SIAM J. DISC. MATH.
Vol. 1, No. 4, November 1988

MATROIDS AND SUBSET INTERCONNECTION DESIGN*

DING-ZHU DU† AND ZEVI MILLER‡

Abstract. A problem arising in the design of vacuum systems and having applications to some natural problems of interconnection design is described as follows.

(1) Given a set X and subsets X_i, Y_i of X, $i = 1, \cdots, n$, satisfying $X_i \cap Y_i = \varnothing$, find a graph G with vertex set X and the minimum number of edges such that for any i, the subgraph induced by $X \setminus Y_i$ has a connected component containing X_i.

Two other problems related to this one are the following ones.

(2) Given a set X and subsets $X_1, X_2, \cdots, X_n$ such that $X = \bigcup_{i=1}^n X_i$, find a graph G with vertex set X and the minimum number of edges such that for any i the subgraph G_i induced by X_i in G is connected.

(3) Given a set X and subsets $X_1, X_2, \cdots, X_n$ such that $X = \bigcup_{i=1}^n X_i$, find a graph G with vertex set X and the minimum number of edges such that for any subset I of $\{1, \cdots, n\}$, the subgraph induced by $\bigcap_{i \in I} X_i$ is connected.

This paper shows that (3) is polynomial-time solvable while (1) and (2) are NP-complete. Also, some heuristics for (1) and (2) are given. The solution of (3) is an interesting application of matroid theory.

Key word. matroids

AMS(MOS) subject classifications. 05B35, 05C65, 05C35

1. Introduction. In a gas, liquid, or vacuum system, valves are used to make various subsets of the system connected. Here, we consider the following problem: Given required connected subsets, how do we design the system by using the minimum possible number of valves?

This problem appeared first in the design of vacuum systems [1], [2]. More generally, it relates to any situation where specified subsets of a superset must be made internally connected. It can be formalized as follows.

PROBLEM 1. *Given a set X and subsets $X_1, Y_1, \cdots, X_n, Y_n$ of X satisfying $X_1 \cap Y_1 = \cdots = X_n \cap Y_n = \varnothing$, find a graph G with the vertex set X and the minimum number of edges such that*

(P1) *For every $i = 1, \cdots, n$, the subgraph G_i induced by $X \setminus Y_i$ has a connected component containing X_i.*

The elements of our vacuum system correspond to the points of X. A valve corresponds to an edge, and the connected subsets correspond to the subgraphs G_i. Our working requirement is that G_i should interconnect the elements that form X_i and should avoid the elements that form the Y_i.

This problem has been shown to be NP-complete [4] and remains NP-complete for the special case $X_1 \cup Y_1 = \cdots = X_n \cup Y_n = X$. In this special case, we can assume $X = \bigcup_{i=1}^n X_i$ without loss of generality since every graph satisfying (P1) has no edge incident to a point in $X \setminus \bigcup_{i=1}^n X_i$ in this case. Thus, the problem in this case can be restated as follows.

* Received by the editors August 25, 1986; accepted for publication May 13, 1988.

† Mathematical Sciences Research Institute, 1000 Centennial Drive, Berkeley, California 94720. Part of this work was performed while the author was at the Department of Mathematics, University of California at Santa Barbara, Santa Barbara, California 93106 and was supported in part by the National Science Foundation under grants DCR83-12472 and 8210790.

‡ Department of Mathematics and Statistics, Miami University, Oxford, Ohio 45056 and Department of Electrical Engineering and Computer Science, University of California, Berkeley, California 94720. This work was partially supported by the Office of Naval Research under grant N00014-85-K-0621.

PROBLEM 2. *Given a set X and subsets $X_1, X_2, \cdots, X_n$ with $X = \bigcup_{i=1}^n X_i$, find a graph G with vertex set X and the minimum number of edges such that*

(P2) *Every subgraph G_i induced by X_i is connected.*

Since neither Problem 1 nor Problem 2 is polynomial-time solvable, we have to study heuristics for them. A strategy to deal with an intractable problem is to find a related problem the solution of which could be a heuristic for the original one. We therefore introduce a new problem.

PROBLEM 3. *Given a set X and subsets $X_1, X_2, \cdots, X_n$ with $X = \bigcup_{i=1}^n X_i$, find a graph G with vertex set X and the minimum number of edges such that*

(P3) *For any subset I of $\{1, \cdots, n\}$, the subgraph G_I induced by $\bigcap_{i \in I} X_i$ is connected. Here $G_i = G_{\{i\}}$.*

The input size for each of these problems is therefore $O(n|X|)$, since we must store X_i and Y_i for each i, $1 \leqq i \leqq n$.

