ON CONSTANT WEIGHT CODES AND HARMONIOUS GRAPHS

by

R. L. Graham
N. J. A. Sloane
Bell Laboratories
Murray Hill, N.J. 07974

Introduction

Very recently a new method has been developed (see [3], [5], [6]) for finding lower bounds on the maximum number of codewords possible in a code of minimum distance d and length n. This method has led in turn to a number of interesting questions in graph theory and additive number theory. In this brief survey we summarize some of these developments.

Background

By a code C of length n over a finite field F = GF(q) we mean a subset of F^n, i.e., a set of n-tuples with entries in F. The most common choice for F is GF(2), and we restrict ourselves to this case for the remainder of the paper (although the same techniques apply to all finite fields). In this case C is called a binary code.

The minimum distance of C is defined to be

$$\min_{\bar{x} \neq \bar{y}} d(\bar{x},\bar{y})$$

where $\bar{x} = (x_1,\ldots,x_n)$ and $\bar{y} = (y_1,\ldots,y_n)$ range over all pairs of codewords (= elements of C) and $d(\bar{x},\bar{y})$ is the Hamming distance between $\bar{x}$ and $\bar{y}$ given by

PROC. WEST COAST CONFERENCE ON COMBINATORICS,
GRAPH THEORY, AND COMPUTING, pp. 25-40.

$$d(\bar{x},\bar{y}) = |\{k: x_k \neq y_k\}|.$$

The weight of a codeword $\bar{x}$, denoted by $w(\bar{x})$, is defined to be its distance from $\bar{0} = (0,0,\ldots,0)$ (which may not be in C).

Two basic quantities studied extensively in coding theory are:

$$A(n,d) \equiv \max\{|C|: \ C \text{ is a binary code of length } n \text{ and minimum distance } d\}$$

and

$$A(n,d,w) \equiv \max\{|C|: \ C \text{ is a binary code of length } n \text{ and minimum distance } d \text{ with all codewords of weight } w\}.$$

(For a fuller treatment of these topics the reader is referred to [11].)

Many upper bounds and some lower bounds for both $A(n,d)$ and $A(n,d,w)$ are available in the literature. For a survey of these the reader is referred to [1] and [5]. In Tables 1 and 2 we give some small values of these functions. Since $A(n-1,2\delta-1) = A(n,2\delta)$ we only list values of $A(n,d)$ for d even.

We should point out that the function $A(n,d,w)$ has been studied under another guise in extremal set theory by Erdös, Hanani, Kalbfleisch, Schönheim, and others (see [4]) in the following context. For given integers t, k, v, let $D(t,k,v)$ denote the maximum number of k-element subsets of

n \ d	4	6	8	10
6	4	2	1	1
7	8	2	1	1
8	16	2	2	1
9	20	4	2	1
10	40	6	2	2
11	72-79	12	2	2
12	144-158	24	4	2

A(n,d)
Table 1

n \ w	2	3	4	5	6	7
4	2	1	1	0	0	0
5	2	2	1	1	0	0
6	3	4	3	1	1	0
7	3	7	7	3	1	1
8	4	8	14	8	4	1
9	4	12	18	18	12	4
10	5	13	30	36	30	13
11	5	17	35	66	66	35
12	6	20	51	74-84	132	73-84

A(n,4,w)
Table 2

a v-element set S such that every t-element subset of S is contained in at most one of the k-element subsets. In fact, it is easy to see that

$$D(t,k,v) = A(v,2k-2t+2,k).$$

We also note for future use that if $w(\overline{x}) = w(\overline{y})$ then $d(\overline{x},\overline{y})$ must be even. Hence

$$A(n,2\delta-1,w) = A(n,2\delta,w).$$

Bounds on A(n,d,w)

While our primary concern will be with lower bounds on A(n,d,w) we mention here for purposes of comparison one of the best upper bounds known (due to S. M. Johnson [9], [10]). It is

$$A(n,2\delta,w) \leq \binom{n}{w-\delta+1} / \binom{w}{w-\delta+1}.$$

From this it follows that, for fixed δ and w,

$$(1) \qquad A(n,2\delta,w) \leq (1+o(1)) \frac{(\delta-1)!n^{w-\delta+1}}{w!} \text{ as } n \to \infty.$$

Of particular interest is the special case $\delta = 2$, when the upper bound becomes

$$(2) \qquad A(n,4,w) \leq \frac{1}{n-w+1} \binom{n}{w}.$$

The following three theorems were given in [5].

