Graphs and Combinatorics (2002) 18:403–413

Graphs and Combinatorics

Geodesics in Graphs, an Extremal Set Problem, and Perfect Hash Families

M. Atici[1] and Andrew Vince[2]

[1] International Computer Institute, Ege University, Izmir, Turkey
[2] Department of Mathematics, University of Florida, Gainesville, FL 32611, USA. e-mail: vince@math.ufl.edu

Abstract. A set U of vertices of a graph G is called a *geodetic* set if the union of all the geodesics joining pairs of points of U is the whole graph G. One result in this paper is a tight lower bound on the minimum number of vertices in a geodetic set. In order to obtain that result, the following extremal set problem is solved. Find the minimum cardinality of a collection $\mathscr{S}$ of subsets of $[n] = \{1, 2, \ldots, n\}$ such that, for any two distinct elements $x, y \in [n]$, there exists disjoint subsets $A_x, A_y \in \mathscr{S}$ such that $x \in A_x$ and $y \in A_y$. This separating set problem can be generalized, and some bounds can be obtained from known results on families of hash functions.

1. Introduction

The topics in this paper were originally motived by a problem in graph theory. The *distance* between two vertices u, v in a graph G is the least number of edges in a path joining u and v. Any such shortest path is called a *geodesic*. A set U of vertices of G is called *geodetic* if the union of all the geodesics joining pairs of points of U is the whole graph G. Let $g(G)$ denote the minimum number of vertices in a geodetic set for G, and call $g(G)$ the *geodetic number* of G. It was originally suggested that $\omega(G)$ is a lower bound on $g(G)$, where $\omega(G)$ is the clique number of G. This, in fact, is far from the situation. In Section 3 we show that

$$g(G) \geq \lceil 3 \log_3 \omega(G) \rceil,$$

and this lower bound is tight in the following sense. 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 problem concerning geodetic sets in graphs is directly related to the following extremal set problem. Let $[n] = \{1, 2, \ldots, n\}$. Find the minimum cardinality $f(n)$ of a collection $\mathscr{S}$ of subsets of $[n]$ with the following property. For any two distinct elements $x, y \in [n]$ there exists disjoint subsets $A_x, A_y \in \mathscr{S}$

such that $x \in A_x$ and $y \in A_y$. In Section 2 an exact formula for $f(n)$ is obtained, namely,

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

for $m \geq 1$. It is easy to see that $f(n)$ is essentially $3 \log_3 n$. More precisely, $f(n) = \lceil 3 \log_3 n \rceil + \epsilon$, where ϵ is 0 or 1. The relation between this extremal set problem and the extremal graph problem concerning geodesics is discussed in Section 3.

The extremal set problem can be generalized as follows. For integers n and $n_1, n_2, \ldots, n_k$, determine the minimum cardinality $f(n; n_1, n_2, \ldots, n_k)$ of a collection $\mathscr{S}$ of subsets of $[n]$ with the following property. For any distinct subsets $U_1, U_2, \ldots, U_k \subseteq [n]$ with $|U_i| = n_i$, there exists pairwise disjoint subsets $A_1, A_2, \ldots, A_k$ in $\mathscr{S}$ such that $U_i \subseteq A_i$ for all i. Except in the case $k = 2, n_1 = n_2 = 1$, we are not able to give an exact formula for $f(n; n_1, n_2, \ldots, n_k)$. In certain other special cases, however, bounds on $f(n; n_1, n_2, \ldots, n_k)$ can be derived from known results on families of hash functions. Hash functions have been extensively studied by computer scientists because of their application in searching a database. These bounds are given in Section 4.

2. Extremal Set Problem

This section contains the proof of the main theorem concerning the extremal set problem. If a collection $\mathscr{S}$ of subsets of $[n]$ has the property that, for any two distinct elements $x, y \in [n]$, there exists disjoint subsets $A_x, A_y \in \mathscr{S}$ such that $x \in A_x$ and $y \in A_y$, we say that $\mathscr{S}$ *separates pairs*.

Theorem 1. *Let $f(n)$ denote the minimum cardinality of a collection $\mathscr{S}$ of subsets of $[n]$ that separates pairs. Then*

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

Proof. First note that $f(n)$ is non-decreasing. If $\mathscr{S}$ is a family of subsets of $[n+1]$ that separates pairs, then, after removing the element $n+1$ from each set in $\mathscr{S}$, the resulting family of subsets of $[n]$ will separate pairs.

