in E, \quad v \in V(e) \\ 0 \leq y_v \leq 1, & \quad \forall\, v \in V. \end{aligned} \tag{6}$$

Let $y_v = 1$ if v is one of the K vertices of largest degree in G and let $y_v = 0$ otherwise. We will show that these values nearly always generate an optimal solution to (5) and (6) when n is large and K is fixed or does not grow too rapidly with n. Similar results for some other combinatorial problems are given in [6].

Let $p_A(n)$ be the fraction of n-vertex graphs that possess a certain property A. We say that *almost all graphs have property A if* $\lim_{n\to\infty} p_A(n) = 1$ [3].

There is also a probabilistic interpretation of $p_A(n)$. Suppose n-vertex graphs are generated randomly by placing an edge between each pair of vertices independently and with probability equal to one-half. With this random model all n-vertex graphs are equally likely and $p_A(n)$ is the probability that a randomly generated n-vertex graph has property A.

Our results are based on the fact that for almost all graphs all of the vertices that have suitably large degrees have sufficiently different degrees. To be precise Cornuejols [1, Theorem 4] has proved

LEMMA 2. *Let G_n be an n-vertex graph with degree sequence $d_1(G_n) \geq d_2(G_n) \geq \cdots \geq d_n(G_n)$. Almost all graphs G_n satisfy $d_k(G_n) - d_{k+1}(G_n) > l$ for all positive integers k and l such that $k \leq n^\alpha$ and $l \leq n^\beta$* where $2\alpha + \beta < 1/2$.

A simple consequence of Lemma 2 is

THEOREM 2. *For almost all graphs G_n and all positive integers $K \leq n^\alpha$, $\alpha < 1/6$, the K vertices of largest degree generate an optimal solution to* (5).

Proof. Assume the vertices $\{1, \cdots, n\}$ are ordered according to the degree sequence. Let $S^* = \{1, \cdots, K\}$ and suppose $S \neq S^*$ is an optimal solution to (5). Consider a pair of vertices $i \in S$, $i > K$ and $j \in S^* - S$ and the solution $S \cup \{j\} - \{i\}$. We have

$$z(S \cup \{j\} - \{i\}) \geq z(S) + d_j(G_n) - (K-1) - d_i(G_n)$$

since the deletion of i uncovers at most $d_i(G_n)$ edges and the addition of j covers at least $d_j(G_n) - (K-1)$ new edges. Take $\alpha = \beta < 1/6$ in Lemma 2. Then for $j \leq K \leq l \leq n^\alpha$ and $i > j$, $d_j(G_n) - d_i(G_n) > l \geq K$ for almost all graphs. So $z(S \cup \{j\} - \{i\}) > z(S) + K - (K-1) > z(S)$, contradicting the fact that S is an optimal solution.

A result more general than Theorem 2, but also more difficult to prove, is

THEOREM 3. *For almost all graphs G_n and all positive integers $K \leq n^\alpha$, $\alpha < 1/6$, the K vertices of largest degree generate an optimal solution to* (6).

Proof. The dual of (6) is

$$
\begin{aligned}
&\min \textstyle\sum_{e\in E} u_e + \sum_{v\in V} s_v + Kt \\
&\qquad u_e + \pi_{ev} \geq 1, \ \forall\ e \in E, \quad v \in V(e) \\
&\qquad -\textstyle\sum_{e\in E(v)} \pi_{ev} + s_v + t \geq 0, \ \forall\ v \in V u_e \geq 0, \\
&\qquad \pi_{ev} \geq 0, \quad s_v \geq 0, \quad \forall\ e \in E, \quad v \in V,
\end{aligned} \tag{7}
$$

where $E(v)$ is the set of edges incident to vertex v.

We will give a solution $[\{\bar{u}_e\}, \{\bar{\pi}_{ev}\}, \{\bar{s}_v\}, \bar{t}]$ and show that it is feasible to (7) for almost all graphs and that $\sum_{e\in E} \bar{u}_e + \sum_{v\in V} \bar{s}_v + K\bar{t}$ equals the number of edges covered by the K vertices of largest degree. The theorem then follows from linear programming duality.

For graph G_n with degree sequence $d_1(G_n) \geq d_2(G_n) \geq \cdots \geq d_n(G_n)$, define $d_i = d_i(G_n)$ and let v_i be the vertex of degree d_i $i = 1, \cdots, n$. Let

$$\bar{u}_e = \begin{cases} 1 & \text{if} \quad e = (v_i, v_j), \quad i \leq K, \quad j \leq K \\ 0 & \text{otherwise.} \end{cases}$$

To satisfy the first set of dual constraints set $\bar{\pi}_{ev} = 1 - \bar{u}_e \geq 0$ for all $e \in E$ and $v \in V(e)$.

