Two results about the hypercube

József Balogh, Tamás Mészáros and Adam Zsolt Wagner

October 25, 2017

Abstract

First we consider families in the hypercube $Q_n$ with bounded VC dimension. Frankl raised the problem of estimating the number $m(n, k)$ of maximal families of VC dimension $k$. Alon, Moran and Yehudayoff showed that

$$n^{(1+o(1))\binom{n}{k}} \leq m(n, k) \leq n^{(1+o(1))\binom{n}{k}}.$$  

We close the gap by showing that $\log(m(n, k)) = (1 + o(1))\binom{n}{k} \log n$ and show how a tight asymptotic for the logarithm of the number of induced matchings between two adjacent small layers of $Q_n$ follows as a corollary.

Next, we consider the integrity $I(Q_n)$ of the hypercube, defined as

$$I(Q_n) = \min\{|S| + m(Q_n \setminus S) : S \subseteq V(Q_n)\},$$

where $m(H)$ denotes the number of vertices in the largest connected component of $H$. Beineke, Goddard, Hamburger, Kleitman, Lipman and Pippert showed that $c_2 \sqrt{n} \leq I(Q_n) \leq C_2 \sqrt{n} \log n$ and suspected that their upper bound is the right value. We prove that the truth lies below the upper bound by showing that $I(Q_n) \leq C_2 \sqrt{n} \log n$.

1 Introduction

Throughout the paper we will use standard notation. For an integer $n \geq 1$ we will write $[n]$ for the set $\{1, 2, \ldots, n\}$, $\mathcal{P}(n)$ for its power set and $\binom{[n]}{k}$ for the collection of subsets of size $k$. For $\bigcup_{i=0}^{k} \binom{[n]}{i}$ (resp. $\sum_{i=0}^{k} \binom{n}{i}$) we will use the shorthand notation $\binom{[n]}{\leq k}$ (resp. $\binom{n}{\leq k}$).

For an integer $n \geq 1$ the graph $Q_n$, the hypercube of dimension $n$, has vertex set $V(Q_n) = \{0, 1\}^n$ and two vertices are connected if they only differ in one coordinate. There is a natural bijection between the vertex set of $Q_n$ and $\mathcal{P}(n)$, and we will use them interchangeably. Here we consider several enumerative and extremal properties of vertex subsets of this graph.
1.1 Enumerative problems

We say that a family $\mathcal{F} \subseteq \mathcal{P}(n)$ shatters a set $S \subseteq [n]$ if for all $A \subseteq S$ there exists a set $B \in \mathcal{F}$ with $B \cap S = A$. Let $\text{Sh}(\mathcal{F}) := \{S \subseteq [n] : \mathcal{F} \text{ shatters } S\}$. The Vapnik-Chervonenkis dimension, $\text{VC-dimension}$ for short, of a family $\mathcal{F} \subseteq \mathcal{P}(n)$ is defined as

$$\text{VC}(\mathcal{F}) = \max \{|S| : \mathcal{F} \text{ shatters } S\}.$$ Pajor’s version [9] of the Sauer-Shelah lemma states that we always have $|\text{Sh}(\mathcal{F})| \geq |\mathcal{F}|$. We say that a family $\mathcal{F} \subseteq \mathcal{P}(n)$ is (shattering-)extremal if $|\text{Sh}(\mathcal{F})| = |\mathcal{F}|$. For example, if $\mathcal{F}$ is a down-set then it is extremal, simply because in this case $\text{Sh}(\mathcal{F}) = \mathcal{F}$. For an integer $k \geq 0$ let $\text{ExVC}(n,k)$ be the number of extremal families in $\mathcal{P}(n)$ with VC dimension at most $k$. The study of these extremal families was initiated by Bollobás, Leader and Radcliffe [5] and since then many interesting results have been obtained in connection with these combinatorial objects.

The Sauer-Shelah lemma [10] states that for any family $\mathcal{F} \subseteq \mathcal{P}(n)$ we have $|\mathcal{F}| \leq {n \choose \text{VC}(\mathcal{F})}$. A family is called maximal if $|\mathcal{F}| = {n \choose \text{VC}(\mathcal{F})}$. Clearly, every maximal family is extremal. Frankl [6] raised the question of estimating $m(n,k)$, the number of maximal families in $\mathcal{P}(n)$ of VC dimension $k$, and showed that

$$2^{\binom{n-1}{k-1}} \leq m(n,k) \leq 2^{n\binom{n-1}{k-1}}.$$\hspace{1cm}(1.1)

Alon, Moran and Yehudayoff [11] showed that for constant $k \geq 2$ we have, as $n \to \infty$, that

$$m(n,k) = n^{\Theta(1)} \binom{n}{k}.$$\hspace{1cm}(1.1)

We close the gap and show that the upper bound of (1.1) is correct, even if we allow $k$ to grow as $k = n^{o(1)}$.