We first show that there exists a collection of subsets of $[n]$ that separates pairs and with cardinality as given by the formula in the statement of the theorem. After that we show that there is no collection with fewer sets that separates pairs.

Let $\overline{n}$ be the least integer greater than or equal to n and of the form $3^m, 4 \cdot 3^{m-1}$ or $2 \cdot 3^m$ for some integer m. Let $\overline{n} = p_1 \dots p_k$, where p_i is either 2 or 3 for each i. Further, let

$$N = \{\alpha := (\alpha_1, \dots, \alpha_k) \mid 1 \leq \alpha_i \leq p_i\}$$

and

$$A_{ij} = \{\alpha \in N \mid \alpha_i = j\}, \qquad \mathscr{S} = \{A_{ij} \mid 1 \leq j \leq p_i, 1 \leq i \leq k\}.$$

Clearly $|N| = \overline{n}$ and

$$|\mathscr{S}| = \sum_i p_i = \begin{cases} 3m & \text{if } \overline{n} = 3^m \\ 3m+1 & \text{if } \overline{n} = 4 \cdot 3^{m-1} \\ 3m+2 & \text{if } \overline{n} = 2 \cdot 3^m. \end{cases}$$

The collection $\mathscr{S}$ of subsets of N separates pairs of elements of N because any two distinct elements of N must differ in at least one coordinate position α_i and hence must belong, respectively, to two distinct elements of $\mathscr{S}$. Let

$$g(\overline{n}) = \begin{cases} 3m & \text{if } \overline{n} = 3^m \\ 3m+1 & \text{if } \overline{n} = 4 \cdot 3^{m-1} \\ 3m+2 & \text{if } \overline{n} = 2 \cdot 3^m. \end{cases}$$

Then we have shown that

$$g(\overline{n}) \geq f(\overline{n}) \geq f(n), \tag{2}$$

the second inequality following from the monotonicity of $f(n)$.

It now suffices to prove, by induction on n, that $f(n) = g(\overline{n})$. This is easy to verify for $n = 2, 3, 4$. Assume it is true for all natural numbers less than n.

Case 0. n is not of the form $3^m + 1, 4 \cdot 3^{m-1} + 1$ or $2 \cdot 3^m + 1$.

Then $g(\overline{n}) \geq f(n) \geq f(n-1) = g(\overline{n-1}) = g(\overline{n})$, the first inequality by formula (2) and the first equality by the induction hypothesis. Therefore $f(n) = g(\overline{n})$.

Case 1. $n = 3^m + 1$ for some m.

Because $3m + 1 = g(\overline{n}) \geq f(n) \geq f(n-1) = g(\overline{n-1}) = 3m$, it must be the case that $f(n) = 3m$ or $f(n) = 3m + 1$. It suffices to show that $f(n) \neq 3m$. Let $\mathscr{S}$ be the collection of subsets of minimal cardinality that separates pairs, and let A be a largest element of $\mathscr{S}$. There must be a $B \in \mathscr{S}, B \cap A = \emptyset$; otherwise A could be removed, contradicting the minimality of $\mathscr{S}$. Let $|A| = a, |B| = b$ and $c = n - a - b$.

If $a \leq 3^{m-1}$ then $a + c = n - b \geq n - a \geq 2 \cdot 3^{m-1} + 1$. This implies, by the induction hypothesis, that $f(a + c) \geq 3m$. This, in turn, implies that $f(n) \geq f(a + c) + 1 \geq 3m + 1$, the +1 in the first inequality being the set B. So, if $a \leq 3^{m-1}$ we are done.

If $a > 4 \cdot 3^{m-2}$ then , by the induction hypothesis, $f(a) \geq 3m - 1$. This implies that $f(n) \geq f(a) + 2 = 3m + 1$, the +2 being the sets A and B. Hence we are done unless

$$3^{m-1} < a \leq 4 \cdot 3^{m-2}.$$

Now $a \leq 4 \cdot 3^{m-2}$ implies that $a + c = n - b \geq n - a \geq 5 \cdot 3^{m-2} > 4 \cdot 3^{m-2}$. Hence, by the induction hypothesis, $f(a + c) \geq 3m - 1$. Assume, by way of contradiction, that $f(n) = 3m$. Then it must be the case that pairs in $[n] \backslash B$ are separated by exactly $3m - 1$ of the sets in $\mathscr{S}$, including the set A. Since $f(a + c) \geq 3m - 1$, the set A is essential; thus there must exists a set $C \in \mathscr{S}, C \neq B, C \cap A = \emptyset$. But $a > 3^{m-1}$ implies, by the induction hypothesis, that $f(a) \geq 3m - 2$, which implies that $f(n) \geq f(a) + 3 = 3m + 1$, the +3 being the sets A, B, C. This contradicts the assumption $f(n) = 3m$.

