A graph-theoretic proof for an upper bound of the maximum block code size

Jeffrey Kam

June 2021

1 Background

Coding theory is an active area in information theory with a wide-array of applications, from data transmission to cryptography. One of the fundamental codes in coding theory is the block code, as defined below along with some of its other properties.

Definition 1.1 Let A be an alphabet set. A *word* is a vector with elements in A. A *code* a set of words and the words are called *codewords*. We say a code C is a *block code* of length n and size M, or $[n, M]$-code in short, if $|C| = M$ and all words in C are of length n.

Definition 1.2 The *distance* between two words u, v in a code is defined by the number of positions in which their elements differ. The *distance* of a code is the minimum distance between all pairs of words in the code. We say a code C is an $[n, M, d]$-code if it is a block code of length n, size M, and distance d.

Definition 1.3 Given $n, q, d \in \mathbb{Z}_{\geq 0}$, we define $A_q(n, d)$ as the largest value M such that there exists an $[n, M, d]$-code over some alphabet A of size q.

Motivated by the applications, researchers have worked on understanding various properties of different codes, such as bounds on the block code capacity. In this short paper, we will prove an upper bound of the rate of a general block code, stated as theorem 2.3. This elementary result is known and has been presented as an exercise in [1], where the implied approach seems to be linear-algebraic. Here, we will instead turn to a graph-theoretic approach and provide an alternative proof that uses the Turán's theorem. It should be noted that this technique is not new, as there have been other instances in coding theory where results have been proved by applying Turán's theorem, such as the generalized Gilbert-Varshamov bound [2].

Theorem 1.4 (Turán, 1941)
The maximum number of edges in a graph of n vertices that does not contain a K_{q+1} is $\frac{q-1}{q}\frac{n^2}{2}$.

2 Proof of the main result

Definition 2.1 Let C be an $[n, M, d]$-code with codewords $\{c_1, \cdots, c_M\}$ over alphabet A, where the size of A is q. Furthermore, for a codeword c, we write $c[i]$ as the i^{th} symbol in codeword c. Then, the *difference graph* of C, denoted G_C, is a graph with $n \times M$ vertices, such that

$$V(G_C) = \{v_{1,1}, \cdots, v_{1,M}, \cdots, v_{n,1}, \cdots, v_{n,M}\},$$

and for each pair of $v_{i,j}, v_{k,l} \in G_C$,

$$v_{i,j}v_{k,l} \in E(G_C) \text{ if and only if } i = k \text{ and } c_j[i] \neq c_l[k].$$

Also, for each $1 \leq i \leq n$, we define $H_i := \{v_{i,1}, \cdots, v_{i,M}\}$ as a subset of vertices of size M. Similarly, for each $1 \leq j \leq M$, we define $J_i := \{v_{1,j}, \cdots, v_{n,j}\}$ as a subset of vertices of size n.

In other words, there is an edge between vertices $v_{i,j}$ and $v_{i,l}$ if the i^{th} symbol of codewords c_j and c_l are different. Also, note that there is no edge between vertices of H_i and H_j if $i \neq j$. Before we prove theorem 2.3, we have to first prove a lemma.

Lemma 2.2 Let C be an $[n, M, d]$-code over alphabet A of size q and G_C be the distance graph of C. Then, K_{q+1} is not a subgraph of G_C, where K_{q+1} is the $(q+1)$-clique.

Proof. Suppose on the contrary that G_C has a $(q+1)$-clique as a subgraph. By construction, edges in G_C only exist between vertices within the same H_i. Let K be a $(q+1)$-clique in H_i for some i, and without loss of generality, we assume $\{v_{i,1}, v_{i,2}, \cdots, v_{i,q}, v_{i,q+1}\}$ are the vertices of K. By definition of edges in G_C, it means that the symbols $c_1[i], c_2[i], \cdots, c_{q+1}[i]$ are all pairwise different. However, this implies there are $q+1$ distinct symbols, contradicting that A is of size q. □

We are now ready to prove the main theorem of this paper.

Theorem 2.3 Let $n, d, q \in \mathbb{Z}_{\geq 0}$. If $d > \frac{n(q-1)}{q}$, then $A_q(n, d) \leq \frac{dq}{dq - n(q-1)}$.

Proof. Fix $n, d \in \mathbb{Z}_{\geq 0}$ and A to be an alphabet set of size q. Let $M \in \mathbb{Z}_{\geq 0}$ and C be an $[n, M, d]$-code over A with codewords $\{c_1, \cdots, c_M\}$. Let G_C be the difference graph of C. Now, suppose, for a contradiction, $\frac{dM(M-1)}{2n} > \frac{q-1}{q}\frac{M^2}{2}$. We will show that G_C has a $(q+1)$-clique, contradicting lemma 2.2.

Since C has distance d, G_C has at least $\sum_{i=1}^{M-1}(d \cdot i) = \frac{d(M-1)M}{2}$ number of edges. This is because if we fixed a codeword c_i, then for every other codeword $c_j \in C$, c_i and c_j will have at least d positions where their symbols differ, and so, it has d edges between vertices in J_i and J_j. From the definition of G_C, each H_i are essentially different components of the graph. In particular, edges in G_C can only be between

vertices within the same H_i. Hence, by the pigeon hole principle, there exists some H_i where its induced subgraph has at least $\lceil \frac{d(M-1)M}{2n} \rceil$ edges.

Let $\tilde{G}$ be the induced subgraph of H_i. We want to show that there is a $(q+1)$-clique in $\tilde{G}$. By our initial assumption, $\frac{dM(M-1)}{2n} > \frac{q-1}{q}\frac{M^2}{2}$, so $\left|E(\tilde{G})\right| > \frac{q-1}{q}\frac{M^2}{2}$. By Turán's theorem, the maximum number of edges in a K_{q+1}-free graph of M vertices is $\frac{q-1}{q}\frac{M^2}{2}$. This means $\tilde{G}$ is not K_{q+1}-free by Turán's theorem. However, this contradicts 2.2, so we have $\frac{dM(M-1)}{2n} \leq \frac{q-1}{q}\frac{M^2}{2}$. Solving for M in the inequality yields

$$\begin{aligned}
\frac{dM(M-1)}{2n} &\leq \frac{q-1}{q}\frac{M^2}{2} \\
\frac{M-1}{M} &\leq \frac{n(q-1)}{dq} \\
1 - \frac{1}{M} &\leq \frac{n(q-1)}{dq} \\
M &\leq \frac{dq}{dq - n(q-1)}.
\end{aligned}$$

Hence, we have $A_q(n,d) \leq \frac{dq}{dq-n(q-1)}$ as desired. □

3 Conclusion

The theorem we have shown above is already a known upper bound of $A_q(n,d)$, but we hope that this new approach will be useful in solving coding theory related problems in the future.

Acknowledgments

This proof is a generalization of my solution to a question posed in my coding theory course, which asked for a bound of $A_2(n,d)$ for small specific parameters n and d. The general result I have proved about has also been found in other coding theory text, including the CMU lecture notes [1], but the approach is different.

References

[1] Venkatesan Guruswami. Notes 4: Elementary bounds on codes, Jan 2010. URL `https://www.cs.cmu.edu/~venkatg/teaching/codingtheory/notes/notes4.pdf`.

[2] L.M.G.M. Tolhuizen. The generalized gilbert-varshamov bound is implied by turan's theorem [code construction]. *IEEE Transactions on Information Theory*, 43(5):1605–1606, 1997. doi: 10.1109/18.623158.