**Theorem 1.1.** Let $k = n^{o(1)}$. Then $m(n,k) = n^{(1+o(1))\binom{n}{k}}$.

Next we consider the problem of finding a small family $\mathcal{F} \subseteq \{0,1\}^n$ such that all connected components of $Q_n \setminus \mathcal{F}$ are small. For a graph $H$ let $m(H)$ denote the maximum number of vertices in a component of $H$. The integrity $I(G)$ of a graph $G$, introduced by Barefoot, Entringer and Swart [2] to measure the vulnerability of a network, is defined by

$$I(G) = \min\{|S| + m(G \setminus S) : S \subseteq V(G)\}.$$\hspace{1cm}(1.2)

In [7] it was conjectured that for the hypercube we have $I(Q_n) = 2^{n-1} + 1$, but Beineke, Goddard, Hamburger, Kleitman, Lipman and Pippert [3] disproved this conjecture and obtained the following bounds:
**Theorem 1.3** (Beineke, Goddard, Hamburger, Kleitman, Lipman, Pippert). There exists constants $c, C > 0$ such that

$$c \frac{2^n}{\sqrt{n}} \leq I(Q_n) \leq C \frac{2^n}{\sqrt{n}} \log n.$$ 

Their upper bound was obtained by a series of ‘orthogonal’ cuts and they suspected it to be of the correct order of magnitude. We show that the true value of $I(Q_n)$ lies below their upper bound.

**Theorem 1.4.** There exists a constant $C > 0$ such that the integrity of the hypercube satisfies

$$I(Q_n) \leq C \frac{2^n}{\sqrt{n}} \log n.$$

This note is organized as follows. We prove Theorems 1.1 and 1.2 in Section 2 and Theorem 1.4 in Section 3. We often omit floor and ceiling signs when they are not crucial, to increase the clarity of our presentation.

## 2 The proofs of Theorems 1.1 and 1.2

**Proposition 2.1.** If $k = n^{o(1)}$ then $m(n, k) \geq \text{IndMat}(n, k) \geq n^{(1+o(1))\binom{n}{k}}$.

**Proof.** Let $S$ be the collection of all induced matchings between layers $\binom{[n]}{k}$ and $\binom{[n]}{k+1}$ and let $M(n, k)$ be the collection of all maximal families $\mathcal{F} \subset \mathcal{P}(n, k)$ of VC-dimension $k$. We will first define an injection $\phi : S \to M(n, k)$ and then show $|S| \geq n^{(1+o(1))\binom{n}{k}}$.

Given an element $S \in S$, define $\phi(S)$ as follows: $\phi(S)$ contains all sets of size at most $k-1$, those sets of size $k$ which are not covered by edges in $S$ and those sets of size $k+1$ which are covered by edges in $S$. Observe that we can reconstruct $S$ from $\phi(S) \cap \left(\binom{[n]}{k} \cup \binom{[n]}{k+1}\right)$ hence we have $\phi(S_1) \neq \phi(S_2)$ for any $S_1, S_2 \in S$ with $S_1 \neq S_2$. As $|\phi(S)| = \binom{n}{\leq k}$ it remains to show that $\text{VC}(\phi(S)) = k$ for all $S \in S$.

As $\phi(S) \subset \binom{[n]}{\leq k+1}$ we have $\text{VC}(\phi(S)) \leq k+1$. Suppose for contradiction that $\phi(S)$ shatters a set $A \in \binom{[n]}{k+1}$. Then since $\phi(S) \subset \binom{[n]}{\leq k+1}$ we must have $A \in \phi(S)$. This means that the induced matching $S$ meets $A$, let the other endpoint of this edge be $B \in \binom{[n]}{k}$. Hence $B \notin \phi(S)$ and as $S$ is induced, for all $A' \in \binom{[n]}{k+1}$ with $A' \neq A$ and $B \subset A'$ we have $A' \notin \phi(S)$. So there is no set $C \in \phi(S)$ with $C \cap A = B$, contradicting the assumption that $A$ is shattered by $\phi(S)$.

Now we turn to showing that $|S| \geq n^{(1+o(1))\binom{n}{k}}$. Let $\varepsilon = \varepsilon(n) = o\left(1/k^2\right)$ be a sequence converging to zero as $n \to \infty$. Let $A = \{1, 2, \ldots, \varepsilon n\}$ and $B = [n] \setminus A$. Call a matching $M$ between layers $\binom{[n]}{k}$ and $\binom{[n]}{k+1}$ good if for every edge $(C_1, C_2)$ of $M$ we have that $C_1 \in \binom{B}{k}$ and $C_2 \in \binom{[n]}{k+1} \setminus \binom{B}{k+1}$. Every good matching is induced, and the number of good matchings that cover $\binom{B}{k}$ is already

$$(\varepsilon n)^{(1-\varepsilon)n} \geq \left(n^{1-o(1)}\right)^{(1-2\varepsilon)k\binom{n}{k}} = n^{(1-o(1))\binom{n}{k}}.$$ 

This completes the proof. □
Proof of Theorem 1.1. Let \( k = n^{o(1)} \). By Proposition 2.1 we have \( n^{(1+o(1))\binom{n}{k}} \leq m(n,k) \) and by (1.1) we have \( n^{(1+o(1))\binom{n}{k}} \geq m(n,k) \).