Case 2. $n = 2 \cdot 3^m + 1$ for some m.

Because $3m + 3 = g(\overline{n}) \geq f(n) \geq f(n - 1) = g(\overline{n - 1}) = 3m + 2$, it must be the case that $f(n) = 3m + 2$ or $f(n) = 3m + 3$. It suffices to show that $f(n) \neq 3m + 2$. As in case 1, let $\mathscr{S}$ be a collection of subsets of minimal cardinality that separates pairs, and let A be a largest element of $\mathscr{S}$. There must be a $B \in \mathscr{S}, B \cap A = \emptyset$; otherwise A could be removed, contradicting the minimality of $\mathscr{S}$. Let $|A| = a, |B| = b$ and $c = n - a - b$.

If $a \leq 2 \cdot 3^{m-1}$ then $a + c = n - b \geq n - a \geq 4 \cdot 3^{m-1} + 1$. This implies, by the induction hypothesis, that $f(a + c) \geq 3m + 2$. This, in turn, implies that $f(n) \geq f(a + c) + 1 \geq 3m + 3$, the +1 in the first inequality being the set B. So, if $a \leq 2 \cdot 3^{m-1}$ we are done.

If $a > 3^m$, then, by the induction hypothesis, $f(a) \geq 3m + 1$. This implies that $f(n) \geq f(a) + 2 = 3m + 3$, the +2 being the sets A and B. Hence we are done unless

$$2 \cdot 3^{m-1} < a \leq 3^m.$$

Now $a \leq 3^m$ implies that $a + c = n - b \geq n - a \geq 3^m + 1$. Hence, by the induction hypothesis, $f(a + c) \geq 3m + 1$. Assume, by way of contradiction, that $f(n) = 3m + 2$. Then it must be the case that pairs in $[n] \setminus B$ are separated by exactly $3m + 1$ of the sets in $\mathscr{S}$, including the set A. Since $f(a + c) \geq 3m + 1$, the set A is essential; thus there must exists a set $C \in \mathscr{S}, C \neq B, C \cap A = \emptyset$. But $a > 2 \cdot 3^{m-1}$ implies, by the induction hypothesis, that $f(a) \geq 3m$, which implies that $f(n) \geq f(a) + 3 = 3m + 3$, the +3 being the sets A, B, C. This contradicts the assumption $f(n) = 3m + 2$.

Case 3. $n = 4 \cdot 3^{m-1} + 1$ for some m.

Because $3m + 2 = g(\overline{n}) \geq f(n) \geq f(n - 1) = g(\overline{n - 1}) = 3m + 1$, it must be the case that $f(n) = 3m + 1$ or $f(n) = 3m + 2$. It suffices to show that $f(n) \neq 3m + 1$. Let the notation be exactly as for cases 1 and 2 above. The proof in case 3 is, however, slightly more involved.

If $a \leq 3^{m-1}$ then $a + c = n - b \geq n - a \geq 3^m + 1$. This implies, by the induction hypothesis, that $f(a + c) \geq 3m + 1$. This, in turn, implies that $f(n) \geq f(a + c) + 1 \geq 3m + 2$, the $+1$ in the first inequality being the set B. So, if $a \leq 3^{m-1}$ we are done.

