Effect of Cardinality and Frequency Restrictions on the Hitting Set and Vertex Cover Problems

Dionysios Kountanis, Jiuqiang Liu and Kenneth Williams
Western Michigan University
Kalamazoo, Michigan 49008

Abstract

The Hitting Set problem is investigated in relation to restrictions imposed on the cardinality of subsets and the frequency of element occurences in the subsets. It is shown that the Hitting Set subproblem where each subset has cardinality C for fixed $C \geq 2$ and the frequency of each element is exactly f for fixed $f \geq 3$ remains NP-complete, but the problem becomes polynomial when $f \leq 2$. The restriction of the Vertex Cover problem to f-regular graphs for $f \geq 3$ remains NP-complete.

1 Introduction

The Hitting Set problem is defined as follows:
Hitting Set (HS)
Instance: Collection D of subsets of a finite set S, positive integer K.
Question: Does S contain a hitting set for D of size K or less, that is, a subset $S' \subseteq S$ with $|S'| \leq K$ such that S' contains at least one element from each subset in D?

It has been proved in [3] that HS is NP-complete even if $|d| \leq 2$ for all d in D. Now, for each $s_i \in S$, we define the frequency of s_i to be $f_i = f(s_i) =$ the count of those subsets in D which contain s_i. This leads to the following definition of (C, f)-Hitting Set which is a subproblem of HS.

(C,f)-Hitting Set
Instance: Collection D of subsets of a set S such that each subset in D has cardinality C and each element of S has frequency f (i.e. each element of S is contained in exactly f subsets in D).
Question: Does S contain a hitting set for D of size K or less?

We begin to develop results on the (C, f)-HS problem by considering Vertex Cover which is defined as follows:
Vertex Cover (VC)
Instance: A graph $G = (V, E)$ and a positive integer $K \leq |V|$.
Question: Is there a vertex cover of size K or less for G, that is, a subset $V' \subseteq V$ such that $|V'| \leq K$ and, for each edge $(u, v) \in E$, at least one of u and v belongs to V'?

It is clear that VC is a subproblem of HS where the cardinality of each subset in HS is exactly 2. Let Max(f)-VC be the subproblem of VC which contains only those instances with the graph having maximum degree less than or equal to f. Let f-VC be the subproblem of VC which contains only those instances with the graphs being f-regular. The formal definitions follow:
Max(f)-Vertex Cover
Instance: A graph $G = (V, E)$ with maximum degree less than or equal to f and a positive integer $K \leq |V|$.
Question: Is there a vertex cover of size K or less for G?
f-Vertex Cover (f-VC)
Instance: An f-regular graph $G = (V, E)$ and a positive integer $K \leq |V|$.
Question: Is there a vertex cover of size K or less for G?

In section 4 we also reference the Edge Cover problem.
Edge Cover (EC)
Instance: A graph $G = (V, E)$ and a positive integer $K \leq |V|$.
Question: Is there an edge cover of size K or less for G, that is, a subset $E' \subseteq E$ with $|E'| \leq K$ such that each vertex $v \in V$ belongs to at least one $e \in E'$?

2 Complexity Results on Vertex Cover

It is known, see [1] and [7], that the CLIQUE problem for graphs (determining whether graph G contains a complete subgraph with at least K vertices) is polynomially equivalent to VC. Also, it is shown in [6] that CLIQUE remains

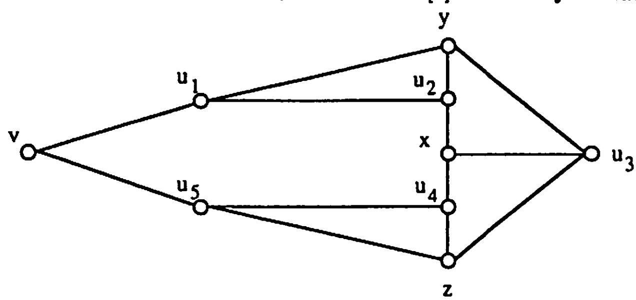

Figure 1: Graph H_1 for Theorem 1

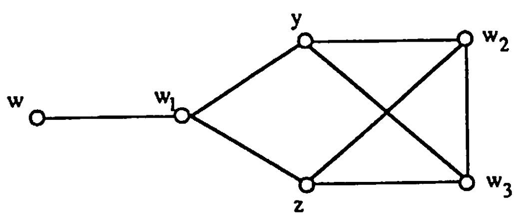

Figure 2: Graph H_2 for Theorem 1

NP-complete for regular graphs, hence VC remains NP-complete for regular graphs. Here we establish the stronger result, that f-VC is NP-complete for all $f \geq 3$. It is already known, see [1] and [7], that the problem we call Max(f)-VC is NP-complete for $f \geq 3$.