**Proof of Theorem 1.2.** Let \( k = n^{o(1)} \). For integers \( n, m \geq 0 \) let \( \text{Conn}(n,m) \) be the number of connected induced subgraphs of \( Q_n \) on exactly \( m \) vertices. First, notice that \( m(n,k) \leq \text{ExVC}(n,k) \) as every maximal family is also extremal. Second, as every extremal family induces a connected subgraph of \( Q_n \) (see e.g. [8]) and as by the Sauer inequality any family of VC dimension \( k \) has size at most \( \binom{n}{k} \) we have

\[
\text{ExVC}(n,k) \leq \sum_{i=0}^{\binom{n}{k}} \text{Conn}(n,i).
\]

We now proceed following the exact same ideas as the ones described in [1]. It is known (e.g., Problem 45 in [4]) that the number of connected subgraphs of size \( \ell \) in a graph of order \( N \) and maximum degree \( D \) is at most \( N(e(D-1)^{\ell-1} \leq N(eD)^{\ell} \). In our case, plugging in \( \ell = i, N = 2^n \) and \( D = n \) yields

\[
\text{ExVC}(n,k) \leq \sum_{i=0}^{\binom{n}{k}} 2^n (en)^i \leq 2^{n+1}(en)^{\binom{n}{k}} = n^{(1+o(1))\binom{n}{k}},
\]

where for the last equality we used that \( k = n^{o(1)} \). Hence together with Proposition 2.1 we have that

\[
n^{(1+o(1))\binom{n}{k}} \leq \text{IndMat}(n,k) \leq m(n,k) \leq \text{ExVC}(n,k) \leq n^{(1+o(1))\binom{n}{k}}.
\]

3 The proof of Theorem 1.4

Our goal in this section is to improve the upper bound from Theorem 1.3. In [3] a set was formed by a series of orthogonal cuts which yielded the upper bound in Theorem 1.3. Instead, we will delete small spheres around some appropriately chosen points.

For a vertex \( x \) of \( Q_n \) and a non-negative integer \( r \) for the ball of radius \( r \) with center \( x \) write

\[
B_r(x) = \{ y \in V(Q_n) \mid d(y,x) \leq r \}
\]

and for the sphere of radius \( r \) with center \( x \)

\[
S_r(x) = \{ y \in V(Q_n) \mid d(y,x) = r \},
\]

where \( d(.,.) \) denotes the Hamming distance.

Let us define \( \alpha \) and \( r_0 \) by the equations

\[
\frac{1}{e^{2\alpha^2}} = \frac{\sqrt{\log n}}{\sqrt{n}} \quad \text{and} \quad r_0 := \left\lfloor \frac{n}{2} - \alpha \sqrt{n} \right\rfloor.
\]

This \( r_0 \) will be the radius of the spheres we delete. Note that \( \alpha \) is very close to \( \sqrt{\log n/2} \). The proof of Theorem 1.4 will be based on the following standard estimates, we provide a full proof for completeness at the end of this section.
Lemma 3.1. For every $x \in \mathcal{P}(n)$

$$|B_{r_0}(x)| = \left( \frac{n}{\sqrt{\log n}} \right)$$

and

$$|S_{r_0}(x)| \geq \left( \frac{n}{r_0} \right) \sqrt{n} = \Theta \left( \frac{2^n}{\sqrt{n}} \right).$$

The proof of Theorem 1.4 follows by repeatedly removing spheres of radius $r_0$ around some appropriately chosen points.

Proof of Theorem 1.4. For a family $\mathcal{F} \subseteq \mathcal{P}(n)$ and $x \in \mathcal{P}(n)$ let $B(\mathcal{F}, x) := B_{r_0}(x) \cap \mathcal{F}$ and $S(\mathcal{F}, x) := S_{r_0}(x) \cap \mathcal{F}$. Fixing the family $\mathcal{F}$ and picking $x \in \mathcal{P}(n)$ uniformly at random, using linearity of expectation and Lemma 3.1, for the expectation $\mathbb{E}_x$ of $|B(\mathcal{F}, x)|$ and $|S(\mathcal{F}, x)|$ we get

$$\mathbb{E}_x(|B(\mathcal{F}, x)|) = \Theta \left( |\mathcal{F}| \frac{\log n}{\sqrt{n}} \right)$$

and

$$\mathbb{E}_x(|S(\mathcal{F}, x)|) = \Theta \left( |\mathcal{F}| \frac{\log n}{n} \right).$$