Theorem 1.

$$(3) \qquad A(n,4,w) \geq \frac{1}{n} \binom{n}{w}.$$

Proof: Let F_w^n denote the set of $\binom{n}{w}$ binary codewords of length n and weight w and let $\mathbf{Z}_n$ denote the integers modulo n. Consider the map $T:F_w^n \to \mathbb{Z}_n$ given by

(4) $$T(\bar{x}) = \sum_{x_i=1} i \pmod{n}$$

for $\bar{x} = (x_1,\ldots,x_n) \in F_w^n$. For $0 \leq i \leq n-1$, let C_i be the code $T^{-1}(i)$. Of course all codewords of C_i have weight w. We claim that the distance between any two distinct codewords of C_i is at least 4. For suppose not, i.e., suppose $\bar{x}, \bar{y} \in C_i$, $\bar{x} \neq \bar{y}$, with $d(\bar{x},\bar{y}) < 4$. Thus $d(\bar{x},\bar{y}) = 2$. This implies that $\bar{x}$ and $\bar{y}$ agree in all but two components, say the r-th and s-th components where $x_r = 1$, $y_r = 0$ and $x_s = 0$, $y_r = 1$. But

$$T(\bar{x}) = T(\bar{y}) = i \text{ so that}$$
$$T(\bar{x}) = a + r \equiv i \pmod{n},$$
$$T(\bar{y}) = a + s \equiv i \pmod{n}$$

for some $a \in \mathbb{Z}_n$. This is impossible since r and s are distinct integers between 1 and n.

Since

$$|C_0| + \ldots + |C_{n-1}| = \binom{n}{w}$$

for at least one j we have

$$|C_j| \geq \frac{1}{n}\binom{n}{w}$$

and the theorem is proved.

Note that this theorem is not completely constructive since we are unable to specify <u>which</u> j it is which has

$|C_j|$ large. A computer search of small cases indicates that any j is probably satisfactory asymptotically, i.e.,

$$|C_i|/|C_j| \to 1$$

for all i, j as $n \to \infty$.

The preceding proof is based on a method given by B. Bose and T. R. N. Rao in [3] in which they prove the slightly weaker bound

$$A(n,4,w) \geq \frac{1}{n+1}\binom{n}{w}.$$

The case of general δ is considered in the next result.

Theorem 2. Let $q \geq n$ be a prime power. Then

$$A(n,2\delta,w) \geq \frac{1}{q^{\delta-1}}\binom{n}{w}.$$

Proof: The proof has a similar structure to that of Theorem 1. Let us label the elements of GF(q) by $\omega_0,\omega_1,\ldots,\omega_{q-1}$. Define a map

$$T:F_n^w \to GF(q)^{\delta-1}$$

by

$$T(\bar{x}) = (T_1(\bar{x}),T_2(\bar{x}),\ldots,T_{\delta-1}(\bar{x}))$$

where

$$T_1(\bar{x}) = \sum_{x_i=1} \omega_i,$$

$$T_2(\bar{x}) = \sum_{\substack{i<j \\ x_i=x_j=1}} \omega_i\omega_j,$$

$$T_3(\bar{x}) = \sum_{\substack{i<j<k \\ x_i=x_j=x_k=1}} \omega_i\omega_j\omega_k$$

$$\vdots$$

for $\bar{x} = (x_1,\ldots,x_n)$. For each $(\delta-1)$-tuple $\bar{v} = (v_1,\ldots,v_{\delta-1}) \in GF(q)^{\delta-1}$ let