Theorem 1 *3-VC is NP-Complete.*

Proof: It is easy to see that 3-VC is in NP. To see that 3-VC is NP-complete we show that Max(3)-VC may be polynomially transformed into 3-VC.

For any instance I of Max(3)-VC, $G = (V, E)$ and $K \in Z^+$, suppose there are a vertices of degree 1 and b vertices of degree 2. The corresponding instance of 3-VC is $I' : G' = (V', E')$ and $K' = K + 5a + 3b$, where G' is obtained from G by attaching one copy of the graph H_1 shown in Figure 1 to each vertex v of degree 1 and one copy of the graph H_2 shown in Figure 2 to each vertex w of degree 2.

Then G' is 3-regular. We now show that G has a vertex cover S_1 with $|S_1| \leq K$ if and only if G' has a vertex cover S_1' with $|S_1'| \leq K'$. Suppose G has a vertex cover S_1 with $|S_1| \leq K$, then we can obtain a vertex cover S_1' of G' by adding to S_1 the five vertices which correspond to $u_1, \ldots, u_5$ in Figure 1 for each attached subgraph isomorphic to H_1 and the three vertices which correspond to w_1, w_2, w_3 in Figure 2 for each attached subgraph isomorphic to H_2. Since $|S_1'| = |S_1| + 5a + 3b \leq K + 5a + 3b = K'$, we get a desired vertex cover of G'.

On the other hand, suppose S_1' is a vertex cover of G' with $|S_1'| \leq K'$. In H_1, even if v and its incident edges are removed, the remaining graph can be partitioned into two triangles, (u_1, u_2, y), (u_4, u_5, z) and an edge (x, u_3), all vertex disjoint, therefore any vertex cover of $H_1 - \{v\}$ must contain at least 5 vertices. Similarly, in H_2, even if w and its incident edges are removed, the remaining graph can be partitioned into a triangle, (y, w_2, w_3) and an edge (w_1, z), which are vertex disjoint, therefore any vertex cover of $H_2 - \{w\}$ must contain at least 3 vertices. This establishes that any vertex cover of H_1 must contain at least 5 vertices of $H_1 - \{v\}$ and any vertex cover of H_2 must contain at least 3 vertices of $H_2 - \{w\}$. Let $S_1 = S_1' \cap V(G)$, then $|S_1| \leq K' - (5a + 3b) = K$ and S_1 is a vertex cover of G. Thus we obtain a desired vertex cover of G.

The above transformation can be computed in polynomial time so we have Max(3)-VC $\propto$ 3-VC. □

Theorem 2 *k-VC is NP-complete for each $k \geq 3$.*

Proof: The proof is by induction on $k \geq 3$. For $k = 3$, the result follows from Theorem 1. Assume the result is true for r-VC, where $3 \leq r \leq k - 1$. We now show that k-VC must be NP-complete, where $k \geq 4$. Clearly k-VC is in NP. We show that (k-1)-VC may be polynomially transformed into k-VC in two cases, depending on whether k is even or odd.
Case I: k is odd.

For any instance I of (k-1)-VC: a (k-1)-regular graph $G = (V, E)$ and a positive integer $K \leq |V|$, where $n = |V|$, the corresponding instance of k-VC is I': $G' = (V', E')$ and positive integer $K' = K + n \cdot k$, where G' is obtained from G by attaching to each vertex v a graph isomorphic to H as shown in Figure 3.