In this paper we will prove that Problem 3 is polynomial-time solvable. As an application, we obtain a polynomial-time solvable special case for Problem 2. These results will form § 2.

A heuristic is called *bounded* if it is within a constant factor from optimal. Although finding a bounded heuristic for Problem 2 is still open, we will show in § 3 that under the restriction that $|X_i| \leqq 3$ for all i, Problem 2 remains NP-complete but has a bounded heuristic with factor two.

We let (P) denote a property. For simplicity, a graph satisfying (P) is also called a (P)-*graph*. For example, a graph satisfying (P2) is also called a (P2)-graph. We will sometimes specify our input by $(\cdot)$, so, for example, we will speak of a (P2)-graph for $(X_1, X_2, \cdots, X_n)$, etc. A (P)-graph is *minimum* if it has the minimum number of edges over all (P)-graphs for the same input. For a graph G, $\|G\|$ denotes the number of edges of G. For a set Y, $|Y|$ denotes the number of elements of Y.

The operations on graphs are defined in the following way. Let $G = (V, E)$ and $G' = (V', E')$ be two graphs. Then, the union, the intersection, and the difference of G and G' are the graphs $G \cup G' = (V \cup V', E \cup E')$, $G \cap G' = (V \cap V', E \cap E')$, and $G \backslash G' = (V, E \backslash E')$, respectively.

A spanning tree for a set X is a graph with vertex set X, which is a tree. If $X \subset Y$ and T is a spanning tree for Y all of whose endpoints are in X, then T is a *Steiner tree* for X. (P1) then implies that G_i contains a Steiner tree for X_i, and (P2) implies that G_i contains a spanning tree for X_i. A problem on Steiner trees is usually more difficult than the corresponding problem on spanning trees. The allowed extra vertices make the search for minimality more difficult. The same thing happens here. We will make some remarks on that in the last section, and also show that there is a heuristic for Problem 1 within $2\sqrt{2n}$ from optimal.

Clearly, (P3) $\Rightarrow$ (P2) $\Rightarrow$ (P1). The paper is organized in this direction.

Finally, we make two conventions. Throughout this paper, we assume $|X_i| \geqq 2$ for all i since X_i with $|X_i| = 1$ can always be deleted. We also assume from now on that every graph G we consider satisfies $G = \bigcup_{i=1}^n G_i$ since an edge not in $\bigcup_{i=1}^n G_i$ can always be deleted when we consider the above three problems.

2. Analysis of Problem 3. We first show that Problem 3 is polynomial-time solvable.

Define $I(x) = \{i: 1 \leqq i \leqq n, x \in X_i\}$ for any $x \in X$ and $I(H) = \bigcap_{x \in V} I(x)$ for a graph H with the vertex set $V \subseteq X$.

LEMMA 2.1. *A graph G satisfies* (P3) *if and only if for any two vertices x and y satisfying $I(x) \cap I(y) \neq \emptyset$, there is a path l from x to y in G such that $I(l) = I(x) \cap I(y)$.*

Proof. The "only if" part is trivial. For the "if" part, consider any subset I of $\{1, \cdots, n\}$. We will show that G_I is connected. Let x and y be any two vertices of G_I. Then, there is a path l from x to y in G such that $I(l) = I(x) \cap I(y)$. Clearly, $G_I = \bigcap_{i \in I} G_i$ and $I(l) = (I(x) \cap I(y)) \supseteq I$. Thus $G_{I(l)} \subseteq G_I$, so l is in G_I. Therefore, G_I is connected. □

We say that a (P3)-graph is minimal if it does not contain a proper subgraph satisfying (P3). For a graph G, an edge u of G is called a *bridge* if $G \backslash u$ has one more connected component than G does. The following lemma states a necessary and sufficient condition for a (P3)-graph to be minimal.

LEMMA 2.2. *A* (P3)*-graph G is minimal if and only if every edge u of G is a bridge in the subgraph $G_{I(u)}$.*

Proof. Note that a graph with a (P3)-subgraph must satisfy (P3). Thus, a (P3)-graph G is minimal if and only if for every edge u of G, $G \backslash u$ does not satisfy (P3), which, by Lemma 2.1, is equivalent to that u is a bridge in the subgraph $G_{I(u)}$. □

The next lemma shows that a minimal (P3)-graph is in fact minimum.