If $a > 2 \cdot 3^{m-1}$, then, by the induction hypothesis, $f(a) \geq 3m$. This implies that $f(n) \geq f(a) + 2 \geq 3m + 2$, the $+2$ being the sets A and B. Hence we are done unless

$$3^{m-1} < a \leq 2 \cdot 3^{m-1}.$$

We now make the following additional assumption:

$$a > 4 \cdot 3^{m-2}. \tag{3}$$

Since $a \leq 2 \cdot 3^m$, we have $a + c = n - b \geq n - a \geq 2 \cdot 3^{m-1} + 1$. Hence, by the induction hypothesis, $f(a + c) \geq 3m$. Assume, by way of contradiction, that $f(n) = 3m + 1$. Then it must be the case that pairs in $[n] \backslash B$ are separated by exactly $3m$ of the sets in $\mathscr{S}$, including the set A. Since $f(a + c) \geq 3m$, the set A is essential; thus there must exists a set $C \in \mathscr{S}, C \neq B, C \cap A = \emptyset$. But, according to the assumption (3), we have $a > 4 \cdot 3^{m-2}$ which implies, by the induction hypothesis, that $f(a) \geq 3m - 1$. This, in turn, implies that $f(n) \geq f(a) + 3 \geq 3m + 2$, the $+3$ being the sets A, B, C. This contradicts the assumption $f(n) = 3m + 1$.

In view of assumption (3) we are now done unless

$$3^{m-1} < a \leq 4 \cdot 3^{m-2}. \tag{4}$$

Since $a \leq 4 \cdot 3^{m-2}$, we have $a + c = n - b \geq n - a \geq 8 \cdot 3^{m-2} + 1$. Hence, by the induction hypothesis, $f(a + c) \geq f(8 \cdot 3^{m-2} + 1) = f(2 \cdot 3^{m-1} + 1) = 3m$. Assume, by way of contradiction, that $f(n) = 3m + 1$. Then it must be the case that pairs in $[n] \backslash B$ are separated by exactly $3m$ of the sets in $\mathscr{S}$, including the set A. Since $f(a + c) \geq 3m$, the set A is essential; thus there must exists a set $C \in \mathscr{S}, C \neq B, C \cap A = \emptyset$. It now must be the case that pairs in $E := [n] \setminus (B \cup C)$ are separated by $3m - 1$ sets in $\mathscr{S}$. Let $e = |E|$. But $e \geq n - 2a \geq 4 \cdot 3^{m-2} + 1$, the last inequality by formula (4). By the induction hypothesis $f(e) \geq 3m - 1$. Therefore the set A is essential in separating pairs in E; so there must exists a set $D \in \mathscr{S}, D \neq B, D \neq C, D \cap A = \emptyset$. Formula (4) implies that $f(a) \geq f(3^{m-1} + 1)$ $= 3m - 2$, the equality by the induction hypothesis. This implies that $f(n) \geq f(a) + 4 \geq 3m + 2$, the $+4$ being the sets A, B, C, D. This contradicts the assumption $f(n) = 3m + 1$. □

3. Geodetic Sets of Graphs

A set U of vertices of a graph G is defined to be *geodetic* if the union of all the geodesics joining pairs of points of U is the whole graph G. This concept derives from a similar notion due to Harary, Loukakis, and Tsours [8]. They define a set U of vertices of G to be *geodetic* if the union of all the geodesics joining pairs of points of U contain all the vertices of G (but not necessarily all the edges). To emphasize the distinction, let $g_v(G)$ denote the minimum number of vertices in a geodetic in the sense of Harary, Loukakis, and Tsours and $g_e(G)$ the minimum in our sense. They will be referred to as the *vertex* and *edge geodetic numbers*, respectively. If G is a nontrivial connected graph, then it is obvious that

$$g_v(G) \leq g_e(G) \leq g_v(G) + (n-2). \tag{5}$$

Equality between $g_v(G)$ and $g_e(G)$ holds in (5), for example, when G is a tree, a cycle, a complete, or a complete bipartite graph. In the case of a tree, a minimum geodetic set is the set of endpoints. In the case of a cycle a minimal geodetic set consists of either two or three vertices, depending on whether the cycle is even or odd, respectively. The geodetic number of K_n is n and the geodetic number of $K_{m,n}$ is $\min\{n, m\}$. In fact, it is easy to show that the geodetic number is n for any n vertex graph that has two adjacent vertices each of which has degree $n-1$. Equality can also hold on the right side of (5). Take $G = K_n - e$, the complete graph with an edge deleted. A minimal vertex geodetic set consists of the two endvertices of the edge e. An edge geodetic set must consists of all n vertices. We refer to [2, 3, 4, 8] for more details on the vertex geodetic number of a graph G.

What follows is the proof of the lower bound given in the introduction for the edge geodetic number (which we hereafter refer to simply as the geodetic number).

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

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

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.

