Discrete Applied Mathematics 161 (2013) 424–429

Contents lists available at SciVerse ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Note

The maximum number of minimal codewords in long codes

A. Alahmadi[a], R.E.L. Aldred[b], R. dela Cruz[c,d], P. Solé[e,a,*], C. Thomassen[f,a]

[a] *Department of Mathematics, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia*
[b] *Department of Mathematics and Statistics, University of Otago, P.O. Box 56, Dunedin, New Zealand*
[c] *Division of Mathematical Sciences, SPMS, Nanyang Technological University, Singapore*
[d] *Institute of Mathematics, University of the Philippines Diliman, Quezon City, Philippines*
[e] *Telecom ParisTech, 46 rue Barrault, 75634 Paris Cedex 13, France*
[f] *Department of Mathematics, Technical University of Denmark, DK-2800 Lyngby, Denmark*

ARTICLE INFO

Article history:
Received 2 April 2012
Received in revised form 24 August 2012
Accepted 18 September 2012
Available online 12 October 2012

Keywords:
Minimal codewords
Intersecting codes
Cycle code of graphs

ABSTRACT

Upper bounds on the maximum number of minimal codewords in a binary code follow from the theory of matroids. Random coding provides lower bounds. In this paper, we compare these bounds with analogous bounds for the cycle code of graphs. This problem (in the graphic case) was considered in 1981 by Entringer and Slater who asked if a connected graph with p vertices and q edges can have only slightly more than 2^{q-p} cycles. The bounds in this note answer this in the affirmative for all graphs except possibly some that have fewer than $2p + 3\log_2(3p)$ edges. We also conclude that an Eulerian (even and connected) graph has at most 2^{q-p} cycles unless the graph is a subdivision of a 4-regular graph that is the edge-disjoint union of two Hamiltonian cycles, in which case it may have as many as $2^{q-p} + p$ cycles.

1. Introduction

A binary linear code of length n and dimension k is a k-dimensional subspace of GF_2^n. The codewords of a binary code form a poset under support inclusion. The non-zero codewords that dominate only the zero element in that partial order are called *minimal*. They include the minimum weight non-zero codewords, but do not coincide, in general, with them. They occur in decoding studies [1,4], and independently in secret sharing schemes based on codes [3]. What is the maximum number $M(C)$ of minimal codewords a binary linear code C might have? If C has dimension k, an immediate upper bound, which we call the *trivial upper bound*, is $M(C) \leq 2^k - 1$. This bound is met with equality for *intersecting codes*, i.e. codes any pair of codewords of which intersect nontrivially [5,6]. Conversely, any code meeting that bound with equality is intersecting.

If G is a connected graph on p vertices with q edges, then its cycle code $C(G)$ has length $n = q$, and dimension $k = q-p+1$. The minimal codewords of $C(G)$ are exactly the incidence vectors of cycles (that is, circuits in the cycle matroid) in G. Thus the above question restricted to the graphic case asks how many cycles a graph with p vertices and q edges can have. (We allow graphs to have multiple edges but no loops.) This question was raised in 1981 by Entringer and Slater [8] who observed that a graph with p vertices and q edges cannot have 2^{q-p+1} cycles because of the trivial upper bound. They gave examples showing that it may have slightly more than 2^{q-p} cycles and asked if these examples were close to best possible. In this paper, we verify this for all graphs except possibly some that are sparse.

[*] Corresponding author at: Telecom ParisTech, 46 rue Barrault, 75634 Paris Cedex 13, France. Fax: +33 01 45 81 71 58.
E-mail addresses: adelnife2@yahoo.com (A. Alahmadi), raldred@maths.otago.ac.nz (R.E.L. Aldred), ROMA0001@e.ntu.edu.sg (R. dela Cruz), sole@enst.fr (P. Solé), C.Thomassen@mat.dt.u.dk (C. Thomassen).