LEMMA 2.3. *All minimal* (P3)*-graphs have the same number of edges.*

Proof. Let G and G' be two minimal (P3)-graphs. By symmetry, to show $\|G\| = \|G'\|$, it suffices to prove $\|G\| \leqq \|G'\|$. We prove this by induction on $\|G \backslash G'\|$. For $\|G \backslash G'\| = 0$, it is obvious that $\|G\| \leqq \|G'\|$. For $\|G \backslash G'\| > 0$, choose an edge u of $G \backslash G'$. Since G' satisfies (P3), by Lemma 2.1, there is a path l in G' connecting the two endpoints of u such that $I(l) = I(u)$.

We claim that l contains an edge v such that $I(v) = I(l)$. For otherwise, suppose that such a v does not exist. Consider the subgraph $H = \bigcup_{j \in \bar{I}} G_{I \cup \{j\}}$ and $H' = \bigcup_{j \in \bar{I}} G'_{I \cup \{j\}}$ where $I = I(u)$ and $\bar{I} = \{1, \cdots, n\} \backslash I$. Since $G_{I \cup \{j\}}$ and $G'_{I \cup \{j\}}$ are connected graphs with the same vertex set, there is a one-to-one correspondence between the connected components of H and the connected components of H' such that corresponding components have the same vertex set.

Now observe that l is in H'. We have $H' = G'_I \cap [\bigcup_{j \in \bar{I}} G'_j]$. Since $l \subset G'_I$ and by assumption $I(v) \cap \bar{I} \neq \emptyset$ for all edges v in l, it follows that l is in both terms of the intersection describing H', and hence $l \subset H'$.

Thus, the two endpoints of u lie in the same connected component of H', and hence in the same connected component of H. However, H is a subgraph of $G_I \backslash u$. Therefore, u is not a bridge in G_I, contradicting the assumption that G is a minimal (P3)-graph. Hence, such a v exists.

Now, define $G'' = (G' \backslash v) \cup u$. Then $\|G \backslash G''\| = \|G \backslash G'\| - 1$ and $\|G'\| = \|G''\|$. It is easy to verify that G'' is still a minimal (P3)-graph. By the inductive hypothesis, $\|G\| \leqq \|G''\| = \|G'\|$. □

Let $\bar{G}$ denote the complement of a graph G. Define $\mathscr{C} = \{\bar{G}: G \text{ satisfies (P3)}\}$. Clearly, $H_1 \subseteq H_2$ and $H_2 \in \mathscr{C}$ imply $H_1 \in \mathscr{C}$. Furthermore, by Lemma 2.3, we have the following theorem.

THEOREM 2.4. *$\mathscr{C}$ is the family of independent sets of a matroid.* (*See* [5] *for the definition of the matroid.*)

Since a maximum independent set in a matroid can be found by the greedy algorithm, Theorem 2.4 yields the following greedy algorithm for finding minimal (P3)-graphs.

ALGORITHM 1.

$G :=$ the complete graph of X;

for each edge u of G *do*
if $G\backslash u$ satisfies (P3) *then* $G := G\backslash u$.

Thus, we have Corollary 2.5.

COROLLARY 2.5. *Problem* 3 *can be solved in* $O(n|X|^3 \log |X|)$ *time*.

Proof. The validity of the algorithm follows from the matroid property.

Consider the time bound. It suffices to show that the time for each iteration of the loop is bounded by $O(n|X| \log |X|)$ since there are at most $|X|^2$ edges, and hence $|X|^2$ iterations.

We now analyze an iteration in which some edge $u = xy$ is removed from the current G, which is assumed to be (P3).

We claim that $G\backslash u$ is a (P3)-graph $\Leftrightarrow (G\backslash u)_I$ is connected, where $I = I(u)$. The $\Rightarrow$ direction is immediate by definition, so we prove the $\Leftarrow$ direction. By definition of (P3) it suffices to show that $(G\backslash u)_{I'}$ is connected for any index set I'. If $I' \not\subseteq I$, then $u \notin G_{I'}$, so $(G\backslash u)_{I'} = G_{I'}$, and hence $(G\backslash u)_{I'}$ is connected. Suppose $I' \subseteq I$. Since by assumption $(G\backslash u)_I$ is connected, u is not a bridge of G_I, and since $G_I \subset G_{I'}$ it also follows that u is not a bridge of $G_{I'}$. Thus $(G\backslash u)_{I'}$ is connected.

