The BCH Family of Storage Codes on Triangle-Free Graphs and Its Relation to $R(3, t)$

Qing Xiang

joint work with Haihua Deng, Hexiang Huang, Guobiao Weng

Department of Mathematics, SUSTech

July 10, 2024

Outline

Storage Codes on Graphs

Storage codes on graphs were independently introduced by Mazumdar [Maz14; Maz15] and K. Shanmugam and A. G. Dimakis [SD14]

Definition 1.

Let Γ be a simple connected graph on n vertices and C a code of length n whose coordinates are indexed by the vertices of Γ. We call C a *storage code* on Γ if, for any codeword $c \in C$, one can recover the information at each coordinate of c by accessing its neighbors in Γ.

Example

The linear code $C_n = \{(c_1, \ldots, c_n) | \sum_{i=1}^n c_i = 0\}$ is a storage code on the complete graph K_n.

Construction of Storage Codes on Graphs

Given a graph Γ on n vertices, we can construct a linear code C on it using $H := A(\Gamma) + I$ as its parity-check matrix, i.e.,

$$C = \{c = (c_1, \ldots, c_n) | Hc^\top = 0\}.$$

Then C is a storage code on Γ since for a codeword $c = (c_1, \ldots, c_n) \in C$, the recovery of any i-th entry of c is feasible through its neighbors, as the i-th row of H suggests a linear equation:

$$c_i = -\sum_{j \in N(i)} c_j,$$

where $N(i)$ denotes the set of neighbors of i in Γ.

Storage Codes on Triangle-Free Graphs

The rate of a linear storage code C, denoted by $R(C)$, is the ratio of its dimension to the dimension of the ambient vector space. For a family of storage codes $\{C_n\}$, where n is the length of C_n, the *rate* of this family is defined as $\lim_{n\to\infty} R(C_n)$, assuming this limit exists.

We would like to construct high-rate storage codes since it represents high probability of all players to correctly guess the colour of their own hat in the model of guessing game.

In the previous example, the graph K_n used to construct the storage code $C_n = \{(c_1, \ldots, c_n) | \sum_{i=1}^{n} c_i = 0\}$ is dense, which prompts the question of the maximum achievable rate of storage codes on graphs without cliques K_t $(t \geq 3)$, i.e., triangle-free graphs.

The Rates of Storage Codes on Triangle-Free Graphs

A triangle-free yet edge-rich graph does not necessarily yield a high-rate storage code. Consider the complete bipartite graph $K_{t,t}$ as an example: It is triangle-free and dense, but any storage code C on it would have $R(C) \leq 1/2$ due to its two independent vertex sets.

An initial conjecture by D. Christofides and K. Markström [CM11] suggested a maximum rate of $1/2$ for triangle-free graphs. Subsequently, P. Cameron, A. Dang, and S. Riis [CDR14] disproved this conjecture by some sporadic examples.

In 2022, A. Barg and G. Zémor [BZ22] introduced four infinite families of storage codes on triangle-free graphs using Cayley graphs.

Storage Codes on Triangle-Free Cayley Graphs

Let $0 \notin S$ be a subset of $\mathbb{F}_2^r$. If the sum of any three distinct vectors in S is nonzero, then the Cayley graph $\Gamma = \mathrm{Cay}(\mathbb{F}_2^r, S)$ is triangle-free. Let C be a binary linear code defined by using $H := A(\Gamma) + I$ as its parity-check matrix. Then C is a storage code on triangle-free graph Γ.

A. Barg and G. Zémor [BZ22] asked whether the rates of storage codes on triangle-free graphs can be arbitrarily close to 1 and the answer is yes.

Even eariler, A. Golovnev and I. Haviv [GH20] introduced a family of storage codes on the generalized Kneser graphs (which are triangle-free) attaining unit rate, albeit using different terminology.

Subsequently, A. Barg, M. Schwartz and L. Yohananov [BSY22], and H. Huang and Q. Xiang [HX23] independently generalized the Hamming family presented in [BZ22].