0166-218X/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2012.09.009

We view this in a coding theoretic context as follows. We denote by $\mu(R)$ the asymptotic exponent of the maximum of $M(C)$ for families of codes C of rate R, where R is the limsup of k/n. Formally, let $\mathbf{C}[n, k]$ denote the set of all $[n, k]$ codes, and let

$$M(n, k) = \max\{M(C) : C \in \mathbf{C}[n, k]\}.$$

We now introduce

$$\mu(R) = \limsup_{n\to\infty} \frac{1}{n} \log_2 M(n, \lceil Rn \rceil).$$

If, in this definition, we replace $\mathbf{C}[n, k]$ by the cycle codes of graphs with p vertices and q edges (where $n = q$, and $k = q - p + 1$) we obtain $M_g(n, k)$, $\mu_g(R)$, respectively.

By the trivial upper bound, $\mu(R) \leq R$. For $R \in [0, 0.5]$ random coding shows that the bound is tight. For $R > 0.5$ the trivial upper bound can be improved using matroid theory for $R \geq R_0$ with $R_0 \approx 0.77$.

In this paper, we prove an upper bound for the number of cycles in a graph with p vertices and q edges. That upper bound implies that

$$\mu_g(R) = R$$

for $R \leq 0.5$, while for $0.5 < R < 1$, $\mu_g(R)$ is continuous, piecewise linear, and

$$\mu_g(R) = -(1 - R)\log_2(1 - R)$$

for infinitely many values of R.

It also implies that every graph with more than $2p + O(\log(p))$ edges has fewer than 2^{q-p} cycles. The graph $2C_p$ (the cycle of length p with every edge doubled) is 4-regular, and has therefore $2p$ edges, and has precisely $2^{q-p} + p = 2^p + p$ cycles. This shows that we cannot omit the $O(\log(p))$ term above.

2. Known bounds on $M(n, k)$, $M(C)$ and $\mu(R)$

In this section, we review some known bounds.

If C is an $[n, k]$ code, then we have $M(C) \leq 2^k - 1$. We call this bound the *trivial upper bound.* It is easy to see that a binary linear code C meets the trivial bound with equality if and only if it is intersecting. For, if C is not intersecting, then two of its codewords, say c and d have disjoint supports. Their sum $c + d$ is nonzero and non-minimal. Hence $M(C) < 2^k - 1$. Conversely, if C has a non-minimal codeword, then it can be written as a sum of at least two disjoint support minimal codewords. Therefore C is not intersecting.

Recall that the trivial upper bound implies that $\mu(R) \leq R$.

It is proved in [7] that, for a binary matroid on n points of rank $n - k$ represented by an $[n, k]$ code C say, we have

$$M(C) \leq \binom{n}{k-1}.$$

This implies that $\mu(R) \leq H(R)$, where H is Shannon's binary entropy function defined for $x \in [0, 1]$ by

$$H(x) = -x\log_2 x - (1 - x)\log_2(1 - x).$$

That upper bound is better than the trivial upper bound for $R > R_0 \approx 0.77$ where $H(R_0) = R_0$.

By averaging arguments (random coding) it was shown in [3, Cor. 2.5] that for $R < 1/2$, we have

$$\mu(R) \geq R,$$

and for $R > 1/2$, we have

$$\mu(R) \geq H(R) - 1 + R.$$

Finally in this section we state, as Theorem 1, a bound for $M(n, k)$ from [1, Theorem 5]. We refer to this bound as the *Agrell upper bound.*

Theorem 1. *For $\frac{k-1}{n} > \frac{1}{2}$ we have*

$$M(n, k) \leq \frac{2^k}{4n(\frac{k-1}{n} - \frac{1}{2})^2}. \quad \Box$$

3. Cycle codes of graphs

It is a long-standing and difficult problem in graph theory to find the maximum number of cycles a connected graph on p vertices and with q edges can have. This problem was raised by Entringer and Slater [8] who observed that no connected graph G can have 2^{q-p+1} cycles. This follows from the trivial upper bound because there is a binary $[q, q-p+1]$ code $C(G)$ called the *cycle code* of the graph. Its codewords are defined on the edge set and are the indicator vectors of the edge disjoint unions of cycles. The minimal codewords of $C(G)$ are the indicator vectors of the cycles of G. Entringer and Slater [8] also observed that there are graphs having slightly more than 2^{q-p} cycles and asked if this is (essentially) the maximum. Finally, Entringer and Slater [8] observed that the maximum is attained for cubic graphs.