Hence in each iteration we only need to check that $(G\backslash u)_I$ is connected for $I = I(u)$. First we must find the vertex set $V(G_I) = \bigcap_{i \in I} X_i$. We can do this by distinguishing some $i_0 \in I$, computing for each $x \in X_{i_0}$ which j satisfy $x \in X_j$, and then including x in G_I if and only if the j for which this holds are precisely the ones in I. This takes $\leqq \log(|X_j|)$ time for a fixed j, and hence $\leqq \sum_j \log |X_j| \leqq n \log |X|$ time over all j. So the vertex set is found in time $O(n|X| \log |X|)$. Finally, to check that $(G\backslash u)_I$ is connected requires linear time (as a function of $|G_I| \leqq O(|X|)$), so the total time per iteration is $O(n|X| \log |X|)$. □

Unfortunately, so far we do not know any nontrivial upper bound for the ratio of the minimum (P3)-graph to the minimum (P2)-graph. This would be an interesting topic for future research.

Although such a nontrivial upper bound has not been found, we can give some reason for the minimum (P3)-graph to yield a good heuristic for the minimum (P2)-graph. First of all, Du and Chen [2] have given two operations for simplifying a (P2)-graph.

(i) If a (P2)-graph has a cycle Q containing an edge u such that $I(Q) = I(u)$, then u can be deleted.

(ii) Suppose a (P2)-graph has a cycle Q containing two vertices x and y such that each of the two paths from x to y in Q has an edge u satisfying $I(u) = I(l)$. Then the graph obtained by deleting both u's and inserting the edge xy is still a (P2)-graph.

It is easy to see that these operations preserve the property (P3). Thus, a minimum (P3)-graph cannot be simplified by these operations.

In the next two lemmas we will show that in a special case, a minimum (P3)-graph is also a minimum (P2)-graph.

First, Du [3] has shown the following. (For the sake of completeness, we include the proof here.)

LEMMA 2.6. (1) *If a graph G satisfies*

(P4) *For any* $i, j = 1, \cdots, n$, *the subgraph* G_{ij} *induced by* $X_i \cap X_j$ *is a tree*, (*note*: $G_{ii} = G_i$) *then G is a minimum* (P2)-*graph for* $(X_1, \cdots, X_n)$.

(2) *If there is a* (P4)-*graph G for* $(X_1, \cdots, X_n)$ *then every minimum* (P2)-*graph for* $(X_1, \cdots, X_n)$ *satisfies* (P4).

Proof. (1) We prove it by induction on n. It is trivial for $n = 1$. Assume $n > 1$. By the induction hypothesis, $\cup_{i=1}^{n-1} G_i$ is a minimum (P2)-graph for $(X_1, \cdots, X_{n-1})$. Let G' be any (P2)-graph for $(X_1, \cdots, X_n)$. Then

$$\left\| \bigcup_{i=1}^{n-1} G'_i \right\| \geqq \left\| \bigcup_{i=1}^{n-1} G_i \right\|.$$

Let k be the number of connected components of X_n obtained by joining all pairs of vertices in $X_i \cap X_n$ for each $i = 1, \cdots, n-1$. Then $\|G' \backslash \cup_{i=1}^{n-1} G'_i\| \geqq k-1$. However, the property (P4) implies that $\|G \backslash \cup_{i=1}^{n-1} G_i\| = k-1$. Therefore

$$\|G'\| = \left\| G' \Big\backslash \bigcup_{i=1}^{n-1} G'_i \right\| + \left\| \bigcup_{i=1}^{n-1} G'_i \right\| \geqq \left\| G \Big\backslash \bigcup_{i=1}^{n-1} G_i \right\| + \left\| \bigcup_{i=1}^{n-1} G_i \right\| = \|G\|.$$

