Complexity of Extremal Set Decision Problem

M. Atici
Department of Computer Science
Western Kentucky University
Bowling Green KY 42101
mustafa.atici@wku.edu

Abstract

Let a set $[n] = \{1, 2, ..., n\}$ be given. Finding a subset $\mathcal{S}$ of $2^{[n]}$ with minimum cardinality such that, for any two distinct elements $x, y \in [n]$, there exists disjoint subsets $A_x, A_y \in \mathcal{S}$ such that $x \in A_x$, $y \in A_y$ is called the *extremal set* problem. In this paper, we define the Extremal Set Decision(ESD) Problem and study its complexity.

Keywords: Extremal set, hash function, complexity

1 Introduction

The combinatorial problem *extremal set* is defined in [1]. We first review the definition of the extremal set problem: Let $[n] = \{1, 2, \ldots, n\}$ be a given set. Find the minimum cardinality of a collection $\mathcal{S}$ of subsets of $[n]$ such that, for any two distinct elements $x, y \in [n]$ there exists $A_x, A_y \in \mathcal{S}$ such that

1. $x \in A_x$ and $y \in A_y$
2. $A_x \cap A_y = \phi$

Such a set $\mathcal{S}$ is called a *separating* set of the given set $[n]$.

The following theorem [1] gives a lower bound on the minimum cardinality of separating set $\mathcal{S} \subset 2^{[n]}$.

Theorem 1 *Let $f(n)$ denote the minimum cardinality of a separating set $\mathcal{S} \subset 2^{[n]}$. Then*

$$f(n) = \begin{cases} 3m & \text{if } 2 \times 3^{m-1} < n \leq 3^m \\ 3m+1 & \text{if } 3^m < n \leq 4 \times 3^{m-1} \\ 3m+2 & \text{if } 4 \times 3^{m-1} < n \leq 2 \times 3^m. \end{cases}$$

Separating sets have application to geodetics in graph theory and perfect hash families in data retrieval. Let G be a simple and unweighted graph. The *distance* between two vertices u, v in G is the minimum number of edges in a path joining the vertex u and the vertex v. Such a shortest path, denoted by $u - v$, is called a *geodesic*. Set $H(u, v)$ is defined to be the set of all edges lying on some $u - v$ geodesic of G. Let S be a subset of vertex set $V(G)$. A subset $H(S)$ of edge set $E(G)$ is defined to be $H(S) = \bigcup_{u,v \in S} H(u, v)$. A set $S \subseteq V(G)$ is called an *edge geodetic set* if $H(S) = E(G)$. *Edge geodetic number* $g_e(G)$ of a given graph G is

$$g_e(G) = min_{S \subseteq V(G)}\{|S| : S \text{ is an edge geodetic set of } G\}.$$

For a given simple graph G, obvious lower and upper bounds of the edge geodetic number $g_e(G)$ are 2 and $|V(G)|$, respectively. The following theorem [1] gives a lower bound of the edge geodetic number of a given graph G. This lower bound is driven from the separating set of $[\omega(G)]$, where $\omega(G)$ is the clique number of the graph G. For more on the edge geodetic number, we refer the reader to [3].

Theorem 2 *If $\omega(G)$ is the clique number and $g_e(G)$ the edge geodetic number of G, then*

$$g_e(G) \geq \lceil 3 \log_3 \omega(G) \rceil.$$

Moreover, for any n there exists a graph G with $\omega(G) = n$ that contains a geodetic set with $\lceil 3 \log_3 n \rceil + \epsilon$ vertices, where ϵ is 0 or 1.

The second application is to construct perfect hash families [2]. A *hash function* is a function $h : \{1, 2, \cdots, n\} \to \{1, 2, \cdots, w\}$, where $n > w$. A hash function h is said to be *perfect* on a subset X of $\{1, 2, \cdots, n\}$ if h is injective on X, i.e., if $h|_X$ is one-to-one. Perfect hash functions are useful for the compact storage and fast retrieval of frequently used data, such as reserved words in programming languages, command names in interactive systems, etc. For more information about perfect hash functions readers can consult the recent survey paper [5].