The BCH family of Storage Codes on Triangle-Free Graphs

In [BZ22], A. Barg and G. Zémor constructed a family of storage codes (BCH family) and posed a question whether it can approach unit rate.

Let $q = 2^m$ with m being an positive integer. Define the vertex set as $G = \mathbb{F}_q^2$ and the connection set as $S_m \backslash \{0\}$, where

$$S_m := \{(a, a^3) | a \in \mathbb{F}_q\} \subseteq G.$$

The resulting graph is $\Gamma = \mathrm{Cay}(\mathbb{F}_q^2, S_m \backslash \{0\})$. We then define $H_m := A(\Gamma) + I$ and construct the binary linear code C_n (with $n = q^2 = 4^m$) using H_m as its parity-check matrix.

The choice of the connection set $S_m \backslash \{0\}$ for Γ, which coincides with the column set of the parity-check matrix for the 2-error-correcting BCH code, underpins our naming of $\{C_n\}$ as the BCH family.

Outline

Polynomial Method

We will apply the *polynomial method* to investigate the intrinsic algebraic structure of the BCH family, which leads us to establish an upper bound on the rank of the parity-check matrix H_m of the BCH family.

We can express the (x, y)-entry of H_m as the value of a polynomial evaluated at (x, y). First, the matrix H_m over $\mathbb{F}_2$ can be formulated as

$$H_m = (a_{x,y})_{x,y \in \mathbb{F}_q^2},$$

where the (x, y)-entry is given by

$$a_{x,y} = \begin{cases} 1, & \text{if } x - y \in S_m, \\ 0, & \text{otherwise.} \end{cases}$$

Polynomial Method

Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Then each entry $a_{x,y}$ can be expressed as the value of a polynomial g evaluated at (x, y):

$$\begin{aligned} a_{x,y} &= ((x_1 - y_1)^3 - (x_2 - y_2))^{q-1} + 1 \\ &= (x_1^3 + x_1 y_1^2 + x_1^2 y_1 + y_1^3 + x_2 + y_2)^{q-1} + 1 \\ &=: g(x_1, x_2, y_1, y_2). \end{aligned}$$

Let $W_m = (a_{x,y} + 1)_{x,y \in \mathbb{F}_q^2}$. Then $W_m = H_m + J$, where J is the all-1 matrix. Hence,

$$\begin{aligned} \operatorname{rank}(W_m) - \operatorname{rank}(J) \le \operatorname{rank}(H_m) &\le \operatorname{rank}(W_m) + \operatorname{rank}(J), \\ \iff \operatorname{rank}(W_m) - 1 \le \operatorname{rank}(H_m) &\le \operatorname{rank}(W_m) + 1. \end{aligned}$$

This means that W_m has almost the same rank as that of H_m.

Simplification

Now the problem is reduced to computing the rank of W_m whose entry is given by

$$h(x_1, x_2, y_1, y_2) = (x_1^3 + x_1 y_1^2 + x_1^2 y_1 + y_1^3 + x_2 + y_2)^{q-1}.$$

Note that for each fixed x_1, the map $(x_1, x_2) \to (x_1, x_2 + x_1^3)$ is a permutation on the rows labeled by (x_1, x_2) where $x_2 \in \mathbb{F}_q$. Thus we have

Proposition 2.

Let $D_m = (f(x, y))_{x, y \in \mathbb{F}_q^2}$, where

$$f(x_1, x_2, y_1, y_2) = (x_1^2 y_1 + x_1 y_1^2 + x_2 + y_2)^{q-1}.$$

Then D_m has the same -rank as that of W_m.

Matrix Factorization

We can expand the polynomial $(x_1^2 y_1 + x_1 y_1^2 + x_2 + y_2)^{q-1}$:

$$f = \sum_{(l_1,l_2,l_3,l_4)\in\Omega} \binom{q-1}{l_1, l_2, l_3, l_4} x_1^{2l_1+l_2} x_2^{l_3} y_1^{l_1+2l_2} y_2^{l_4},$$

where $\Omega := \left\{ (l_1, l_2, l_3, l_4) \middle| \sum_{i=1}^4 l_i = q-1, 0 \le l_i \le q-1, \forall i \right\}$.