(2) We will use the fact that for $n \leqq 2$ a (P2)-graph G' is minimum if and only if G' satisfies (P4). Suppose then that a (P2)-graph G' does not satisfy (P4). By rearranging indices of the X_i's, we can assume that either G'_1 or G'_{12} is not a tree. Letting G satisfy (P4) (as in (1)), the case $n \leqq 2$ gives $\|G'_1 \cup G'_2\| > \|G_1 \cup G_2\|$. From the proof of (1), we have $\|\cup_{j=1}^{i} G'_j \backslash \cup_{j=1}^{i-1} G'_j\| \geqq \|\cup_{j=1}^{i} G_j \backslash \cup_{j=1}^{i-1} G_j\|$. Hence

$$\begin{aligned} \|G'\| &= \sum_{i=3}^{n} \left\| \bigcup_{j=1}^{i} G'_j \Big\backslash \bigcup_{j=1}^{i-1} G'_j \right\| + \|G'_1 \cup G'_2\| \\ &> \sum_{i=3}^{n} \left\| \bigcup_{j=1}^{i} G_j \Big\backslash \bigcup_{j=1}^{i-1} G_j \right\| + \|G_1 \cup G_2\| = \|G\|. \end{aligned}$$

But since G satisfies (P4) we know by (1) that G also satisfies (P2). Hence G' is not a minimum (P2)-graph, a contradiction. □

We have the following similar result.

LEMMA 2.7. (1) *A* (P4)-*graph is a minimum* (P3)-*graph.*

(2) *If there is a* (P4)-*graph for* $(X_1, \cdots, X_n)$ *then every minimum* (P3)-*graph for* $(X_1, \cdots, X_n)$ *satisfies* (P4).

Proof. We first prove that (P4) implies (P3). Let G be a (P4)-graph for $(X_1, \cdots, X_n)$. Consider any two vertices x and y of G. For any $i \in I(x) \cap I(y)$, since $G_i = G_{ii}$ is a tree, there is an unique path l_i from x to y in G_i. Similarly, for any $i, j \in I(x) \cap I(y)$, since G_{ij} is a tree, there is a path l_{ij} from x to y in G_{ij}. By the uniqueness of l_i and l_j, we have $l_i = l_{ij} = l_j$. From this fact, we can see that all l_i for $i \in I(x) \cap I(y)$ are identical, say to l. Hence we have $I(l) = I(x) \cap I(y)$. Therefore by Lemma 2.1, G satisfies (P3).

Consider (1). Clearly every (P3)-graph is a (P2)-graph. Hence, if a (P3)-graph is a minimum (P2)-graph then it is also a minimum (P3)-graph. Now, a (P4)-graph satisfies (P3) by the paragraph above and is also a minimum (P2)-graph by Lemma 2.6. Hence it is a minimum (P3)-graph.

To see (2), we only need to notice that the existence of a (P4)-graph implies the existence of a minimum (P3)-graph which is also a minimum (P2)-graph. But a minimum (P2)-graph is a (P4)-graph by Lemma 2.6 and the existence of a (P4)-graph. □

By Lemmas 2.7 and 2.6, we now have the following corollary of Theorem 2.4.

COROLLARY 2.8. *The question of whether a* (P4)-*graph for* $(X_1, \cdots, X_n)$ *exists can be answered in* $O(n|X|^3 \log |X|)$ *time, and if a* (P4)-*graph for* $(X_1, \cdots, X_n)$ *exists then a minimum* (P2)-*graph can be found in* $O(n|X|^3 \log |X|)$ *time.*

An idea for constructing a heuristic can be motivated from the above results. Start by choosing a largest subset I_1 of $\{1, \cdots, n\}$ such that there exists a (P4)-graph for

$(X_i,\ i \in I_1)$. In each subsequent step k, choose a largest subset I_k of $\{1, \cdots, n\} \setminus (\bigcup_{i=1}^{k-1} I_i)$ such that there exists a (P4)-graph for I_k. For each I_k, construct a minimum (P2)-graph for $(X_i,\ i \in I_k)$ and then join them. As will be seen below, this idea results in a heuristic for Problem 2 within a factor of $\sqrt{2n}$ from optimal.

A simple sufficient condition for an input $(X_1, \cdots, X_n)$ to have a (P4)-graph is that all X_i, $i = 1, \cdots, n$, have a point x in common. Then the tree consisting of an internal vertex, which is x, and $|X| - 1$ leaves, which form the set $X \setminus \{x\}$, where $X = \bigcup_{i=1}^{n} X_i$, is a (P4)-graph, and hence a minimum (P2)-graph. Now, for an input $(X_1, \cdots, X_n)$, we choose $I_1, \cdots, I_k$ in the following way.

ALGORITHM 2.