Let $\mathcal{F} = \{h | h : \{1, 2, ..., n\} \to \{1, 2, ..., w\}\}$ be a set of hash functions. $\mathcal{F}$ is called (n, w, k)- *perfect hash family* if there exists at least one $h \in \mathcal{F}$ such that $h|_X$ is one-to-one for any $X \subset \{1, 2, ..., n\}$ with $|X| = k$. The notation PHF$(N; n, w, k)$ is used to denote an (n, w, k)-perfect hash family with $|\mathcal{F}| = N$.

Let $\mathcal{S}$ be a separating set of given set $[n]$ with minimum cardinality $f(n)$. The following theorem [2] shows the existence of a perfect hash family $\mathcal{F}$ with certain values of w and k.

Theorem 3 *Let $n > 0$ be given integer. Then there exist PHF$(\frac{f(n)}{3}; n, 3, 2)$ if $2 \times 3^{m-1} < n \leq 3^m$ or PHF$(\lceil \frac{f(n)}{3} \rceil; n, 3, 2)$ if $3^m < n \leq 2 \times 3^m$.*

2 Special Separating Sets

In this section, we study separating set $\mathcal{S}$ which has restrictions on its elements. That is, the separating set $\mathcal{S}$ contains subsets of $[n]$ with certain cardinalities. The following is an obvious observation.

Lemma 4 *Let $[n] = \{1, 2, 3, ..., n\}$ be a given set and let $\mathcal{S} = \{A \subset [n] : |A| = 1\}$ be a separating set of $[n]$. Then $|\mathcal{S}| = n$.*

Next, assume $n = k \times m$, where $2 \leq k \leq m$. We are interested in the minimum cardinality of a separating set $\mathcal{S}$ of $[n] = [k \times m]$ with the following property: $\mathcal{S}$ contains a specific partition of set $[n]$ i.e. $\mathcal{S}$ contains B_i's such that $|B_i| = m$ for all i, $B_i \cap B_j = \phi$ for any $i \neq j$, and $\bigcup_{i=1}^{k} B_i = [n]$. A general description of such a separating set $\mathcal{S}$ can be given as

$$\mathcal{S} = \{X, B_i : |X| \geq 1, |B_i| = m, B_i \cap B_j = \phi \text{ for any } i \neq j, \text{ and } \bigcup_{i=1}^{k} B_i = [n]\}.$$

The following theorem gives the cardinality of such a separating set $\mathcal{S} \subset 2^{[n]}$.

Theorem 5 *Let $n = k \times m$, where $2 \leq k \leq m$ and let $\mathcal{S} = \{X, B_i : |X| \geq 1, |B_i| = m, B_i \cap B_j = \phi$ for any $i \neq j$, and $\bigcup_{i=1}^{k} B_i = [n]\}$. Then $|\mathcal{S}| = m + k$.*

Proof: First we construct a separating set $\mathcal{S}$ with cardinality at most $m+k$. In order to construct such a separating set $\mathcal{S}$; we construct a $0-1$ matrix D of size $(m+k) \times n$, that is, D is $(m+k) \times n$ matrix with 0 and 1 entries. We convert the matrix D into the separating set $\mathcal{S}$. Let I_j be a $m \times m$ identity(or permutation) matrix, $ONE = [1, 1, ..., 1]$ be $1 \times m$ all 1's matrix, and $ZERO = [0, 0, ..., 0]$ be $1 \times m$ all 0's matrix. Then we construct matrix D of size $(m + k) \times n$ as follows:

$D =$

I_1	I_2	...	I_k
ONE	$ZERO$	...	$ZERO$
$ZERO$	ONE	...	$ZERO$
.....	.....	.....	.....
$ZERO$	$ZERO$	...	ONE

Example Let $n = 12 = 3 \times 4$. Then

$D =$

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	0	0	0	1	0	0	0	1	0	0	0
2	0	1	0	0	0	1	0	0	0	1	0	0
3	0	0	1	0	0	0	1	0	0	0	1	0
4	0	0	0	1	0	0	0	1	0	0	0	1
5	1	1	1	1	0	0	0	0	0	0	0	0
6	0	0	0	0	1	1	1	1	0	0	0	0
7	0	0	0	0	0	0	0	0	1	1	1	1

m columns, which contains sub matrix I_i, of the matrix D is called *block i*. So the above matrix D consists of k blocks. The set $\mathcal{S}$ is constructed from the matrix $D = (d_{i,j})$, where $i = 1, 2, ..., m+k$ and $j = 1, 2, ..., n$ as follows:

$$\mathcal{S} = \{A_i : i = 1, 2, ..., m+k\}, \text{ where } A_i = \{j : d_{i,j} = 1\}.$$