Therefore, we can factor D_m as the product of two matrices

$$\begin{aligned} D_m &= LR \\ &= \begin{bmatrix} \cdots & \binom{q-1}{l_1,l_2,l_3,l_4} x_1^{2l_1+l_2} x_2^{l_3} & \cdots \end{bmatrix} \begin{bmatrix} \vdots \\ y_1^{l_1+2l_2} y_2^{l_4} \\ \vdots \end{bmatrix}, \end{aligned}$$

where the rows of L and columns of R are indexed by elements of $\mathbb{F}_q^2$, the columns of L and rows of R are indexed by elements of Ω.

Lucas' Theorem

Let N_m be the number of distinct nonzero monomials in L. That is,

$$N_m := \left| \left\{ (2l_1 + l_2, l_3) : \binom{q-1}{l_1, l_2, l_3, l_4} \equiv 1 \pmod 2 \right\} \right|.$$

We then have an upper bound for $\operatorname{rank}(D_m)$:

$$\operatorname{rank}(D_m) \leq \operatorname{rank}(L) \leq N_m.$$

Theorem 3 (Lucas' Theorem).

Let p be a prime, and express the non-negative integers $n, l_1, l_2, \ldots, l_s$ in base p as $n = \langle n_k, n_{k-1}, \ldots, n_1, n_0 \rangle_p$; $l_i = \langle l_{i,k}, l_{i,k-1}, \ldots, l_{i,1}, l_{i,0} \rangle_p$, where $0 \leq n_j, l_{i,j} \leq p-1$ for $j = 0, 1, \ldots, k$ and $i = 1, 2, \ldots, s$. Then

$$\binom{n}{l_1, l_2, \ldots, l_s} \equiv \prod_{j=0}^{k} \binom{n_j}{l_{1,j}, l_{2,j} \ldots, l_{s,j}} \pmod p.$$

Lucas' Theorem

Let $a, b, c \in \mathbb{Z}_{\geq}0$. We write $a + b \lessdot c$, if the following conditions hold:

$$a_i + b_i \leq c_i \text{ for all } i = 0, \ldots, k,$$

where $a = \langle a_k \cdots a_1 a_0 \rangle_2$, $b = \langle b_k \cdots b_1 b_0 \rangle_2$, $c = \langle c_k \cdots c_1 c_0 \rangle_2$.

For $0 \leq s \leq q-1$, define

$$B_s := \left\{ 2l_1 + l_2 \; : \; \binom{q-1}{l_1, l_2, q-1-s, l_4} \equiv 1 \pmod 2 \text{ for some } l_4 \right\}$$

and $b_s := |B_s|$. Note that $N_m = \sum_{s=0}^{q-1} b_s$. By Lucas' Theorem,

$$\binom{q-1}{l_1, l_2, q-1-s, l_4} \equiv 1 \pmod 2$$

if and only if the addition $l_1 + l_2 + (q-1-s) + l_4 = q-1$ involves no carries, which is equivalent to $l_1 + l_2 \lessdot s$ and $l_4 = s - l_1 - l_2$.

Properties of B_s, b_s

Proposition 4.

Let $s = \langle \alpha_1, \alpha_2, \ldots, \alpha_i, \beta_1, \beta_2, \ldots, \beta_j \rangle$. Then

$$B_s = B_{s_1} \times 2^j + B_{s_2} := \left\{ r2^j + t : r \in B_{s_1}, t \in B_{s_2} \right\},$$

where $s_1 = \langle \alpha_1, \alpha_2, \ldots, \alpha_i \rangle$ and $s_2 = \langle \beta_1, \beta_2, \ldots, \beta_j \rangle$.

Proposition 5.

Let i be a positive integer. Then $b_{2^{i-1}-1} = 2^i - 1$.

Proposition 6.

Let $s = \langle \alpha_1, \alpha_2, \ldots, \alpha_i, 0, \beta_1, \beta_2, \ldots, \beta_j \rangle$. Then $b_s = b_{s_1} b_{s_2}$, where $s_1 = \langle \alpha_1, \alpha_2, \ldots, \alpha_i \rangle$ and $s_2 = \langle \beta_1, \beta_2, \ldots, \beta_j \rangle$.