Choose x_1 such that $|I(x_1)| = \max_{x \in X} |I(x)|$; $I_1 := I(x_1)$; $I := I(x_1)$; $k := 2$;
while $\{1, \cdots, n\} \setminus I \neq 0$ *do begin*
choose x_k such that

$$|I(x_k) \setminus I| = \max_{x \in X} |I(x) \setminus I|;$$

$I_k := I(x_k) \setminus \bigcup_{j=1}^{k-1} I_j$;
$I := I \cup I_k$;
$k := k + 1$;
end.

From the algorithm, we can see that the I_i, $i = 1, \cdots, k$, form a partition of $\{1, \cdots, n\}$, and for each $i = 1, \cdots, k$ and all $j \in I_i$ the sets X_j have the point x_i in common. In addition, $|I_i| = \max_{x \in X} |I(x) \setminus \bigcup_{j=1}^{i-1} I_j|$. Let H be the union over all $i = 1, \cdots, k$ of minimum (P2)-graphs for $(X_j, j \in I_i)$. Clearly,

$$\|H\| \leqq \sum_{i=1}^{k} \left(\left| \bigcup_{j \in I_i} X_j \right| - 1 \right).$$

The next lemma states that $\|H\|$ is within a factor of $\sqrt{2n}$ from an optimal solution to Problem 2.

LEMMA 2.9. *Let G^* be a minimum* (P2)-*graph for* $(X_1, \cdots, X_n)$. *Then*

$$\sum_{i=1}^{k} \left(\left| \bigcup_{j \in I_i} X_j \right| - 1 \right) < \sqrt{2n}\, \|G^*\|.$$

Proof. Let $G^{(i)} = \bigcup_{j \in I_i} G_j^*$ and $V^{(i)} = \bigcup_{j \in I_i} X_j$. Since X_j, $j \in I_i$, have a point in common, $G^{(i)}$ is a connected graph with the vertex set $V^{(i)}$. Therefore,

$$|V^{(i)}| - 1 \leqq \|G^{(i)}\|.$$

Let t be the maximum number of $V^{(i)}$'s that can have a point in common. We first prove $t < \sqrt{2n}$. Let x be a common point of $V^{(i_1)}, V^{(i_2)}, \cdots, V^{(i_t)}$ with $i_1 < i_2 < \cdots < i_t$. Since all I_i, $i = 1, \cdots, t$, are disjoint, $I(x)$ contains at least t indices not in $\bigcup_{j=1}^{i_1-1} I_j$. Hence, $|I_{i_1}| = \max_{x \in X} |I(x) \setminus \bigcup_{j=1}^{i_1-1} I_j| \geqq t$. Similarly, we can prove that $|I_{i_2}| \geqq t - 1, \cdots, |I_{i_t}| \geqq 1$. Therefore, $n \geqq |I_{i_1}| + |I_{i_2}| + \cdots + |I_{i_t}| \geqq t + (t-1) + \cdots + 1 = \frac{1}{2}t(t+1) > \frac{1}{2}t^2$. Hence, $t < \sqrt{2n}$.

Now, since every point x can be contained by at most t of the $V^{(i)}$'s, every edge of G^* can belong to at most t of the $G^{(i)}$'s. Thus, we have

$$\sum_{i=1}^{k} (|V^{(i)}| - 1) \leqq \sum_{i=1}^{k} \|G^{(i)}\| \leqq t\|G^*\| < \sqrt{2n}\, \|G^*\|. \qquad \square$$

THEOREM 2.10. *There is a heuristic for Problem 2 that runs in $O(n|X|^2)$ time and is within a factor $\sqrt{2n}$ from optimal.*

It is worth mentioning that so far we do not know how to find a largest subset I of $\{1, \cdots, n\}$ such that a (P4)-graph for $(X_i, i \in I)$ exists. A better heuristic for Problem 2 may be found by solving this problem.

3. On a bounded heuristic for Problem 2. We have presented a heuristic for Problem 2 within a factor $\sqrt{2n}$ from optimal. Is there a heuristic within a constant factor from optimal? We know of such a heuristic only for the case that for every $i = 1, \cdots, n$, $|X_i| \leqq 3$. There is an interesting relation between this case and the vertex covering problem (denoted VC). A vertex cover of a graph G is a subset S of $V(G)$ such that every edge in G is incident on at least one point of S. The problem VC is as follows. Given a graph G and positive integer k, does G have a vertex cover of size at most k?