Such a set $\mathcal{S}$ has m elements of size k and k elements of size m. That is

$$\begin{aligned}
&A_1 = \{1, m+1, ..., (k-1)m+1\} \\
&A_2 = \{2, m+2, (k-1)m+2\} \\
&... \\
&A_m = \{m, 2m, ..., km\} \\
&A_{m+1} = \{1, 2, 3, ..., m\} \\
&A_{m+2} = \{m+1, m+2, ..., 2m\} \\
&... \\
&A_{m+k} = \{(k-1)m+1, (k-1)m+2, ..., km\}.
\end{aligned}$$

We show the construction of the set $\mathcal{S}$ by using the above example. $A_1 = \{1, 5, 9\}$,$A_2 = \{2, 6, 10\}$, $A_3 = \{3, 7, 11\}$, $A_4 = \{4, 8, 12\}$ $A_5 = \{1, 2, 3, 4\}$, $A_6 = \{5, 6, 7, 8\}$, and $A_7 = \{9, 10, 11, 12\}$. Hence $\mathcal{S} = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7\}$.

We now show that the set $\mathcal{S}$, which is constructed from the matrix D, is a separating set for $[n] = [k \times m]$: Let x, y be two distinct elements of $[n]$. Then x, y correspond column indexes of the matrix D:

(i) If x, y are in same block, say they are in block i: Since I_i is a permutation matrix there are two rows, say r, s such that $d_{r,x} = 1$ and $d_{s,y} = 1$ in I_i. Hence $x \in A_r$ and $y \in A_s$ and obviously from the construction A_i's, we have $A_r \cap A_s = \phi$.

(ii) If x, y are in two different blocks, that is, x in block i and y is in block j: In this case, row x_i of the $i-$th block and row y_j of the $j-$th, where $x_i \neq y_j$ and $m < x_i, y_j \leq m+k$, contain sub matrix ONE. Hence $x \in A_{x_i}$, $y \in A_{y_j}$, and $A_{x_i} \cap A_{y_j} = \phi$. Therefore $\mathcal{S}$ is a separating set with cardinality $m+k$.

We need to show that any separating set $\mathcal{S}$ with the given property must have at least $k+m$ elements in it. Let $\mathcal{S}$ be a such separating set. That is, $\mathcal{S}$ contains partition $A_1, A_2, ..., A_k$ of $[n] = [k \times m]$, where $|A_i| = m$ for $i = 1, 2, ..., k$, $A_i \cap A_j = \phi$ for any $i \neq j$, and $\cup_{i=1}^{k} A_i = [n]$. Without loss of generality we can assume that $A_1 = \{x_1, x_2, ..., x_m\}$. Hence there must be at least m elements, say B_j's, in $\mathcal{S}$ such that $x_1 \in B_1$ and $x_j \in B_j$ for $j = 2, 3, ..., m$ such that $B_1 \cap B_j = \phi$ for $j = 2, 3, ..., m$. Hence $|\mathcal{S}| \geq m + k$. □

3 Complexity of an Extremal Set Decision Problem

In this section we first define the EXTREMAL SET DECISION PROBLEM and study its complexity.

EXTREMAL SET DECISION PROBLEM (ESD):

INSTANCE : Set $[n] = \{1, 2, ..., n\}$, $\mathcal{S}' \subseteq 2^{[n]}$, and integer k.

QUESTION : Is there a $\mathcal{S} \subset \mathcal{S}'$ with $|\mathcal{S}| = k$ such that, for any two distinct elements $x, y \in [n]$, there exists disjoint subsets $A_x, A_y \in \mathcal{S}$ such that $x \in A_x$, $y \in A_y$?