The first bound significantly below the trivial bound was obtained by Aldred and Thomassen [2] who proved that no connected graph G can have more than $\frac{15}{16}2^{q-p+1}$ cycles. This is the best known upper bound for cubic graphs. But, for graphs of average degree > 4 there are better bounds, and in fact, the question by Entringer and Slater has been answered for all graphs of average degree slightly greater than 4. The Agrell bound (Theorem 1) immediately implies the following.

Corollary 1. *If G is a connected graph on p vertices and with q edges satisfying $q > 2p$, then its number of cycles is at most*

$$\frac{q2^{q-p+1}}{(q-2p)^2}. \quad \square$$

We shall here prove the following.

Theorem 2. *If G is a connected graph on p vertices and with q edges, and we write*

$$q-1 = (p-1)m + r$$

where m, r are nonnegative integers, and $0 \le r < p-1$,
then its number of cycles is at most

$$qm^{p-1-r}(m+1)^r. \quad \square$$

If we fix the ratio q/p and let p tend to infinity, then the bounds in Corollary 1 and Theorem 2 are essentially exponential functions. The exponential function in Theorem 2 is in a sense best possible, as we point out below. Also, both results answer the question of Entringer and Slater (asking if a graph can have significantly more than 2^{q-p} cycles) for graphs of average degree slightly more than 4. Corollary 1 shows that counterexamples (if any) can have at most $2p+O(\sqrt{p})$ edges. Theorem 2 goes further and says that they must have at most $2p+O(\log(p))$ edges. It would be interesting to answer the question for all graphs of average degree at least 4.

Theorem 2 is a consequence of Theorem 3 below.

A *path* is a graph with vertices $v_1, v_2, \dots, v_p$ and edges $v_1v_2, v_2v_3, \dots, v_{p-1}v_p$. A *multipath* is obtained from a path by replacing some edges with multiple edges. Let $f(q,p)$ denote the maximal number of paths from v_1 to v_p in a multipath with p vertices and q edges. This maximum is attained if no two edge multiplicities differ by more than 1. So, if $p-1$ divides q, then

$$f(q,p) = (q/(p-1))^{p-1}.$$

Lemma 1. *If x, y are vertices in a graph G with p vertices and q edges, then G has at most $f(q,p)$ paths from x to y.*

Proof. The proof is by induction on p. If $p=2$ the statement is trivial. So assume $p>2$. Let d denote the degree of x, and let $x_1, x_2, \dots, x_d$ be the neighbors of x. (Some of these neighbors may be identical.) By induction, $G-x$ has at most $f(q-d, p-1)$ paths from x_i to y, for each $i = 1, 2, \dots, d$. So G has at most $df(q-d, p-1)$ paths from x to y. As $f(q-d, p-1)$ is the number of paths between the ends in a multipath with $p-1$ vertices and $q-d$ edges, we may interpret $df(q-d, p-1)$ as the number of paths between the ends in a multipath with p vertices and q edges, where the first edge multiplicity in the multipath is d. By the maximum property of $f(q,p)$, we have

$$df(q-d, p-1) \le f(q,p),$$

which completes the proof. $\square$

Theorem 3. *Let p, q be natural numbers ≥ 2.*
There exists a graph with p vertices, q edges, and at least $f(q-1,p)$ cycles.
If G is any graph with p vertices and q edges, then G has at most

$$qf(q-1,p)$$

cycles.

Proof. By the definition of f, there exists a multipath with p vertices, $q-1$ edges, and precisely $f(q-1,p)$ paths between the ends. If we add an edge between the ends we get a graph with p vertices, q edges, and at least $f(q-1,p)$ cycles.

To prove the last statement, consider any edge $e = xy$ in G. The number of cycles in G containing e is the number of paths in $G - e$ from x to y. By Lemma 1, this number is at most

$$f(q-1, p).$$

This completes the proof. □

Because of the logarithm in the definition of μ, Theorem 3 gives the right μ-value for the graphs of a fixed average degree. It is still interesting, though, to decide if the bound $qf(q-1,p)$ can be lowered to about $(q/p)^p$. Does the maximum number of cycles occur in graphs similar to tC_p? Are there graphs of average degree $2t$ without multiple edges that have the same number or a larger number of cycles? These questions are open even for cubic graphs. As mentioned earlier, Entringer and Slater [8] observed that a cubic graph on p vertices may have as many as $2^{p/2}$ cycles. It has been open for several years if this is close to the right number. For planar cubic graphs this was verified by Aldred and Thomassen [2]. For general cubic graphs they lowered the trivial upper bound $2^{p/2+1}$ to $(15/16) \cdot 2^{p/2+1}$.