A Recurrence Relation of N_m

By the aforementioned properties and some calculations, we have

Proposition 7.

The sequence of numbers N_m satisfies the following recurrence relation:

$$N_m = 4N_{m-1} - 2N_{m-2}, \quad m \geq 0.$$

By solving this linear recursion, we can obtain the formula of N_m:

$$N_m = \frac{1+\sqrt{2}}{2}(2+\sqrt{2})^m + \frac{1-\sqrt{2}}{2}(2-\sqrt{2})^m, \quad m \geq 0.$$

An upper bound

Theorem 8.

Let D_m be defined as above with $m \geq 1$. Then

$$\mathrm{rank}(D_m) \leq \frac{1+\sqrt{2}}{2}(2+\sqrt{2})^m,$$

and so

$$R(D_m) = \frac{\mathrm{rank}(D_m)}{4^m} \leq \frac{1+\sqrt{2}}{2}\left(\frac{2+\sqrt{2}}{4}\right)^m \to 0$$

Therefore, the BCH family is of unit rate.

Outline

The Generalized BCH Family

We can generalize the BCH family by setting the connection set to be

$$S_{k,m} := \{(a, a^k) : a \in \mathbb{F}_q\} \subseteq \mathbb{F}_q^2,$$

where k is a fixed odd integer and $1 < k \le q-1$. Then we obtain the *generalized BCH family* F_k on the graph $\Gamma_{k,m} = \mathrm{Cay}(\mathbb{F}_q^2, S_{k,m} \backslash \{0\})$.

Remark: In the above generalization, we may require k to be odd. In fact, the matrix $H_{k,m}$ has the same rank as $H_{k/2,m}$ when k is even, where $H_{k,m}$ denotes the coset matrix of $S_{k,m}$ in $\mathbb{F}_q^2$.

We hope that the generalized BCH family F_k is also of unit rate. Before that, we want the graph $\Gamma_{k,m} = \mathrm{Cay}(\mathbb{F}_q^2, S_{k,m} \backslash \{0\})$ to be connected and triangle-free.

The Condition for $\Gamma_{k,m}$ to be Connected and Triangle-Free

When m is large enough, the graph $\Gamma_{k,m}$ is connected.

Theorem 9.

Let $k > 1$ be an odd integer. If $2^{\frac{m}{2}} + 1 > k$, then $S_{k,m}$ contains an $\mathbb{F}_2$-basis for $\mathbb{F}_q^2$; and the graph $\Gamma_{k,m}$ is connected.

The following statement gives a necessary and sufficient condition for $\Gamma_{k,m}$ to be triangle-free.

Lemma 10.

The graph $\Gamma_{k,m}$ is triangle-free if and only if the equation $(x+1)^k = x^k + 1$ in $\mathbb{F}_q$ only has solutions $x = 0, 1$.

Some Special Cases: $k = 2^r + 1$

We will consider the case when $k = 2^r + 1$ and some three bit k (more precisely, $k = 7, 11, 13$) and show that in these cases, the generalized BCH family F_k is of unit rate. For fixed k, we may omit the subscript k.

When $k = 2^r + 1$, we can deduce from Lemma 10 that it is triangle-free iff $\gcd(r, m) = 1$. Define:

$$N_m := \left| \left\{ (2^r l_1 + l_2, l_3) \, : \, \binom{q-1}{l_1, l_2, l_3, l_4} \equiv 1 \pmod 2 \right\} \right|.$$

Although we haven't found the formula of N_m, we can give an upper bound for N_m.

Theroem 11.

We have

$$N_m \leq \left(\frac{15}{16} \right)^{\frac{m}{r+1}} 4^m.$$

Some Special Cases: $k = 7, 11, 13$

In $D_m := (f(x,y))_{x,y \in \mathbb{F}_q^2}$, the (x,y)-entry is given by

$$f(x,y) = \left(x_1^{2^r} y_1 + x_1 y_1^{2^r} + x_2 + y_2\right)^{q-1},$$

which is a $(q-1)$-power of $x_1^{2^r} y_1 + x_1 y_1^{2^r} + x_2 + y_2$.