Theorem 6 *If $\mathcal{S}' = 2^{[n]}$ in* ESD, *then* ESD *has a polynomial solution.*

Proof: Let $[n] = \{1, 2, ..., n\}$ be a given set. Using Theorem 1 determine the value of integer m. Then set N to be 3^m or $4 \times 3^{m-1}$ or 2×3^m, that is, $N = q_1 \times q_2 \times ... \times q_k$, where q_i is 2 or 3 for $i = 1, 2, ..., k$. Then here is the algorithm to construct a separating set $\mathcal{S}$ for $[n]$.

ESD([N],$q_1, q_2, ..., q_k$)

1. Construct $N \times k$ matrix $M = (m_{s,t})$ as follows:

```
j=1
while(j ≤ N)
   for(i_1 = 1 to q_1)
      for(i_2 = 1 to q_2)
      .....
         for(i_k = 1 to q_k)
            m_{j,1} = i_1
            m_{j,2} = i_2
```

.....

$m_{j,k} = i_k$

j=j+1

2. Define $A_{ij} = \{r : m_{r,i} = j, 1 \leq i \leq k, 1 \leq j \leq q_i, 1 \leq r \leq N\}$
3. Define $\mathcal{S}(N) = \{A_{ij} : i = 1, 2, ..., k \text{ and } 1 \leq j \leq q_i\}$
4. Define $\mathcal{S} = \{B | B = A - \{x : n+1 \leq x \leq N\}, \text{ where } A \in \mathcal{S}(N)\}$
5. If $|\mathcal{S}| = k$ return "YES"
   Else return "NO"

We need to prove that the set $\mathcal{S}$ constructed by the above algorithm is a separating set for $[n]$. It has been proven in [1] that $\mathcal{S}(N)$ is a separating set of $[N]$. Suppose $x \neq y \in [n] = \{1, 2, ..., \} \subset \{1, 2, ..., N\}$. So there exist $A_x, A_y \in \mathcal{S}(N)$ such that $x \in A_x$, $y \in A_y$ and $A_x \cap A_y = \phi$. By the definition of $\mathcal{S}$ in line 4, $x \in B_x = A_x - \{w : n+1 \leq w \leq N\}$, $y \in B_y = A_y - \{w : n+1 \leq w \leq N\}$, and $B_x \cap B_y = \phi$. Hence $\mathcal{S}$ is an separating set of $[n]$. Run time of the above algorithm is $O(Nlog(N))$, where $N = |[N]|$. □

Example: Let $n = 7$. Then $7 < 3^2$ so $m = 2$ and $N = 3^2 = 9$. Hence $N = q_1 \times q_2 = 3 \times 3$. Therefore matrix M is:

$M = (m_{s,i})_{9 \times 2} =$

	1	2
1	1	1
2	1	2
3	1	3
4	2	1
5	2	2
6	2	3
7	3	1
8	3	2
9	3	3

Hence $\mathcal{S}(9) = \{A_{ij} | 1 \leq i \leq k, 1 \leq j \leq q_i\}$,where the A_{ij}'s are:

$A_{11} = \{r : m_{r,1} = 1, 1 \leq r \leq 9\} = \{1, 2, 3\}$
$A_{12} = \{r : m_{r,1} = 2, 1 \leq r \leq 9\} = \{4, 5, 6\}$
$A_{13} = \{r : m_{r,1} = 3, 1 \leq r \leq 9\} = \{7, 8, 9\}$
$A_{21} = \{r : m_{r,2} = 1, 1 \leq r \leq 9\} = \{1, 4, 7\}$
$A_{22} = \{r : m_{r,2} = 2, 1 \leq r \leq 9\} = \{2, 5, 8\}$
$A_{23} = \{r : m_{r,2} = 3, 1 \leq r \leq 9\} = \{3, 6, 9\}$