Then G' is k-regular. We now show that G has a vertex cover S_1 with $|S_1| \leq K$ if and only if G' has a vertex cover S_1' with $|S_1'| \leq K'$. Suppose that G has a vertex cover S_1 with $|S_1| \leq K$, then we can form a vertex cover S_1' of

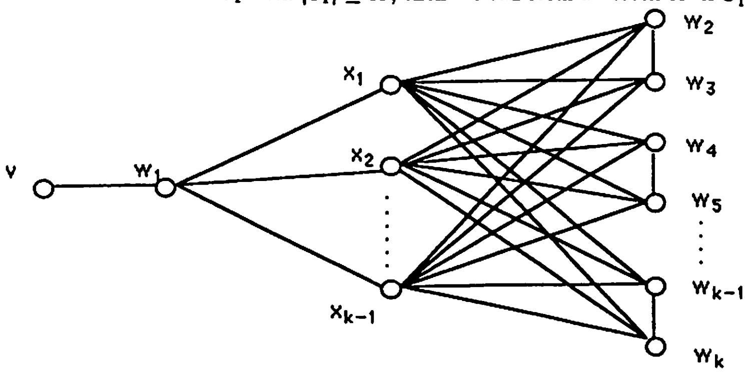

Figure 3: Graph H for odd values of k

G' by adding to S_1 the k vertices corresponding to $w_1, w_2, \ldots, w_k$ in Figure 3 for each attached subgraph. Since $|S_1'| = |S_1| + n \cdot k \leq K + n \cdot k = K'$ and S_1' is a vertex cover of G', we have a desired vertex cover for G'. On the other hand, suppose that G' has a vertex cover S_1' with $|S_1'| \leq K'$. Note that H has the property that each vertex cover of H contains at least k vertices of $H - \{v\}$, in fact the only set of k vertices that will suffice is the set $\{w_1, w_2, \ldots, w_k\}$. Let $S_1 = S_1' \cap V(G)$, then $|S_1| \leq |S_1'| - n \cdot k \leq K' - n \cdot k = K$. S_1 then, is a vertex cover of G of the desired form.

Case II: k is even.

For any instance I of (k-1)-VC: a (k-1)-regular graph $G = (V, E)$ and a positive integer $K \leq |V|$, where $n = |V|$, since k-1 is odd n must be even. Let $n = 2 \cdot m$, and $V(G) = \{v_1, v_2, \ldots, v_{2m}\}$. The corresponding instance of k-VC is I': $G' = (V', E')$ and positive integer $K' = K + n \cdot k$, where G' is obtained from G by attaching to each pair of vertices of G, v_{2i-1} and v_{2i}, a graph isomorphic to H as shown in Figure 4.

Then G' is k-regular. Note that H has the property that each vertex cover of H contains at least $2k$ vertices of $H - \{v_{2i-1}, v_{2i}\}$, in fact the only set of order $2k$ that will suffice is $\{w_1, w_2, \ldots, w_k, u_1, u_2, \ldots, u_k\}$. Then we can show in a manner similar to Case I that G has a vertex cover of size K or less if and only if G' has a vertex cover of size K' or less.

Note that the above transformations can be carried out in polynomial time. Thus we obtain a polynomial transformation from (k-1)-VC to k-VC for all values of k which are greater or equal to 4. Since (k-1)-VC is NP-complete, k-VC is also NP-complete. □

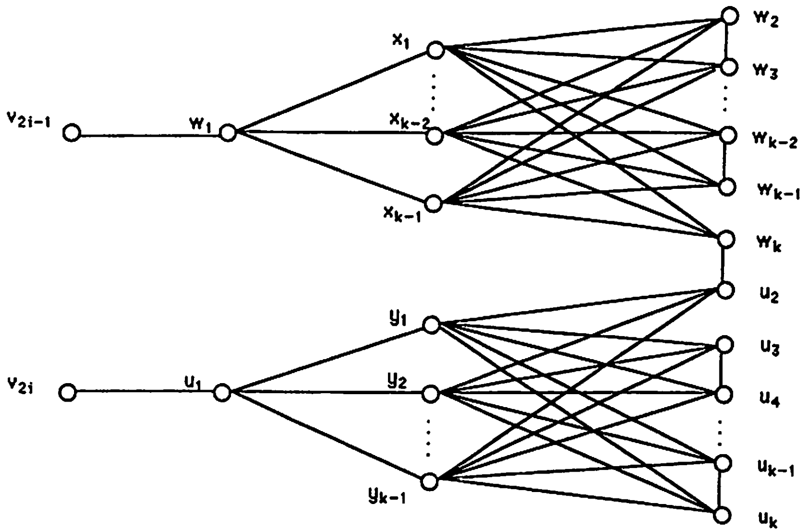

Figure 4: Graph H for even values of k

3 Complexity Results for (C, f)-Hitting Set

Since f-VC is equivalent to $(2, f)$-HS we know from the results in the previous section that $(2, f)$-HS is NP-complete for all $f \geq 3$. In this section we show that (C, f)-HS is NP-complete for all values of $C \geq 2$ and $f \geq 3$.