$$C_{\bar{v}} = T^{-1}(\bar{v}).$$

Thus, for some $\bar{v}$,

$$|C_{\bar{v}}| \geq \frac{1}{q^{\delta-1}} \binom{n}{w}.$$

We claim that $C_{\bar{v}}$ has distance 2δ. Suppose not, i.e., suppose there exist $\bar{x}$, $\bar{y} \in C_{\bar{v}}$, $\bar{x} \neq \bar{y}$, with $d(\bar{x},\bar{y}) = 2\gamma \leq 2\delta-2$. Thus there are 2δ distinct coordinates $r_1,\ldots,r_\gamma$, $s_1,\ldots,s_\gamma$ such that

$$x_{r_1} = \ldots = x_{r_\gamma} = y_{s_1} = \ldots = y_{s_\gamma} = 0,$$

$$x_{s_1} = \ldots = x_{s_\gamma} = y_{r_1} = \ldots = y_{r_\gamma} = 1$$

and $x_i = y_i$ for all other i. Since $T(\bar{x}) = T(\bar{y})$, the first δ elementary symmetric function σ_j, $0 \leq j \leq \delta-1$, of $\{\omega_{r_1},\ldots,\omega_{r_\gamma}\}$ and $\{\omega_{s_1},\ldots,\omega_{s_\gamma}\}$ agree. Thus the polynomial

$$x^\gamma - \sigma_1 x^{\gamma-1} + \sigma_2 x^{\gamma-2} - \ldots + (-1)^\gamma \sigma_\gamma$$

has all the ω_{r_i} and ω_{s_j} as roots. This is impossible since in any field a polynomial of degree m cannot have more than m roots. This proves the theorem.

Another Construction

Let us call an n-element subset $S \subseteq \mathbb{Z}_n$ an S_t-set of size n and modulus m if all the sums

$$s_{i_t} + s_{i_2} + \ldots + s_{i_t}$$

with $i_1 < i_2 < \ldots < i_t$ are distinct modulo m. These sets have been studied in the combinatorial literature (see [7]) and can also be used to obtain good lower bounds on $A(n,2\delta,w)$.

Theorem 3. If there exists an $S_{\delta-1}$-set of size n and modulus m then

$$A(n,2\delta,w) \geq \frac{1}{m}\binom{n}{w}.$$

The proof is similar to that of Theorem 2 but using the map

$$T:F_w^n \to \mathbb{Z}_m$$

given by

$$T(\bar{x}) = \sum_{x_i=1} s_i \pmod m.$$

As before, the codes are $C_i = T^{-1}(i)$, one of which must have as many codewords as the average $\frac{1}{m}\binom{n}{w}$.

From known results for S_t-sets it follows that if $q \geq n-1$ is a prime power and $\delta \geq 3$ then

$$A(n,2\delta,w) \geq \frac{q-1}{q^{\delta}-1}\binom{n}{w}. \tag{5}$$

Harmonious Graphs

Note that if S is an S_t-set of size n and modulus m then

$$m \geq \binom{n}{t}. \tag{6}$$

For the remainder of the paper, we restrict ourselves to the case t = 2. Equation (6) then becomes

$$m \geq \binom{n}{2}. \tag{6'}$$

Equality can be achieved in (6') for small n by the following examples.

$S = \{0,1\}$ for $n = 2$, $m = 1$,

$S = \{0,1,2\}$ for $n = 3$, $m = 3$,

$S = \{0,1,2,4\}$ for $n = 4$, $m = 6$.

However these are the only values of n for which equality can occur.

We can translate this situation into the following equivalent form. S is an S_2-set of size n and modulus $\binom{n}{2}$ iff it is possible to label the vertices of K_n, the complete graph on n vertices, with the elements of S so that if each edge of K_n is assigned the sum modulo $\binom{n}{2}$ of the two values assigned to its endpoints, then all edge values are distinct (and so represent a complete residue system modulo $\binom{n}{2}$). In Figure 1 we show the labelled complete graphs corresponding to the three extremal sets S given above.