Therefore

$$B_1 = A_{11} - \{x : 8 \leq x \leq 9\} = \{1, 2, 3\}$$
$$B_2 = A_{12} - \{x : 8 \leq x \leq 9\} = \{4, 5, 6\}$$
$$B_3 = A_{13} - \{x : 8 \leq x \leq 9\} = \{7\}$$
$$B_4 = A_{21} - \{x : 8 \leq x \leq 9\} = \{1, 4, 7\}$$
$$B_5 = A_{22} - \{x : 8 \leq x \leq 9\} = \{2, 5\}$$
$$B_6 = A_{23} - \{x : 8 \leq x \leq 9\} = \{3, 6\}$$

and $\mathcal{S} = \{B_1, B_2, B_3, B_4, B_5, B_6\}$ is a separating set of $[7] = \{1, 2, 3, 4, 5, 6, 7\}$.

If a separating set $\mathcal{S}$ is constructed from a strict subset of $2^{[n]}$, then the problem becomes very difficult to solve.

Theorem 7 *If $\mathcal{S}' \subsetneq 2^{[n]}$ in ESD, then ESD is an NP-Complete problem.*

Proof: First we need to show that ESD is in the NP class. A nondeterministic polynomial algorithm can be given as follows:

1. Pick a random subset $\mathcal{S}$ of size k from $\mathcal{S}'$
2. Check whether $\mathcal{S}$ is a separating set or not for $[n]$

Line 1 can be done in k steps. For given set $[n]$, there are $\binom{n}{2}$ tuples $(x, y) \in [n] \times [n]$, where $x \neq y$. So Line 2 can be checked in $O(k^2 n^2) = O(n^2)$ steps. Therefore the above algorithm is polynomial in terms of $|[n]| = n$.

Next we give a deterministic and polynomial transformation from 3-DIMENSIONAL MATCHING(3DM) to ESD. Let us first review the definition of 3DM decision problem.

3-DIMENSIONAL MATCHING(3DM):

INSTANCE : A set $M \subseteq U = \{u_1, u_2, ..., u_m\} \times V = \{v_1, v_2, ..., v_m\} \times W = \{w_1, w_2, ..., w_m\}$, where U, V, and W are mutually disjoint sets.

QUESTION : Does M contain a *matching*, that is, a subset $E \subseteq M$ such that $|E| = m$ and no two elements of E agree in any coordinate?

We define deterministic and polynomial algorithm f : 3DM $\rightarrow$ ESD as follows: Let $M \subseteq U \times V \times W$, that is, $M = \{m_1 = (u_{i_1}, v_{j_1}, w_{k_1}), m_2 = (u_{i_2}, v_{j_2}, w_{k_2}), m_3 = (u_{i_3}, v_{j_3}, w_{k_3}), ..., m_t = (u_{i_t}, v_{j_t}, w_{k_t})\}$ be an instance of 3DM. Then $f(M)$, which is an instance of ESD, is defined as follows:

1. $f(M) = (m_{ij})_{(t+3)\times 3m}$ is $0 - 1$ matrix
2. The r-th row of $f(M)$ are all 0 except the entries corresponding the columns $u_{i_r}, v_{j_r}, w_{k_r}$, where $m_r = (u_{i_r}, v_{j_r}, w_{k_r}) \in M$ for $r = 1, 2, ..., t$
3. The last 3 rows; row $t + 1$, row $t + 2$, and row $t + 3$ are ONE ZERO ZERO, ZERO ONE ZERO, and ZERO ZERO ONE, respectively.

Here ONE is $1 \times m$ matrix of all 1's and ZERO is $1 \times m$ matrix of all 0's. Hence for any given instance of M, $f(M)$ is $0-1$ matrix of size $(t+3) \times 3m$.

$f(M) =$

	$u_1 \ldots u_m$	$v_1 \ldots v_m$	$w_1 \ldots w_m$
1	r_{11}	r_{12}	r_{13}
2	r_{21}	r_{22}	r_{23}
..	...	...	...
t	r_{t1}	r_{t2}	r_{t3}
$t+1$	ONE	ZERO	ZERO
$t+2$	ZERO	ONE	ZERO
$t+3$	ZERO	ZERO	ONE