Let $S = \{s_1, s_2, \ldots, s_l\}$ and $D = \{d_1, d_2, \ldots, d_r\} \subseteq P(S)$ such that each element of S appears in exactly three subsets in D and $|d_i| = 2$ for $1 \leq i \leq r$. Then $3 \cdot l = 2 \cdot r$, that is $l = (2 \cdot r)/3$. It follows that $r \equiv 0 \bmod 3$. For $r = 3 \cdot t$ we obtain the following Lemma.

Lemma 1 *Let S, D, r and t be as described above. Then the elements of D may be reordered as d_1, $d_2, \ldots$, d_{2t-1}, d_{2t}, d_{2t+1}, $\ldots$, d_{3t} such that $d_{2i-1} \cap d_{2i} \neq \phi$ for $i = 1, 2, \ldots, t$.*

Proof: First we form a bipartite graph $G = (V, E)$ as follows. $V = V_1 \cup V_2$, where $V_1 = \{s_1, s_2, \ldots, s_l\}$ and $V_2 = \{d_1, d_2, \ldots, d_r\}$. s_i is adjacent to d_j if and only if $s_i \in d_j$. Then $deg(s_i) = 3$ for $1 \leq i \leq l$ and $deg(d_j) = 2$ for $1 \leq j \leq r$. Furthermore it is easy to see that it suffices to prove that G has t vertex-disjoint subgraphs of the form $H = (\{s_i, d_j, d_k\}, \{(s_i, d_j), (s_i, d_k)\})$ to establish the desired result. Suppose, to the contrary, that G has at most h vertex-disjoint subgraphs $B_1, B_2, \ldots, B_h$ of the form H, where $h < t$. Let $V_2' = V_2 \cap V(B)$ and $V_2^* = V_2 - V_2'$, where

$$B = \bigcup_{j=1}^{h} B_j.$$

Then G', the induced subgraph on G with vertex set $V - V_2'$, is a bipartite graph with partites V_1 and V_2^* and each vertex in V_2^* has degree 2 in G'. By the assumption for G, each vertex of V_1 has degree at most 1 in G' for otherwise there exists an $s_i \in V_1 - V(B)$ such that $deg_{G'}(s_i) \geq 2$, which then implies G has another subgraph B_{h+1} of the form H with $V(B_{h+1}) \cap V(B) = \phi$. Thus

$$\sum_{i=1}^{l} deg_{G'}(s_i) \leq l = (2r)/3 = 2t.$$

Since $h < t$, $|V_2^*| \geq 3t - 2h \geq t + 1$. It follows that

$$\sum_{v \in V_2^*} deg_{G'}(v) \geq 2 \cdot (t+1) > 2t \geq \sum_{i=1}^{l} deg_{G'}(s_i),$$

a contradiction. Therefore G has at least t vertex disjoint subgraphs of the form H, and the lemma follows. □

The reordering of elements in D may be carried out by an algorithm which arbitrarily selects the new d_1, then with a linear search finds d_2. A linear search (of remaining elements) can find a d_3, then another linear search can find a d_4 etc. This must be done t times. An appropriate data structure to maintain, for each item in S, the number of times the item has been "used" in a new d_i will be necessary. The overall reordering of elements of D can thus be carried out with $O(r^2)$ time complexity.

Theorem 3 *(3,3)-HS is NP-complete.*

Proof: Clearly (3,3)-HS is in NP. The proof is completed by a polynomial transformation from (2,3)-HS.

Let $I = (S = \{s_1, s_2, \ldots, s_l\}, D = \{d_1, d_2, \ldots, d_r\}, K)$ be an arbitrary instance of (2,3)-HS. Note that $r \equiv 0 \bmod 3$ since $r = (3 \cdot l)/2$. Therefore $t = r/3$ is an integer.