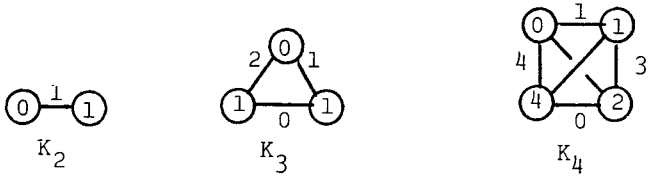


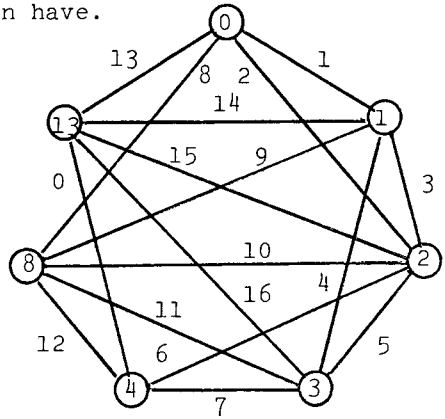
Figure 1

This interpretation prompts the following definition (see [6] for further information):

Definition. A graph G with e edges is called harmonious if it is possible to label the vertices of G with distinct

values from $\mathbf{Z}_e$ so that every element of $\mathbb{Z}_e$ occurs uniquely as an edge sum of G.

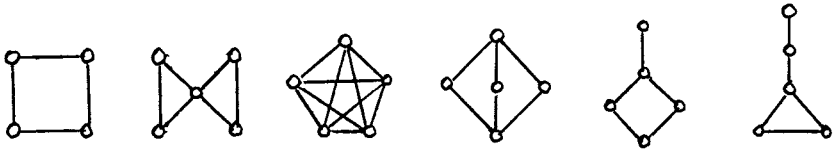
For example, we show in Figure 2 a harmonious graph with 7 vertices and 17 edges. It turns out (see [6]) that this is the maximum number of edges a harmonious graph on 7 vertices can have.



A Harmonious Graph with 7 Nodes and 17 Edges

Figure 2

In Figure 3 we give the connected graphs on at most 5 vertices which are <u>not</u> harmonious.

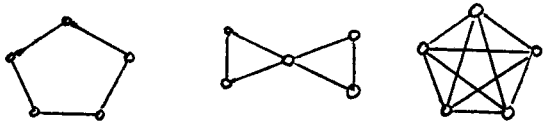


Nonharmonious Graphs.

Figure 3

A curious geometrical interpretation can be given to the condition that a graph G be harmonious. Let P_e denote a fixed regular e-gon embedded in the plane. Then G is harmonious iff the vertices of G can be embedded into the vertices of P_e so that no two edges of the embedded copy of G are parallel. This follows from the observation that if the vertices of P_e are labelled cyclically by 0,1,...,e-1, then the direction of the chord joining i and j depends only on i + j (mod e).

A related concept which has appeared frequently in the graph theory literature is that of a graceful graph (see [2]). A graph G with e edges is said to be graceful if it is possible to assign distinct values from {0,1,...,e} to the vertices of G so that the absolute values of the edge differences are all distinct (and therefore all values in {1,2,...,e} occur uniquely). In Figure 4 we list the connected graphs on 5 vertices which are not graceful.



Nongraceful Graphs

Figure 4

While it can be observed that Figures 3 and 4 contain two common graphs, in general the concepts of being

graceful and being harmonious are rather independent. For example, cycles of length n have the following properties:

n (mod 4)	harmonious	graceful
0	no	yes
1	yes	no
2	no	no
3	yes	yes

Similarly, complete bipartite graphs, which are known to be graceful, are never harmonious. This result has a remarkably short proof.

Theorem 4. $K_{r,s}$ is not harmonious.

Proof: Suppose a harmonious labelling of $K_{r,s}$ exists. This is equivalent to a direct sum decomposition of $\mathbb{Z}_{rs} = A \oplus B$ where A and B are disjoint subsets of $\mathbb{Z}_{rs}$ with $|A| = r$, $|B| = s$. Since all $a + b$ (modulo rs), $a \in A$, $B \in B$, are distinct then so are all differences $a - b$ (modulo rs). But there are $|A||B| = rs$ differences. Hence $0 = a - b$ must occur exactly once and therefore A and B are not disjoint. □

We extract an interesting corollary from the proof.