THEOREM 3.1. *Problem 2 is* NP-*complete, even when all subsets satisfy $|X_i| \leqq 3$.*

Proof. We reduce the vertex covering problem (VC) to Problem 2. Let G and $k \in \mathbb{Z}^+$ be an instance of VC. Let x^* be a new point not in $V(G)$, and for any edge $u = xy \in E(G)$ let $X_u = \{x, y\}$ and $X_u^* = \{x, y, x^*\}$. The corresponding instance of Problem 2 is the set system $(X_u, X_u^*; u \in E(G))$ and integer $\|G\| + k$ (where the question is whether the system has a (P2)-graph with at most $\|G\| + k$ edges).

It is easy to verify that a subset Y of V is a vertex-covering of G if and only if the graph $G \cup (\bigcup_{x \in Y} (x^*, x))$ is a (P2)-graph for $(X_u, X_u^*, u \in E)$. □

THEOREM 3.2. *There is a heuristic within a factor of* 2 *from optimal for Problem* 2 *with the restriction that all X_i's are of size at most* 3.

Proof. We provide a polynomial time reduction preserving the approximation performance from this special case of Problem 2 to VC. Since there is a polynomial approximation for VC to within a factor of 2 [7], the theorem follows.

For our instance of Problem 2 let $X_1, X_2, \cdots, X_n$ be sets of points of size at most 3. Without loss of generality, assume $|X_1| = \cdots = |X_m| = 3$ and $|X_{m+1}| = \cdots = |X_n| = 2$. The graph G for the corresponding instance of VC is constructed as follows.

(1) The vertices of G are all subsets of size 2 of $X = \bigcup_{i=1}^n X_i$.

(2) Two vertices A, B of G are joined by an edge if and only if neither A nor B is in $\{X_{m+1}, \cdots, X_n\}$ and there is an X_i, $1 \leqq i \leqq m$, such that $A \subseteq X_i$ and $B \subseteq X_i$.

Let S be a subset of vertices of G disjoint from $\{X_{m+1}, \cdots, X_n\}$. Define H to be the graph with the vertex set X and the edge set $\{X_{m+1}, \cdots, X_n\} \cup \{(x, y): \{x, y\} \in S\}$. We will prove that S is a vertex-covering of G if and only if H is a (P2)-graph for $(X_1, \cdots, X_n)$.

First, suppose that S is a vertex-covering. To see that H satisfies (P2), it suffices to show that every X_i of size 3 contains at least two subsets belonging to

$$S \cup \{X_{m+1}, \cdots, X_n\}.$$

Let A, B, and C be the three subsets of size two of X_i. If two of A, B, and C belong to $\{X_{m+1}, \cdots, X_n\}$ then we are done. If one of A, B, and C, say A, belongs to $\{X_{m+1}, \cdots, X_n\}$, then the edge BC exists in G, so either B or C is in S since S is a vertex-covering. Hence X_i contains the two required subsets A, and B or C. If none of A, B, and C belongs to $\{X_{m+1}, \cdots, X_n\}$, then the edges AB, BC, and CA all exist in G. To cover them, S must contain at least two of A, B, and C.

Next, suppose that H is a (P2)-graph for $(X_1, \cdots, X_n)$. Consider each edge AB of G. There exists X_i, $1 \leqq i \leqq m$, such that $A \subseteq X_i$ and $B \subseteq X_i$. Since H_i is connected and X_i is of size three, H_i has at least two edges of which at least one corresponds to A or B. Hence, either A or B belongs to $S \cup \{X_{m+1}, \cdots, X_n\}$. However, neither A nor B is in

$\{X_{m+1}, \cdots, X_n\}$. Therefore, either A or B is in S. It follows that S is a vertex-covering of G.

Note that $\|H\| = n - m + |S|$. Hence, S is a minimum vertex-covering of G if and only if H is a minimum (P2)-graph for $(X_1, \cdots, X_n)$ and if S is a heuristic within two from optimal for VC then so is H for the minimum (P2)-graph. This completes our reduction and the theorem is proved. □

4. Remarks on Problem 1. As we have indicated in the Introduction, Problem 1 seems much more difficult than Problem 2. A result similar to Lemma 2.6 has been given by Chao and Du [1]. (See Theorem 4.1 below.) However, the proof is different and more difficult. Surprisingly, it seems the simpler method in the proof of Lemma 2.6 does not work here.