Example Let $U = \{u_1, u_2, u_3\}$, $V = \{v_1, v_2, v_3\}$, and $W = \{w_1, w_2, w_3\}$ be three mutually disjoint sets and $M = \{m_1 = (u_1, v_1, w_1), m_2 = (u_2, v_3, w_2), m_3 = (u_3, v_2, w_3), m_4 = (u_1, v_3, w_2), m_5 = (u_2, v_1, w_1)\}$ be a subset of $U \times V \times W$. Then $f(M)$ is:

$f(M) =$

	u_1 u_2 u_3	v_1 v_2 v_3	w_1 w_2 w_3
1	1 0 0	1 0 0	1 0 0
2	0 1 0	0 0 1	0 1 0
3	0 0 1	0 1 0	0 0 1
4	1 0 0	0 0 1	0 1 0
5	0 1 0	1 0 0	1 0 0
6	1 1 1	0 0 0	0 0 0
7	0 0 0	1 1 1	0 0 0
8	0 0 0	0 0 0	1 1 1

Now let M be a YES instance of a 3DM problem. There exists $E \subseteq M$ with $|E| = m$ such that $E = \{e_j = (u_{i_j}, v_{i_j}, w_{i_j}) : j = 1, 2, ..., m\}$ is a matching. We construct $f(M)$ as described above. Using construction given in Theorem 5, the set $\mathcal{S} = \{A_i : i = 1, 2, ..., t, t+1, t+2, t+3\}$ is constructed from $f(M)$. We need to show that $\mathcal{S}$ is a separating set of $U \cup V \cup W$. Let x, y be two distinct elements of the set $U \cup V \cup W$. If $x, y \in U$(or in V or in W), then there exist $e_i, e_j \in E$ such that $e_i = (x, v_{i_k}, w_{i_l})$ and $e_j = (y, v_{i_s}, w_{i_r})$, where $v_{i_k} \neq v_{i_s}, w_{i_l} \neq w_{i_r}$. Therefore there exist two disjoint sets A_i, A_j in $\mathcal{S}$ such that $x \in A_i, y \in A_j$. If $x \in U$ and $y \in V$(or W). Then $x \in A_{t+1}, y \in A_{t+2}$. By the construction, sets A_{t+1} and A_{t+2} are disjoint subsets of $\mathcal{S}$.

Conversely, let $\mathcal{S}$ be a separating set of $U \cup V \cup W$ such that $\mathcal{S}$ contains partition A_1, A_2, A_3, where $|A_i| = m$ for $i = 1, 2, 3$. Without loss of generality we can assume that $A_1 = U$, $A_2 = V$, and $A_3 = W$. By Theorem 5 we know that $|\mathcal{S}| = m + 3$. Hence $\mathcal{S}$ must have m elements beside A_1, A_2, A_3. Assume that these are $B_1, B_2, ..., B_m$. Since $|A_i| = m$ for $i = 1, 2, 3$ and

S is a separating set, then each B_i must contain exactly one element from each U, V, W. That is nothing but a matching with m elements.

The above argument shows that M is YES instance of 3DM $\Leftrightarrow$ $f(M)$ is YES instance of ESD. The algorithm f is deterministic and polynomial which takes $O(tm)$ times, where $t = |M|, m = |U| = |V| = |W|$. We have showed that

1. ESD is in NP-class
2. f : 3DM $\rightarrow$ ESD is deterministic and polynomial

This concludes that ESD is NP-complete since 3DM is NP-complete [4]. □

References

[1] M. Atici and A. Vince. Geodetics in Graphs, an Extremal Set Problem, and Perfect Hash Families, *Graphs and Combinatorics* **18**:403-413, 2002.

[2] M. Atici. Data Retriavel and Extremel Set, *International Journal: Mathematical Manuscript(IJMM),* No.1, Vol **1**: 27-32, 2007.

[3] M. Atici. On the Edge Geodetic Number of a Graph, *Intern. J. Computer Math.,* No. 7,Vol **80**:853-861, 2003.

[4] M. R. Garey and D. S. Johnson. Computers and Intractability: A Guide to the Theory of NP-completeness. Freeman 1979.

[5] Z. J. Czech, G. Havas and B. S. Majewski. Perfect Hashing, *Theoretical Computer Science* **182**:1-143, 1997.