To form the corresponding instance of (3,3)-HS we first reorder the elements of D using the result of Lemma 1 so that $d_{3i-2} \cap d_{3i-1} \neq \phi$ for $i = 1, 2, \ldots, t$. For each three consecutive subsets in D of the form $d_{3i-2}, d_{3i-1}, d_{3i}$ for $i = 1, 2, \ldots, t$ we introduce the following 13 new elements of S': x_{3i-2}, x_{3i-1}, x_{3i}, b_1, b_2, b_3, a, c, e, f, g, y, z. In addition to these 13 new elements for each 3 consecutive subsets in D of the above form, S' will also contain all the original elements of S. For each 3 consecutive subsets in D of the above form, we include the following 15 new subsets in D': $d_{3i-2} \cup \{x_{3i-2}\}$, $d_{3i-1} \cup \{x_{3i-1}\}$, $d_{3i} \cup \{x_{3i}\}$, $\{x_{3i-2}, y, z\}$, $\{x_{3i-1}, y, z\}$, $\{x_{3i}, y, z\}$, $\{x_{3i-2}, a, b_1\}$, $\{x_{3i-1}, a, b_2\}$, $\{x_{3i}, a, b_3\}$, $\{b_1, c, e\}$, $\{b_2, c, f\}$, $\{b_3, c, g\}$, $\{b_1, f, g\}$, $\{b_2, e, g\}$, $\{b_3, e, f\}$. Then each variable in S' occurs in exactly 3 subsets in D' and the order of each subset in D' is exactly 3.

The complete instance, I', of (3,3)-HS consists of S', D' and $K' = K + 4t$. Now we show that I has a hitting set of order less than or equal to K if and only if I' has a hitting set of order less than or equal to K'.

Assume I has a hitting set S_1 with $|S_1| \leq K$. Then set S_1, together with $\{y, b_1, b_2, b_3\}$ for each triple $d_{3i-2}, d_{3i-1}, d_{3i}$, will hit each subset in D' so I' has a hitting set of order less than or equal to K'.

Assume I' has a hitting set $S_1' \subseteq S'$ with $|S_1'| \leq K + 4t$. We claim that I' must then have a hitting set S^* such that $|S^*| \leq K + 4t$ and no x_i is in S^* for $1 \leq i \leq 3t$. Let S^* be a hitting set of I' such that $|S^*| \leq K + 4t$ and S^* contains a minimum number of elements from $\{x_i | 1 \leq i \leq 3t\}$. Suppose, to the contrary, that the claim is not true and S^* does contain an x_i. Then there will be an i_0 such that at least one of x_{3i_0-2}, x_{3i_0-1}, x_{3i_0} is in S^*. Consider those subsets in D' corresponding to the 3 subsets in D: d_{3i_0-2}, d_{3i_0-1}, d_{3i_0}. It is easy to verify that at least 3 elements from $\{b_1, b_2, b_3, c, e, f, g\}$ are required to hit the subsets $\{b_1, c, e\}$, $\{b_2, c, f\}$, $\{b_3, c, g\}$, $\{b_1, f, g\}$, $\{b_2, e, g\}$, $\{b_3, e, f\}$. Whatever 3 of these elements are included in S^*, we replace them and instead include $\{b_1, b_2, b_3\}$. Then the remaining subsets that must be hit are $d_{3i_0-2} \cup \{x_{3i_0-2}\}$, $d_{3i_0-1} \cup \{x_{3i_0-1}\}$, $d_{3i_0} \cup \{x_{3i_0}\}$, $\{x_{3i_0-2}, y, z\}$, $\{x_{3i_0-1}, y, z\}$, $\{x_{3i_0}, y, z\}$.

If there are at most two elements of $\{x_{3i_0-2}, x_{3i_0-1}, x_{3i_0}\}$ in S^*, then either y or z must also be in S^*. In this case, we can replace each x_j of S^* by an arbitrary element of d_j, where $3i_0 - 2 \leq j \leq 3i_0$, to obtain a hitting set for I' which has at most $K + 4t$ elements and contains fewer elements from $\{x_i | 1 \leq i \leq 3t\}$ than S^*, contradicting the choice of S^*. If all three of x_{3i_0-2}, x_{3i_0-1}, x_{3i_0} are in S^*, then we may replace them by the following three elements:

1) y

2) an arbitrary element of d_{3i_0}

3) an element s such that $s \in d_{3i_0-2} \cap d_{3i_0-1}$

to obtain a hitting set for I' which has at most $K + 4t$ elements and contains fewer elements from $\{x_i | 1 \leq i \leq 3t\}$ than S^*, again contradicting the choice of S^*. Therefore, the claim is true.