THEOREM 4.1 (Chao and Du [1]). *If G is a* (P4)*-graph for* $(X \setminus Y_1, \cdots, X \setminus Y_n)$ *and for every edge u of G, there exists an index i such that two endpoints of u are in X_i, then G is a minimum* (P1)*-graph for* $(X, X_1, Y_1, \cdots, X_n, Y_n)$.

Another example that can explain the difference between these two problems is the special case where all $X \setminus Y_i$ have a point in common. For Problem 2, this is equivalent to saying that all X_i have a point in common, so an optimal solution can be found easily as shown in § 2. However, for Problem 1, it is still an unsolved case. Let x^* be the common point of all $X \setminus Y_i$, and let G be the graph formed by unioning all edges xx^* for $x \in \bigcup_{i=1}^n X_i$. The trouble is that G may not be a minimum (P1)-graph for $(X, X_1, Y_1, \cdots, X_n, Y_n)$. In fact, let G^* be a minimum (P1)-graph, then $\|G\| - \|G^*\|$ can be made to be any positive integer by choosing the input properly. However, G is a bounded heuristic for G^*.

LEMMA 4.2. $\|G\| \leq 2\|G^*\|$.

Proof. Let G^* be a minimum (P1)-graph such that the degree of x^* in G^* is a maximum over all minimum (P1)-graphs. Let H and K be connected components of G^* such that x^* is in H and is not in K. It is easy to see that H is a tree with only one internal vertex that is x^*. Since every X_i is of size at least two, K contains at least one edge. Deleting all edges in K and connecting x^* to very vertex in K increases the total number of edges by one per edge in K. Therefore, $\|G\| \leq 2\|G^*\|$. □

We have Theorem 4.3, which is similar to Theorem 2.10.

THEOREM 4.3. *There is a heuristic for Problem* 1 *that runs in*

$$O(\min\{n|X|^2, n^2|X|\})$$

time and is within the factor $2\sqrt{2n}$ *from optimal.*

Proof. Define $I(x) = \{i: 1 \leq i \leq n, x \in X \setminus Y_i\}$ for any $x \in X$ and $I(H) = \bigcap_{x \in V} I(x)$ for a graph H with the vertex set $V \subseteq X$. Choose the index sets $I_1, \cdots, I_k$ by Algorithm 2. Let G^* be a minimum (P1)-graph. Let $G^{(i)} = \bigcup_{j \in I_i} G_j^*$ where G_j^* is the subgraph of G^* induced by $X \setminus Y_j$. By the same argument as in the proof of Lemma 2.9, we can obtain $\sum_{i=1}^k \|G^{(i)}\| < \sqrt{2n}\|G^*\|$. By Lemma 4.2, for each I_i, we can construct a (P1)-graph $H^{(i)}$ for $(X, X_j, Y_j, j \in I_i)$ such that $\|H^{(i)}\| \leq 2\|G^{(i)}\|$. Therefore, the heuristic solution $H = \bigcup_{i=1}^k H^{(i)}$ satisfies

$$\|H\| \leq \sum_{i=1}^{k} \|H^{(i)}\| \leq 2 \sum_{i=1}^{k} \|G^{(i)}\| < 2\sqrt{2n}\|G^*\| \qquad \square$$

Acknowledgments. Some ideas included in this paper were obtained in a class offered by Professor Ronald V. Book at the University of California, Santa Barbara, CA during 1984–1985. Thus the authors thank Professor Book and several classmates for helpful discussions.

REFERENCES

[1] S.-C. CHAO AND D.-Z. DU, *A sufficient condition for optimal solutions of the valve-placement problem*, J. North-East Heavy Industry Institute, 4 (1983). (In Chinese.)

[2] D.-Z. DU AND Y.-M. CHEN, *Placement of valves in vacuum systems*, J. Light Sources, 4 (1976). (In Chinese.)

[3] D.-Z. DU, *An optimization problem on graphs*, Discrete Appl. Math., 14 (1986), pp. 101–104.

[4] ———, *Generalized complexity cores and levelability of intractable sets*, Ph.D. thesis, University of California, Santa Barbara, CA, 1985.

[5] M. GONDRAN AND M. MINOUX, *Graphs and Algorithms*, John Wiley, New York, 1984.

[6] M. R. GAREY AND D. S. JOHNSON, *Computers and Intractability, A Guide to the Theory of* NP-*Completeness*, Freeman, San Francisco, CA, 1979.

[7] C. H. PAPADIMITRIOU AND K. STEIGLITZ, *Combinatorial Optimization, Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, NJ, 1982.