We can now answer the question by Entringer and Slater [8] for all graphs with average degree slightly more than 4.

Theorem 4. *Let G be a graph with p vertices and q edges. If $q > 2p + 3\log_2(3p)$, then G has at most 2^{q-p} cycles.*

Proof. Consider first the case where $q = 2p$. The upper bound on the number of cycles provided by Theorem 2 is $(2p)\cdot 3 \cdot 2^{p-2}$, which is more than $2^{q-p} = 2^p$. However, if we increase q to $2p+r$, then the upper bound increases to $(2p+r)\cdot 3\cdot 2^{p-2}\cdot(3/2)^r$. This number is $\leq 2^{p+r}$ for $r > 3\log_2(3p)$.

This completes the proof. □

4. Applications to $\mu_g(R)$

We can now determine $\mu_g(R)$ completely.

Theorem 5. *For $0 < R \leq 0.5$,*

$$\mu_g(R) = R.$$

If R is of the form $1 - 1/t$ where t is a natural number ≥ 2, then

$$\mu_g(R) = -(1-R)\log_2(1-R).$$

The function $\mu_g(R)$ is continuous and linear in each closed interval from $1 - 1/t$ to $1 - 1/(t+1)$, where t is a natural number ≥ 2.

Proof. Assume first that $0 < R < 0.5$. For any two natural numbers p, r we let $C_{p,r}$ be obtained from $2C_p$ by subdividing one edge r times. Then $C_{p,r}$ has $p + r$ vertices and $n = 2p + r$ edges. Thus the dimension of the cycle code is $k = p + 1$, and the rate of the cycle code is $k/n = (p+1)/(2p+r)$. For each natural number p we let r be the largest natural number such that $k/n = (p+1)/(2p+r) \geq R$. Then $k = \lceil Rn \rceil$. Also, $r = \lfloor (p+1)/R - 2p \rfloor$. Recall that the number of cycles in $2C_p$ and hence also in $C_{p,r}$ is $> 2^p$. Substituting these values in the definition of μ_g and letting p tends to infinity, we conclude that $\mu_g(R) \geq R$.

The trivial upper bound shows that this inequality is, in fact, an equality.

Consider next the case where $R = 1 - 1/t$ for where t is the natural number ≥ 2. Let tC_p be the cycle of length p where each edge has been duplicated t times. This graph has p vertices and $q = pt$ edges. The number of cycles in this graph is

$$t^p + p\binom{t}{2},$$

the first term counting cycles of length p and the second cycles of length 2. The graph is regular of degree $2t$. Hence it has $n = pt$ edges and has rate $(pt - p + 1)/pt = 1 - 1/t + 1/pt$. Hence its cycle code is an $[n, Rn + 1]$-code. If we delete an edge, then we get an $[n-1, \lceil R(n-1) \rceil]$-code. Deleting an edge reduces the number of cycles only slightly. Letting p and hence also n tend to infinity, we conclude that

$$\mu_g(R) \geq -(1-R)\log_2(1-R).$$

Now assume that $1 - 1/t < R < 1 - 1/(t+1)$, where t is a (fixed) natural number ≥ 2. Then we let $G(p, R, r)$ denote the graph obtained from tC_p by adding r edges between neighboring vertices such that all edge multiplicities are t or $t+1$. The resulting graph $G(p, R, r)$ has p vertices and $n = pt + r$ edges. Thus the dimension of the cycle code is $k = p(t-1) + r + 1$, and the rate of the cycle code is $k/n = (p(t-1) + r + 1)/(pt + r)$.