Now let S^* be a hitting set of I' such that $|S^*| \leq K + 4t$ and S^* contains no element from $\{x_i | 1 \leq i \leq 3t\}$. Note that each hitting set of I' contains at least four of the 13 new elements for each triple $d_{3i-2}, d_{3i-1}, d_{3i}$. Let $S_1 = S^* \cap S$, then $|S_1| \leq K$ and S_1 is a hitting set of I. □

A generalized technique similar to that used in the above proof can be used to polynomially transform an arbitrary instance of (C, f)-HS to $(C + 1, f)$-HS for all $C \geq 2$ and $f \geq 3$. To apply the technique to $I = (S = \{s_1, s_2, \ldots, s_l\}, D = \{d_1, d_2, \ldots, d_r\}, K)$ we must have $r \equiv 0 \mod f$ so that $t = r/f$ will be an integer. If this is not the case an intermediate (C, f)-HS problem consisting of f copies of the entire original problem where $K' = f \cdot K$ will be required.

We now have complete results on (C, f)-HS for $f \geq 3$. In the next section we develop results for $f \in \{1, 2\}$ and establish the equivalence of HS to the Decision form of the Query Optimization Problem.

4 Relationship of the Query Optimization problem to Hitting Set

The problem of Query Optimization using prime keys was introduced in [2]. The formal description of the problem is: Let $Y = \{k_1, k_2, \ldots, k_n\}$ be the set of keys and let $A = \{(k_i, l_i, a_i), 1 \leq i \leq n\}$ be the directory, where k_i is a key in Y, l_i indicates the number of records in the file which contains k_i and a_i points to the first record of the list for k_i. For each key k_i, all records in the file containing k_i are linked. Let

$$Q = (k_{1,1} \wedge k_{1,2} \wedge \ldots \wedge k_{1,n_1}) \vee (k_{2,1} \wedge k_{2,2} \wedge \ldots \wedge k_{2,n_2}) \vee \ldots \vee (k_{m,1} \wedge k_{m,2} \wedge \ldots \wedge k_{m,n_m})$$

be a query over Y, where each $k_{i,j} \in Y$ and let $Q_i = \{k_{i,1}, \ldots, k_{i,n_i}\}$ for $1 \leq i \leq m$. A record always contains some subset of Y. A *file*, F,, over Y is a subset of $P(Y)$, i.e. a set of records over Y. Record R is said to *satisfy* Q if there exists some i, such that $Q_i \subseteq R$.

Assume that f_i = the count of those Q_j which contain k_i for $1 \leq i \leq n$ and $1 \leq j \leq m$. Let W = { all retrieval processes associated with Q and F }, V = $\{S \mid S \subseteq R$ and $S \cap Q_j \neq \phi$ for $1 \leq j \leq m\}$, then there is a one-to-one correspondence between W and V. Let $u \in W$ and S be the element of V that corresponds to u. We define $cost(u) = l_{i_1} + l_{i_2} + \ldots + l_{i_r}$ where S = $\{ k_{i_1}, k_{i_2}, \ldots, k_{i_r} \}$.

Decision form of Query Optimization Problem (DQOP)

Instance: A set of keys, $Y = \{k_1, k_2, \ldots, k_n\}$; a set of records over Y, $F = \{R_1, R_2, \ldots, R_s\}$; a directory set $A = \{(k_i, l_i, a_i), 1 \leq i \leq n\}$, for F; a query

$$Q = (k_{1,1} \wedge k_{1,2} \wedge \ldots \wedge k_{1,n_1}) \vee (k_{2,1} \wedge k_{2,2} \wedge \ldots \wedge k_{2,n_2}) \vee \ldots \vee (k_{m,1} \wedge k_{m,2} \wedge \ldots \wedge k_{m,n_m})$$

over Y ; and a bound $B \in Z^+$.

Question: Is there a retrieval process $u \in W$ with cost(u) $\leq B$ which is guaranteed to retrieve all records which satisfy Q ?