Define $\bar{\rho}_{v_i} = \sum_{e\in E(v_i)} \bar{\pi}_{ev_i}$. Note that

$$d_i - K + 1 \leq \bar{\rho}_{v_i} \leq d_i. \tag{8}$$

Now choose $\alpha = \beta < 1/6$ in Lemma 2. Then if $i \leq K \leq n^{\alpha}$, we have $d_i - d_{i+1} > K$ for almost all graphs. Thus, from (8),

$$\bar{\rho}_{v_1} > \bar{\rho}_{v_2} > \cdots > \bar{\rho}_{v_K} > \max_{j>K} \bar{\rho}_{v_j} \tag{9}$$

for almost all graphs.

Define $\bar{t} = \bar{\rho}_{v_K}$ and define

$$\bar{s}_{v_i} = \begin{cases} \bar{\rho}_{v_i} - \bar{t}, & i = 1, \cdots, K \\ 0, & i > K. \end{cases}$$

(9) implies that $\bar{s}_{v_i} \geq 0$, $i = 1, \cdots, K$ and that $-\bar{\rho}_{v_i} + \bar{s}_{v_i} + \bar{t} \geq 0$, $i > K$ for almost all graphs. Thus $[\{\bar{u}_e\}, \{\bar{\pi}_{ev}\}, \{\bar{s}_v\}, \bar{t}]$ is feasible to (7) for almost all graphs if $K \leq n^{\alpha}$, $\alpha < 1/6$.

The value of this dual solution is

$$
\begin{aligned}
&\textstyle\sum_{e\in E} \bar{u}_e + \sum_{i=1}^{K} (\bar{\rho}_{v_i} - \bar{\rho}_{v_K}) + K\bar{\rho}_{v_K} = \sum_{e\in E} \bar{u}_e + \sum_{i=1}^{K} \bar{\rho}_{v_i} \\
&\qquad = \textstyle\sum_{e\in E} \bar{u}_e + (\sum_{i=1}^{K} d_i - 2\sum_{e\in E} \bar{u}_e) = \sum_{i=1}^{K} d_i - \sum_{e\in E} \bar{u}_e,
\end{aligned}
$$

which is precisely the number of edges covered by the K vertices of largest degree.

As a consequence of Theorem 3 and its proof we see that the primal and dual problems (6) and (7) have integral optimal solutions for almost all graphs and $K \le n^{\alpha}$, $\alpha < 1/6$. The integral dual problem is to choose a set of edges $E' \subset E$ to minimize $|E'|$ plus the sum of the K largest degrees in the graph obtained by deleting E' from G. Thus in purely graphical terms our result can be stated as

THEOREM 4. *For almost all graphs and any $K \le n^{\alpha}$, $\alpha < 1/6$, the maximum number of edges that can be covered by K vertices equals the minimum of the cardinality of a set of edges plus the K largest degrees in the subgraph induced by deleting these edges.*

The greedy algorithm applied to (5) chooses a vertex of maximum degree first; at subsequent stages it chooses a vertex of maximum degree in the subgraph induced by deleting the vertices already selected. An obvious consequence of Lemma 2 is that for almost all graphs and any $K \le n^{\alpha}$, $\alpha < 1/6$, the greedy algorithm for (5) chooses the set of K vertices of largest degree. Also, the proof of Theorem 2 shows that the interchange algorithm chooses the K vertices of largest degree for almost all graphs. Thus

COROLLARY 1. *For almost all graphs G_n and positive integers $K \le n^{\alpha}$, $\alpha < 1/6$, the greedy and interchange algorithms yield an optimal solution to* (6).

This excellent asymptotic performance of the greedy and interchange algorithms on almost all graphs is in sharp contrast with worst-case performance. There is a family of graphs $\{G_K\}_{K=2}^{\infty}$ for which the performance of the greedy algorithm on (5) is $(e-1)/e$ and another family [2] for which the performance of the interchange algorithm is one-half.

By Theorem 2, most instances of problem (5) for $K \le n^{\alpha}$, $\alpha < 1/6$, are trivial to solve; nevertheless (5) appears to have very difficult instances even for these values of K since

THEOREM 5. *Given a graph G on n vertices and an integer $K \le n^{1/q}$, the problem of finding a subset of K vertices that cover the maximum number of edges is NP-hard for any positive integer q.*

Proof. Call the problem in the statement of the theorem (Q_1), and consider the problem (Q_2):

Given a graph G on n vertices and an integer $K \le n^{1/q}$ where q is any positive integer, does there exist a subset of K vertices that cover all of the edges of G?