Theorem 12.

Let A, B be two matrices. Then

$$\mathrm{rank}(A \otimes B) = \mathrm{rank}(A) \cdot \mathrm{rank}(B).$$

Corollary 13.

Let A, B be two $m \times n$ matrices. Then

$$\mathrm{rank}(A \circ B) \leq \mathrm{rank}(A)\mathrm{rank}(B).$$

Reduction of Power

Let $(h(x_1, x_2, y_1, y_2))_{(x_1,x_2),(y_1,y_2)\in\mathbb{F}_q^2}$ denote the matrix in which the $((x_1, x_2), (y_1, y_2))$-entry is $h(x_1, x_2, y_1, y_2)$. We may simply write $(h(x_1, x_2, y_1, y_2))_{(x_1,x_2),(y_1,y_2)\in\mathbb{F}_q^2}$ as (h).

Lemma 14.

Let i be a non-negative integer. Then

$$\mathrm{rank}((h)) = \mathrm{rank}((h^{2^i})).$$

Proof.

Note that

$$h(x_1, x_2, y_1, y_2)^{2^i} = h(x_1^{2^i}, x_2^{2^i}, y_1^{2^i}, y_2^{2^i}).$$

Furthermore, this expression represents a permutation of both the rows and columns of (h). Thus the result follows. □

Reduction of Power

Combining Corollary 13 and Lemma 14, we have

Proposition 15.

Let $d(x_1, x_2, y_1, y_2) \in \mathbb{F}_q[x_1, x_2, y_1, y_2]$ and m, t be positive integers with $m > t$. Then

$$\mathrm{rank}((d^{2^m-1})) \leq c \cdot \left(\mathrm{rank}((d^{2^t-1}))\right)^{\frac{m}{t}},$$

where $c = \max\{\mathrm{rank}((d^{2^i-1})) : 0 \leq i < t\}$.

Remark: The above proposition tells us that the rank of $A_m = (d^{2^t-1})_{\mathbb{F}_{2^m}^2 \times \mathbb{F}_{2^m}^2}$ will give an upper bound for the rank of (d^{2^m-1}). However, the matrix A_m is changing as m increases and $\mathrm{rank}(A_m)$ would not change when m is sufficiently large.

Rank of a Polynomial

When the field size q is larger than the largest individual degree d of a polynomial $h(x_1, x_2, y_1, y_2)$, the rank of the matrix $(h(x_1, x_2, y_1, y_2))$ is invariant, depending on the polynomial h. Thus, we may call it the rank of h, denoted by $\mathrm{rank}(h)$.

Lemma 16.

Let $h \in \mathbb{F}_2[x_1, x_2, y_1, y_2]$. Assume that $d = \max\{\deg_{x_1} h, \deg_{x_2} h, \deg_{y_1} h, \deg_{y_2} h\}$, where $\deg_{x_1} h$ is the degree of h in variable x_1. If $q > d$, then

$$\mathrm{rank}((h)_{\mathbb{F}_q^2 \times \mathbb{F}_q^2}) = \mathrm{rank}(h).$$

Proof Using a Computer

According to Proposition 15 and Lemma 16,

Theorem 17.

If there exists a positive integer t such that

$$\mathrm{rank}(d^{2^t-1}) < 4^t,$$

then the generalized BCH family F_k is of unit rate.

Using Magma, we know that $\mathrm{rank}(d^{2^6-1}) = 3256 < 4096 = 4^6$ for F_7,
$\mathrm{rank}(d^{2^7-1}) = 15018 < 16384 = 4^7$ for F_{11}, and
$\mathrm{rank}(d^{2^7-1}) = 14442 < 16384 = 4^7$ for F_{13}.

Corollary 18.