Proof. Assume that the largest clique in G, denoted K_n, has order n. Let U be a geodetic set of G. For any geodesic γ between $u \in U$ and $v \in U$, denote by u_γ the first vertex of γ, starting at u, that lies in K_n (if it exists), and let γ_u denote the subpath of γ from u to u_γ.

Since U is a geodetic set, the set of geodesics joining pairs of vertices of U must cover every edge of K_n. From here on we consider only geodesics that contain an edge of K_n. We claim that, if α and β are two such geodesics with the same endvertex $u \in U$, then the lengths of α_u and β_u are equal. Assume not, that α_u is longer than β_u. Let geodesic α join vertices $u, v \in U$, and let w be the last vertex on α that is also in K_n. Consider the path consisting, in sequence, of β_u, followed by the edge of K_n from u_β to w, then followed by the subpath of α from w to v. This is a path from u to v that is shorter than α, contradicting the fact that α is a geodesic. Let

$$A_u = \{u_\gamma \mid \gamma \text{ is a geodesic with endvertex } u\}, \qquad \mathcal{S} = \{A_u \mid u \in U\}.$$

We have just proved that the distances from u to the vertices in A_u are all the same.

Denote the vertices of K_n by $\{1,2,\ldots,n\}$, so that $\mathscr{S}$ can be considered as a collection of subsets of $[n]$. We next show that, in the terminology of Section 2, the collection $\mathscr{S}$ separates pairs in $[n]$. Let i,j be any two distinct vertices of K_n. Since U is a geodetic set, there is a geodesic γ from, say $u \in U$, to, say $v \in U$, containing the edge $\{i,j\}$. Then $u_\gamma = i$ and $v_\gamma = j$ (or vice versa); otherwise the path obtained by concatenating the shortest path from u to u_γ, followed by the edge $\{i,j\}$, followed by the shortest path from v_γ to v would be shorter than γ, contradicting the fact that γ is a geodesic. Therefore $i \in A_u$ and $j \in A_v$. Moreover, $A_u \cap A_v = \emptyset$. Otherwise, if there exists a vertex $w \in A_u \cap A_v$, then, by what we proved in the paragraph above, there would be a path from u to w with the same length as the path from u to i along γ; and a path from v to w with the same length as the path from v to j along γ. This would imply that the concatenated path from u to v through w would be shorter than γ, contradicting the fact that γ is a geodesic. Therefore $|U| = |\mathscr{S}| \geq f(n)$ by Theorem 1. It follows that $g(G) \geq |U| \geq f(n) = f[\omega(G)] \geq \lceil 3\log_3 \omega(G) \rceil$.

We next show that this lower bound is tight. In fact, this bound can be realized as follows. Consider a complete graph K_n together with a set U of $f(n)$ additional vertices. According to Theorem 1, there exists a collection $\mathscr{S} = \{A_u | u \in U\}$ of subsets of $[n]$, i.e., subsets of vertices of the complete graph K_n, that separate pairs. For each $u \in U$, join vertex u, by a single edge, to each vertex in A_u. Call the resulting graph G. Since $A_u \neq [n]$ for all $u \in U$, we have $\omega(G) = n$. Moreover, for each edge $\{i,j\}$ of K_n there exist $u, v \in U$ such that $i \in A_u, j \in A_v$ and $A_u \cap A_v = \emptyset$. Hence there exists a geodesic between u and v of length 3 containing $\{i,j\}$. So U is a geodetic set of cardinality $f(n)$; in other words $g(G) = f(n)$. □

4. Set Separation and Perfect Hash Families

In this section a few comments are made on the following generalization of the extremal set problem. For integers n and $n_1, n_2, \ldots, n_k$, determine the minimum cardinality $f(n; n_1, n_2, \ldots, n_k)$ of a collection $\mathscr{S}$ of subsets of $[n]$ with the following property. For any distinct subsets $U_1, U_2, \ldots, U_k \subseteq [n]$ with $|U_i| = n_i$, there exists pairwise disjoint subsets $A_1, A_2, \ldots, A_k$ with $U_i \subseteq A_i$ for all i. Two special cases are considered in this section:

1. $n_1 = n_2 = \cdots = n_k = 1$
2. $k = 2$

In the first case, a collection $\mathscr{S}$ that satisfies the required property will be said to *separate k-tuples*, and the minimum size of such a collection will be denoted $f(n,k)$. Note that $f(n,2) = f(n)$ in the notation of Theorem 1. To obtain a bound on $f(n,k)$ we consider families of hash functions. An (n,m,k)-*perfect hash family* (PHF) is a set $\mathscr{F}$ of functions such that $h : [n] \to [m]$ for each $h \in \mathscr{F}$, and, for any $X \subseteq [n]$ such that $|X| = k$, there exists an least one $h \in \mathscr{F}$ such that $h|_X$ is one-to-one. Note that, if

$$\mathscr{S}(\mathscr{F}) = \{h^{-1}(i) \mid h \in \mathscr{F}, 1 \leq i \leq m\},$$

then $\mathscr{S}$ separates k-tuples. For example, the $(9,3,2)$-PHF

$\mathscr{F} =$

h	1	2	3	4	5	6	7	8	9
h_1	1	1	1	2	2	2	3	3	3
h_2	1	2	3	1	2	3	1	2	3

provides the following collection of subsets of [9] that separates pairs:

$$\mathscr{S}(\mathscr{F}) = \{A_1, A_2, A_3, A_4, A_5, A_6\}$$

where

$$\begin{aligned} A_1 &= h_1^{-1}(1) = \{1,2,3\}, \\ A_2 &= h_1^{-1}(2) = \{4,5,6\}, \\ A_3 &= h_1^{-1}(3) = \{7,8,9\}, \\ A_4 &= h_2^{-1}(1) = \{1,4,7\}, \\ A_5 &= h_2^{-1}(2) = \{2,5,8\}, \\ A_6 &= h_2^{-1}(3) = \{3,6,9\}. \end{aligned}$$

In general, the following relationship between the k-tuple separation problem and perfect hash families will be used in this section.

Lemma 3. *If $P(n,m,k)$ denotes the minimum number of hash functions in an (n,m,k)-PHF, then*

$$f(n,k) \leq \min_{m \geq 2} \, [m \cdot P(n,m,k)].$$

It is known [1] that the minimal number of functions in a perfect hash family, with $k = 2$, is $\lceil \log n / \log m \rceil$. Lemma 3 implies the bound

$$f(n) = f(n,2) \leq \min_{m \geq 2} m \lceil \log n / \log m \rceil \tag{6}$$

for the original extremal set problem. Note, however, that equality does not, in general, hold in inequality (6). For example, if $n = 12$, then $f(n) = 7$ by Theorem 1, but $\min_{m \geq 2} m \lceil \log n / \log m \rceil = 8$ (with $m = 2$).

Although results on perfect hash function do not improve or, for that matter, do not imply Theorem 1, they provide upper bounds for values of $k \geq 3$. Using elementary counting, Mehlhorn [9] shows that there exist (n,m,k)-PHFs of size at most

$$\left\lceil \frac{\log \binom{n}{k}}{\log(m^k) - \log\left(m^k - k!\binom{m}{k}\right)} \right\rceil,$$

which implies the simpler (but less accurate) bound

$$\lceil ke^{k^2/m} \ln n \rceil .$$

By Lemma 3, this implies $f(n,k) \leq \lceil \min_m mke^{k^2/m} \ln n \rceil$. The minimum is attained when $m = k^2$, giving the following upper bound.

Corollary 4. $f(n,k) \leq \lceil k^3 e \ln n \rceil$.

Thus, for fixed k, we have $f(n,k) = O(\log n)$. The proof of Melhorn's bound, however, is not constructive, and, even for $k = 3$, we have no examples of families $\mathscr{S}$ of size $O(\log n)$ with the separating property.

The authors of [1] do construct (n,m,k)-PHFs, but the size is polynomial in $\log n$ rather than linear. To be precise, for arbitrarily large values of n, they construct an (n, m_0, k)-PHF of size

$$\frac{N_0}{(\log n_0)^{\log\left(\binom{k}{2}+1\right)}} (\log n)^{\log\left(\binom{k}{2}+1\right)}. \tag{7}$$

The log is base 2, and the constants N_0, n_0, m_0 can take any values for which there exists a (n_0, m_0, k)-PHF of size N_0 where $\gcd(n_0, \binom{k}{2}!) = 1$. Lemma 3 then implies the following result, which is constructive but fairly far from the theoretical bound of Corollary 4.