We first choose any p so large that the rate of tC_p, which is $(pt - p + 1)/pt = 1 - 1/t + 1/pt$, is smaller than R. Then we let r be the largest natural number such that the rate of $G(p, R, r)$, which is $(p(t-1) + r + 1)/(pt + r)$, is smaller than or equal to R. That is, $r = -pt + \lfloor (p-1)/(1-R) \rfloor$. Then the cycle code of $G(p, R, r)$ is an $[n, \lceil Rn \rceil]$-code, where $n = pt + r = \lfloor (p-1)/(1-R) \rfloor$ is the number of edges of $G(p, R, r)$.

The number of cycles in $G(p, R, r)$ is $> t^{p-r}(t+1)^r$.

If we substitute these values in the definition of μ_g and let p tend to infinity, then we conclude that

$$\mu_g(R) \geq ((1-R)(1+t)-1)\log(t) + (-t(1-R)+1)\log(t+1).$$

The right hand side is clearly a linear function. For R equal to $1-1/t$ or $1-1/(t+1)$ the right hand side has the same values as the lower bounds we obtained for those two values of R. So we have obtained a lower bound for $\mu_g(R)$ which is continuous and piecewise linear.

We claim that this lower bound is also an upper bound. We used the graph $G(p, R, r)$ above. If we put $q = n = pt + r$, then $G(p, R, r)$ has the maximum number of cycles among those graphs with q edges which are obtained from a cycle of length p by duplicating edges. The graphs used to give the lower bound $f(q-1, p)$ in Theorem 3 are also graphs of this type. Hence $G(p, R, r)$ has at least $f(q-1, p)$ cycles. On the other hand, Theorem 3 says that any graph with p vertices and q edges has at most $qf(q-1, p)$ cycles. So no graph with p vertices and q edges has more than q times as many cycles as $G(p, R, r)$. Hence the lower bound for $\mu_g(R)$ obtained from the graphs $G(p, R, r)$ is also an upper bound. □

The function $\mu_g(R)$ is less than the matroid upper bound for all R and also less than the random upper bound for $R > 0.5$. The Agrell upper bound gives the same upper bound on μ as the trivial upper bound. Fig. 1 shows these bounds.

5. Cycle codes of 4-regular graphs

Cycle codes of cubic (that is, 3-regular) graphs have enjoyed particular attention because in order to answer the aforementioned question by Entringer and Slater [8], it suffices to consider cubic graphs. The investigations in this paper indicate that the 4-regular graphs also deserve attention. First of all, their cycle codes have rate $R \approx 0.5$, and this is the smallest value of R for which the function $\mu_g(R)$ changes shape. Here we shall provide another reason. Although it is merely an observation we call it a theorem because of the striking exceptions that appear in the statement. A graph is *Eulerian* if it is connected and all vertices have even degree. These graphs are particularly interesting in the present context because the vector **1** consisting of ones is a code word. Recall that a *Hamiltonian cycle* in a graph is a cycle containing all vertices.

Theorem 6. *If G is an Eulerian graph with p vertices and q edges, then G has at most* 2^{q-p} *cycles unless G is a subdivision of a 4-regular graph which is a union of two Hamiltonian cycles.*

Proof. The map sending a codeword $\mathbf{x}$ into the codeword $\mathbf{x}+\mathbf{1}$ is a map from the cycle code into itself. If the codeword of every cycle is mapped into a codeword which does not correspond to a cycle, then at most half of the 2^{q-p+1} codewords correspond to cycles, and the result follows. So we may assume that G has a cycle C such that both its codeword $\mathbf{x}$ and also the codeword $\mathbf{x}+\mathbf{1}$ correspond to cycles, say C, C'. Hence G cannot have a vertex of degree 6 or more. Each of C, C' contains precisely two edges incident with each vertex of degree 4. And each edge of G belongs to precisely one of C, C'. So C, C' is a partition of G into two Hamiltonian cycles after 'suppressing' the vertices of degree 2 (i.e. for each vertex v of degree 2 and neighbors u, w, delete the vertex v and add the edge uw). □

6. Conclusion and open problems

The function $\mu_g(R)$ is known exactly. The function μ seems more problematic. Thus $\mu(R) = R$ for $R \leq 0.5$, while for $R > 0.5$ we only have the bounds

$$H(R) - 1 + R \leq \mu(R) \leq \min(R, H(R))$$

the best known upper bound for $R > 0.5$. These bounds can be generalized to linear codes over a non-binary alphabet.