Let DQOPL be the subproblem of DQOP which contains only those instances with $l_1 = l_2 = \ldots = l_n = L$.

Let *DQOP1* be the subproblem of DQOPL which contains only those instances with $f_1 = f_2 = \ldots = f_n = 3$.

Let *DQOP2* be the subproblem of DQOPL which contains only those instances with $f_i \leq 2$ for all $1 \leq i \leq n$.

It was shown in [4] that DQOP1 is an NP-complete problem while DQOP2 has a polynomial solution.

Recall that there is a one-to-one correspondence between W and V. This implies the following equivalence between HS and DQOPL.

Theorem 4 *HS is polynomially equivalent to DQOPL.*

Theorem 4 is easily shown to be true for the following correspondence.

For each instance $I' = (Y, Q, F, A, B)$ of DQOPL, the corresponding instance of HS is $I = (S, D, K)$, where $S = Y$, $D = \{Q_1, Q_2, \ldots, Q_m\}$, and $K = B/L$. Going in the other direction, for each instance $I = (S, D, K)$ of HS, with $|S| = n$ and $|D| = m$, the corresponding instance of DQOPL is $I' = (Y, Q, F, A, B)$, where $Y = S$, $Q = C_1 \vee C_2 \vee \ldots \vee C_m$ with each C_i being the conjunction of

C \ f	1	2	3	4	......
1	P	X	X	X	......
2	P	P	NP-C	NP-C	......
3	P	P	NP-C	NP-C	......
⋮	⋮	⋮	⋮	⋮	⋱

Figure 5: Table summarizing (C, f)-Hitting Set complexity results (X means "meaningless".)

the elements of $d_i \in D$, $F \subseteq P(S)$ such that each element of S is present in L subsets in F, $A = \{(s_i, L, a_i), 1 \leq i \leq n\}$, and $B = K \cdot L$. □

As a consequence of Theorem 4 all the complexity results established for HS are also true for the corresponding DQOPL problems. Also, the following corollary follows directly from a result in [4].

Corollary 1 *The subproblem of HS where each element belongs to at most 2 subsets is polynomial.*

Alternatively, Corollary 1 can be established by a reduction of HS, where each element belongs to at most 2 subsets, to EC. Except for the issue of subsets containing only elements of frequency one, this restriction of HS is exactly EC. It is well known (see [1] and [5]) that EC can be solved in polynomial time. To reduce the restricted HS to EC, let D be an instance of restricted HS. First remove from D any subset which contains only elements of frequency one; for each such subset d removed, add any element of d to the hitting set. View the remaining subsets as a graph where the vertices are the subsets, and where d_i and d_j are connected by an edge if and only if $d_i \cap d_j \neq \phi$. A minimum edge cover of this graph corresponds to a minimum hitting set.

5 Conclusions

Each (C, f)-HS problem, for $C \geq 2$ and $f \in \{1, 2\}$, is a special case of the HS subproblem addressed in Corollary 1, therefore (C, f)-HS is polynomial for all $C \geq 2$ and $f \in \{1, 2\}$.

The table shown in Figure 5 summarizes many of the results established in this paper relating to the (C, f)-Hitting Set problem.

Acknowledgements: The authors thank an anonymous referee for the alternate proof of Corollary 1 and for other worthwhile suggestions.

References

[1] M. Garey and D. Johnson, "*Computers and Intractability*", Freeman and Company, 1979.

[2] D. Hsiao and F. Harary, "A formal system for information retrieval from files", *CACM*, vol. 13, pp. 67-73, 1970.

[3] R.M. Karp, "Reducibility among combinatorial problems", in R.E. Miller and J.W. Thatcher(eds.), *Complexity of Computer Computations*, Plenum Press, New York, 1972.

[4] D. Kountanis, J. Liu and K. Williams, "The Query Optimization Problem for a Generalized File Organization", *Proceedings of the Eighth International Conference on Mathematical and Computer Modelling*, 1991.

[5] E. Lawler, "*Combinatorial Optimization: Networks and Matroids*", Holt, Rinehart and Winston, 1976.

[6] U. Manber, "*Introduction to Algorithms*", Addison-Wesley, 1989.

[7] K. Mehlhorn, "*Graph Algorithms and NP-completeness*", Springer-Verlag, 1984.