Corollary 5. *There are explicitly constructed families of subsets of* $[n]$, *for arbitrarily large values of n, for which*

$$f(n,k) = O\Big((\log n)^{\log(\binom{k}{2}+1)}\Big).$$

The method used in [1] for constructing perfect hash families of size given in formula (7) is recursive, constructing a (n^2, m, k)-PHF of size $\left(\binom{k}{2} + 1\right)N$ from a (n, m, k)-PHF of size N. As examples, consider the following $(5,3,3)$-PHF and $(5,4,4)$-PHF, respectively.

h	1	2	3	4	5
h_1	1	2	3	1	2
h_2	3	1	2	2	3
h_3	2	3	1	3	1

h	1	2	3	4	5
h_1	1	2	3	4	4
h_2	1	2	2	3	4
h_3	1	1	2	3	4

Using these perfect hash families as seed, formula (7) implies the following result.

Corollary 6. *There are explicitly constructed families of subsets of* $[n]$, *for arbitrarily large values of n, for which*

$$f(n,3) \leq \frac{9}{(\log 5)^2}(\log n)^2 \approx 1.66(\log n)^2$$

$$f(n,4) \leq \frac{12}{(\log 5)^{\log 7}}(\log n)^{\log 7} \approx 1.12(\log n)^{\log 7}.$$

We now turn to the second case of the general problem, to determine $f(n; n_1, n_2)$. In this case we must provide a collection of subsets of $[n]$ sufficient to separate any two subsets of given sizes n_1 and n_2, respectively. This problem is closely related to a problem of Friedman, Graham, and Ullman [7] that was motivated by considerations involving asynchronous sequential circuits. An (n_1, n_2)-*separating system* is a family $\mathscr{P} = \{(P_i, Q_i)\}$ of partitions of $[n]$ into two blocks satisfying the following property. For each pair of disjoint subsets U_1, U_2 of $[n]$ there is at least one partition $(P, Q) \in \mathscr{P}$ such that $U_1 \subseteq P$ and $U_2 \subseteq Q$ or $U_1 \subseteq Q$ and $U_2 \subseteq P$. Using probabilistic methods, Fredman and Komlós [6] prove that the minimum size of a (n_1, n_2)-separating system is

$$O(Z(n_1, n_2) \log n)$$

where

$$Z(i,j) = \frac{(i+j)^{i+j+1}}{i^i j^j}$$

and $\log n$ is base 2. This immediately implies the following bound.

Corollary 7. $f(n; n_1, n_2) = O(Z(n_1, n_2) \log n)$.

Thus, for n_1, n_2 fixed, $f(n; n_1, n_2)$ is $O(\log n)$. As in the first special case of the general separating set problem, except when $n_1 = n_2 = 1$, the result of Corollary 7 is not constructive. We have no constructive example of a collection of subsets of $[n]$, for arbitrarily large n, whose size is $O(\log n)$.

References

1. Atici, M., Magliveras, S.S., Stinson D.R., Wei, W.D.: Some recursive constructions for perfect hash families. J. Comb. Designs **4**, 353–363 (1996)
2. Atici, M.: Graph operations and their geodetic number. Congr. Numerantium **141**, 95–110 (1999)
3. Chartrand, G., Harary, F., Zhang, P.: Geodetic number of a graph. Networks 39, 1–6 (2002)
4. Chartrand, G., Harary, F., Zhang, P.: Geodetic sets in graphs. Discuss. Math Graph Theory 20, 129–138 (2000)
5. Czech, Z.J., Havas, G., Majewsi, B.S.: Perfect hashing. Theor. Comput. Sci. **182**, 1–143 (1997)

6. Fredman, M.L., Komlós, J.: On the size of separating systems and families of perfect hash functions. SIAM J. Algebraic Discrete Methods **5**, 61–68 (1984)
7. Friedman, A.D., Graham, R.L., Ullman, J.D.: Universal single transition time asynchronous state assignments. IEEE Trans. Comput. **C-18**, 541–547 (1969)
8. Harary, F., Loukakis, E., Tsours, C.: The geodetic number of a graph. Math. Comput. Modeling **17**, 89–95 (1993)
9. Mehlhorn, K.: On the program size of perfect and universal hash functions. In Proc. 23rd Annual IEEE Symp. Foundations of Computer Science pp. 190–175, 1982

Received: May 19, 2000
Final version received: July 5, 2001