It is obvious that Q_2 reduces to Q_1.

Next consider the problem (Q_3):

Given a graph G on n vertices and an integer $K \le n$ does there exist a subset of K vertices that cover all of the edges of G?

Q_3 is NP-complete [5]. Now as Q_2 is obviously in NP, it suffices to show that Q_3 is polynomially transformable to Q_2.

Let G' be an n'-vertex graph that is the union of G and a graph $\hat{G} = (V, \phi)$ on n^q vertices and no edges. The vertices of G are $\hat{G}$ are not connected. All the edges of G' can be covered with K vertices if and only if all the edges of G can be covered with K vertices. Moreover $K \leq n < (n')^{1/q}$.

Theorem 3 says that for almost all graphs and $K \leq n^{\alpha}$, $\alpha < 1/6$, $Z = Z^{LP}$ (these quantities are the optimum values to (5) and (6), respectively). However, it is easy to see that this result cannot hold for all larger values of K.

THEOREM 6. *For almost all graphs and every* $\delta > 0$ (i) *if* $n/2 \leq K < n - (2 + \delta)\log_2 n$ *then* $Z < Z^{LP}$ *and* (ii) *if* $K > n - (2 - \delta)\log_2 n$ *then* $Z = Z^{LP}$.

Proof. If $K \geq n/2$ then any $\{y_v\}$ that satisfies $\sum_{v \in V} y_v = K$ and $y_v \geq 1/2 \ \forall \ v \in V$ is an optimal solution to (6) with value $Z^{LP} = |E|$; i.e., all edges are covered. Grimmett and McDiarmid [4] have shown that for almost all graphs and every $\delta > 0$ fewer than $n - (2 + \delta)\log_2 n$ vertices will not cover all edges and more than $n - (2 - \delta)\log_2 n$ will.

A slightly different version of Theorem 6 is proved by Pulleyblank [10]. He states that for almost all graphs the optimal solution in real variables to the problem of covering all edges of a graph by a minimum number of vertices is $y_v = 1/2$, $\forall \ v \in V$.

ACKNOWLEDGMENT

This work has been supported by National Science Foundation Grant ENG 75-00568 to Cornell University and a NATO Systems Science Special research grant to the Center for Operations Research and Econometrics of the University of Louvain.

REFERENCES

1. G. CORNUEJOLS, "Degree Sequences of Random Graphs," Discussion Paper No. 7818, Center for Operations Research and Econometrics, University of Louvain, 1978.
2. G. CORNUEJOLS, M. L. FISHER AND G. L. NEMHAUSER, "Location of Bank Accounts to Optimize Float: An Analytic Study of Exact and Approximate Algorithms," *Mgmt. Sci.* **23,** 789–810 (1977).
3. P. ERDOS AND J. SPENCER, *Probabilistic Methods in Combinatorics*, Academic Press, New York, 1974.
4. E. R. GRIMMETT AND C. J. H. MCDIARMID, "On Colouring Random Graphs," *Math. Proc. Comb. Phil. Soc.* **77,** 313–324 (1975).
5. R. M. KARP, "On the Computational Complexity of Combinatorial Problems," *Networks* **5,** 45–68 (1975).
6. R. M. KARP, "The Probabilistic Analysis of Some Combinatorial Search Algorithms," in *Algorithms and Complexity: New Directions and Recent Results*, pp. 1–19, J. F. Traub (ed.), Academic Press, New York, 1976.

7. G. L. NEMHAUSER AND L. A. WOLSEY, "Best Algorithms for Approximating the Maximum of a Submodular Set Function," *Math. Opns. Res.* **3,** 177–188 (1978).
8. G. L. NEMHAUSER AND L. A. WOLSEY, "Maximizing Submodular Set Functions: Formulations and Analysis of Algorithms," Tech. Rep. No. 398, School of Operations Research and Industrial Engineering, Cornell University, 1978.
9. G. L. NEMHAUSER, L. A. WOLSEY AND M. L. FISHER, "An Analysis of Approximations for Maximizing Submodular Set Functions-I," *Math. Programming* **14,** 265–294 (1978).
10. W. R. PULLEYBLANK, "Minimum Node Covers and 2-Bicritical Graphs," *Math. Programming* **17,** 91–103 (1979).