Corollary. If $\mathbb{Z}_n = A \oplus B$ then $|A \cap B| = 1$.

In fact most graphs are neither harmonious nor graceful. More precisely, it can be shown using the probability method (see [6]) that the fraction of all graphs on n vertices which are harmonious (or graceful) tends to 0 exponentially with n.

Let us define H(n) to be the maximum number of edges a harmonious graph on n vertices can have (with G(n)

defined similarly for graceful graphs). In Table 3 we list some of the known values.

n	H(n)	G(n)
2	1	1
3	3	3
4	6	6
5	9	9
6	13	13
7	17	17
8	24	23
9	30	29
10	36	36

Table 3

Asymptotically it can be shown [6] that

$$\frac{5}{18} n^2 \leq H(n) \leq \frac{1}{2} n^2.$$

It is especially annoying that we cannot prove that $H(n) < \left(\frac{1}{2} - \varepsilon\right)n^2$ for some $\varepsilon > 0$. The lower bound depends on recent results of Hämmerer and Hofmeister [7] who showed that it is possible to select n nonnegative integers $a_1 < a_2 < \ldots < a_n$ such that all integers up to $\frac{5}{18} n^2$ can be represented as $a_i + a_j$.

Some Questions

(1) A well known conjecture of Ringel and Kotzig asserts that all trees (= acyclic connected graphs) are graceful. We make the corresponding conjecture that all trees are harmonious, where we have to modify the definition slightly so as to allow one vertex label to be repeated. This is true for all trees with at most nine vertices.

(2) Is there a polynomial-time algorithm to determine if G is harmonious? We conjecture that there is not.

(3) Is $H(n) \sim cn^2$? Is $c < 1/2$? We think that the answer to both questions is in the affirmative.

(4) How large must m be for an S_t-set of size n and modulus m to exist? For $t = 2$, it is known that the answer is $(1+o(1))n^2$. We conjecture that $m \geq (1+o(1))n^t$.

(5) What is the value of $A(n,d,w)$? From our results it follows that $(1+o(1)) \frac{n^{w-\delta+1}}{w!} \leq A(n,2\delta,w) \leq (1+o(1)) \frac{(\delta-1)!n^{w-\delta+1}}{w!}$. The upper bound is known to be correct if $w = 4$ and $\delta = 2$ or 3. Is it always correct?

REFERENCES

[1] M. R. Best, A. E. Brouwer, F. J. MacWilliams, A. M. Odlyzko, and N. J. A. Sloane, "Bounds for binary codes of length less than 25," IEEE Trans. Inf. Th. IT-24 (1978), 81-93.

[2] G. S. Bloom, "A chronology of the Ringel-Kotzig conjecture and the continuing quest to call all trees graceful," Annals of the N.Y. Acad. Sci. 326 (1979) 32-51.

[3] D. Bose and T. R. N. Rao, "On the theory of unidirectional error correcting/detecting codes," Technical Report CS-7817, Dep. of Computer Science,.Southern Methodist University, Dallas, Texas, September 1978.

[4] P. Erdös and H. Hanani, "On a limit theorem in combinatorical analysis," Pub. Math. Debrecen, 10 (1963), 10-13.

[5] R. L. Graham and N. J. A. Sloane, "Lower bounds for constant weight codes," (to appear in IEEE Trans. Inf. Th.).

[6] R. L. Graham and N. J. A. Sloane, "On additive bases and harmonious graphs" (to appear).

[7] H. Halberstam and K. F. Roth, Sequences, Vol. 1, Oxford University Press, Oxford, 1966.

[8] N. Hämmerer and G. Hofmeister, "Zu einer Vermutung von Rohrbach," J. reine angew. Math. 286/287 (1976), 239-247.

[9] S. M. Johnson, "A new upper bound for error-correcting codes," IRE Trans. Inf. Th., IT-8 (1962), 203-207.

[10] S. M. Johnson, "Upper bounds for constant weight error-correcting codes," Discrete Math. 3 (1972), 109-124.

[11] F. J. MacWilliams and N. J. A. Sloane, The Theory of Error-Correcting Codes, North Holland, Amsterdam, 1977.