By Cohen and Lempel [5] long linear intersecting codes can exist only for $R < 0.283$. Thus the codes of rate between that value and 0.5 provided by random coding are "almost" intersecting.

Open problem 1. *Is $\mu(R)$ a continuous function of R? Is it concave? For which R is it maximum?*

Analogous questions have been considered by Manin [9].

We have answered Open problem 1 for $\mu_g(R)$. This function has maximum for $R = 2/3$. The 6-regular graphs have rate $\approx 2/3$, so maybe also the 6-regular graphs are worth studying in more detail.

As mentioned earlier, the 4-regular graph $2C_p$ has $2^p + p$ cycles, and no 4-regular graph has more than 2^{p+1} cycles, by the trivial upper bound.

Open problem 2. *Does there exist a 4-regular graph with p vertices and more than $2^p + p$ cycles?*

In case the answer is negative, then it seems that the 4-regular graphs are the only regular graphs for which there is a simple expression for the maximum number of cycles.

Open problem 3. *Does there exist a real number $c < 2$ such that every 4-regular graphs with p vertices and no multiple edges has less than c^{p+1} cycles?*

As mentioned above, tC_p has p vertices, $q = pt$ edges, and $t^p + p\binom{t}{2}$ cycles.

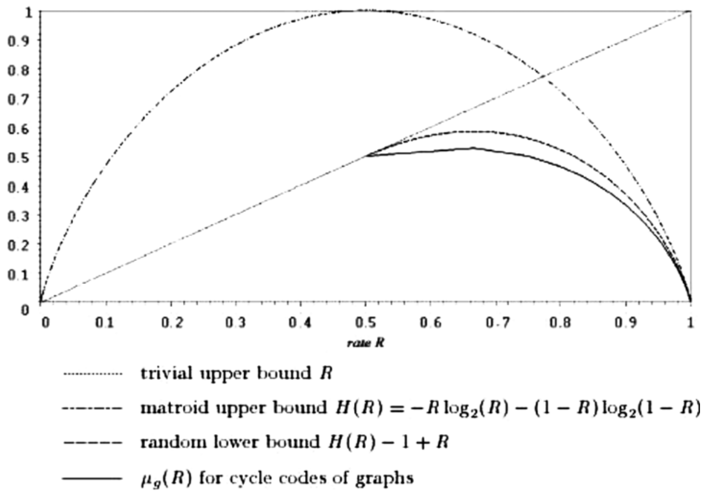

Fig. 1. $\mu(R)$ versus *R*.

Open problem 4. *Does there exist a graph with p vertices, $q = pt$ edges (where t is a natural number), and more than $t^p + p\binom{t}{2}$ cycles?*

Acknowledgment

The authors are indebted to G. D. Cohen and H. Randriam for helpful discussions.

References

[1] E. Agrell, On the Voronoi neighbor ratio for binary linear codes, IEEE Trans. Inform. Theory (1998) 3064–3072.
[2] R.E.L. Aldred, C. Thomassen, On the maximum number of cycles in a planar graph, J. Graph Theory 53 (2008) 255–264.
[3] A. Ashikmin, A. Barg, Minimal vectors in linear codes, IEEE Trans. Inform. Theory (1998) 2010–2017.
[4] A. Ashikmin, A. Barg, G. Cohen, L. Huguet, Variations on Minimimal Codewords in Linear Codes, in: LNCS, vol. 948, Springer, 1995, pp. 96–105.
[5] G. Cohen, A. Lempel, Linear intersecting codes, Discrete Math. 56 (1985) 35–43.
[6] G. Cohen, G. Zemor, Intersecting codes and independent families, IEEE Trans. Inform. Theory (1994) 1872–1881.
[7] Gy. Dosa, I. Szalkai, C. Laflamme, The maximum and minimum number of circuits and bases of matroids, PU. M. A. 15 (2004) 383–392.
[8] R.C. Entringer, P.J. Slater, On the maximum number of cycles in a graph, Ars Combin. 11 (1981) 289–294.
[9] Y.I. Manin, What is the maximum number of points on a curve over F2? J. Fac. Sci. Univ. Tokyo 1A 28 (1981) 715–720.