The generalized BCH families F_7, F_{11} and F_{13} are all of unit rate. $\square$

Outline

Relation to $R(3, t)$

As we shall see, the storage codes on triangle-free graphs are related to the Ramsey number $R(3, t)$. If we use $\alpha(\Gamma)$ to denote the independence number of the graph Γ, then we have the following result.

Lemma 19.

Let C be an $[n, k]_q$ storage code on a graph Γ. Then we have

$$\alpha(\Gamma) \leq n - k.$$

Hence, if we have an upper bound for the rank of the parity-check matrix of C, we then can bound the independence number $\alpha(\Gamma)$. Employing the BCH family F_3, we obtain a constructive lower bound for $R(3, t)$:

$$R(3, t) \geq \Omega(t^{\log_{2+\sqrt{2}} 4}).$$

Constructive Lower Bounds for $R(3, t)$

Utilizing the same Cayley graph Γ, Noga Alon provided a better upper bound in [Alo95], applying the Carlitz-Uchiyama bound [CU57] for the eigenvalues of Γ and then using Hoffman's ratio bound. This approach led Alon to a constructive lower bound $R(3, t) \geq \Omega(t^{4/3})$ which is better than the result we obtained above.

Currently, the best-known constructive lower bounds of $R(3, t)$ are $\Omega(t^{3/2})$, as seen in [Alo94; KPR10]. Note that $R(3, t) \sim t^2 / \log t$ [Kim95].

As a consequence, the rate of convergence of $1/(1 - R(C_n))$ cannot exceed $\sqrt{n}/\sqrt{\log n}$.

References I

[Maz14] Arya Mazumdar. "On a duality between recoverable distributed storage and index coding". In: *2014 IEEE International Symposium on Information Theory*. IEEE. 2014, pp. 1977–1981.

[Maz15] Arya Mazumdar. "Storage capacity of repairable networks". In: *IEEE Transactions on Information Theory* 61.11 (2015), pp. 5810–5821.

[SD14] Karthikeyan Shanmugam and Alexandros G Dimakis. "Bounding multiple unicasts through index coding and locally repairable codes". In: *2014 IEEE International Symposium on Information Theory*. IEEE. 2014, pp. 296–300.

[CM11] Demetres Christofides and Klas Markström. "The guessing number of undirected graphs". In: *the electronic journal of combinatorics* (2011), p.192–p.192.

[CDR14] Peter J Cameron, Anh N Dang, and Soren Riis. "Guessing games on triangle-free graphs". In: *arXiv preprint arXiv:1410.2405* (2014).

[BZ22] Alexander Barg and Gilles Zémor. "High-rate storage codes on triangle-free graphs". In: *IEEE Transactions on Information Theory* 68.12 (2022), pp. 7787–7797.

[GH20] Alexander Golovnev and Ishay Haviv. "The (generalized) orthogonality dimension of (generalized) Kneser graphs: Bounds and applications". In: *arXiv preprint arXiv:2002.08580* (2020).

[BSY22] Alexander Barg, Moshe Schwartz, and Lev Yohananov. "Storage codes on triangle-free graphs with asymptotically unit rate". In: *arXiv preprint arXiv:2212.12117* (2022).

References III

[HX23] Hexiang Huang and Qing Xiang. "Construction of storage codes of rate approaching one on triangle-free graphs". In: *Designs, Codes and Cryptography* (2023), p.3901–p.3913.

[Alo95] Noga Alon. "Tough Ramsey graphs without short cycles". In: *Journal of Algebraic Combinatorics* 4 (1995), pp. 189–195.

[CU57] Leonard Carlitz and Saburo Uchiyama. "Bounds for exponential sums". In: (1957).

[Alo94] Noga Alon. "Explicit Ramsey graphs and orthonormal labelings". In: *the electronic journal of combinatorics* (1994), R12–R12.

[KPR10] Alexandr Kostochka, Pavel Pudlák, and Vojtech Rödl. "Some constructive bounds on Ramsey numbers". In: *Journal of Combinatorial Theory, Series B* 100.5 (2010), pp. 439–445.

References IV

[Kim95] Jeong Han Kim. "The Ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$". In: *Random Structures & Algorithms* 7.3 (1995), pp. 173–207.