In particular there is an absolute constant $C > 0$ such that for every $n$ and $\emptyset \neq \mathcal{F} \subseteq \mathcal{P}(n)$ there exists an $x = x(\mathcal{F})$ satisfying

$$|S(\mathcal{F}, x)| \leq |B(\mathcal{F}, x)| C \frac{\log n}{\sqrt{n}}.$$ 

Now to formalize our strategy, fix such a function $x : \mathcal{P}(\mathcal{P}(n)) \setminus \{\emptyset\} \to \mathcal{P}(n)$. Let $\mathcal{F}_0 = 2^{[n]}$, $S_0 = \emptyset$ and for $i = 0, 1, \ldots$ set $\mathcal{F}_{i+1} := \mathcal{F}_i \setminus B(\mathcal{F}_i, x(\mathcal{F}_i))$ and $S_{i+1} := S_i \cup S(\mathcal{F}_i, x(\mathcal{F}_i))$. So, each time peel off new components by removing some appropriately chosen sphere of radius $r_0$. Let $\ell$ be the least integer such that $\mathcal{F}_\ell = \emptyset$.

$$|S_\ell| = \sum_{i=0}^{\ell-1} S(\mathcal{F}_i, x(\mathcal{F}_i)) \leq C \frac{\log n}{\sqrt{n}} \sum_{i=0}^{\ell-1} B(\mathcal{F}_i, x(\mathcal{F}_i)) = C \frac{\log n}{\sqrt{n}} 2^n.$$

Deleting $S_\ell$ from $Q_n$ leaves a graph with all components being contained in some ball of radius $r_0$, and by Lemma 3.1 having size at most $\left( \frac{n}{r_0} \right) = \Theta \left( \frac{2^n}{\sqrt{n}} \right).$  

All that remains is to prove Lemma 3.1.

Proof of Lemma 3.1. As for $n$ fixed, the function $f(x) = \binom{n}{x} \binom{n}{x-1} = \frac{n-x+1}{x}$ is decreasing we have that

$$\left( \frac{n}{r_0} \right) \left( \frac{n}{r_0 - \frac{\sqrt{n}}{\alpha}} \right)^{-1} = \prod_{i=r_0-\frac{\sqrt{n}}{\alpha}+1}^{r_0} \left( \frac{n}{i} \right) \left( \frac{n}{i-1} \right)^{-1} \geq \left( \frac{n}{r_0} \right) \left( \frac{n}{r_0 - 1} \right)^{-1} \sqrt{n/\alpha}$$

$$= \left( \frac{n - r_0 + 1}{r_0} \right) \sqrt{n/\alpha} \geq \left( \frac{n/2}{n/2 - \alpha \sqrt{n}} \right) \sqrt{n/\alpha} \geq \left( 1 + \frac{\alpha}{\sqrt{n}} \right) \sqrt{n/\alpha} \geq 1.1.$$ 

Also note that, again using that $f(x)$ is decreasing, the same lower bound holds if we replace $r_0$ by any other value $r$ with $\frac{\sqrt{n}}{\alpha} \leq r \leq r_0$. 

5
Similarly
\[
\binom{n}{r_0} \binom{r_0 - \sqrt{n}}{\sqrt{n}/\alpha} = \prod_{i=r_0-\sqrt{n}+1}^{r_0} \binom{n}{i} \frac{1}{i-i} \leq \left( \binom{r_0 - \sqrt{n}/\alpha}{r_0 - \sqrt{n}/\alpha} \right) \leq \left( \frac{n/2 + 2\alpha \sqrt{n}}{n/2 - 2\alpha \sqrt{n}} \right) \leq \frac{1 + \frac{10\alpha}{\sqrt{n}}}{\sqrt{n}} \leq 10^{10}.
\]

Accordingly,
\[
\left( \frac{n}{r_0} \right) \geq \sum_{i=r_0-\sqrt{n}/\alpha}^{r_0} \frac{\sqrt{n}}{\alpha} \left( \frac{n}{r_0 - \sqrt{n}/\alpha} \right) \geq \frac{\sqrt{n}}{\alpha} \left( \frac{n}{r_0} \right) \left( \frac{n}{r_0 - \sqrt{n}/\alpha} \right) \geq \Omega \left( \frac{n}{r_0} \frac{\sqrt{n}}{\log n} \right)
\]
and
\[
\left( \frac{n}{r_0} \right) = \frac{\sqrt{n}}{\alpha} \sum_{i=r_0-j\sqrt{n}/\alpha}^{r_0} \frac{\sqrt{n}}{\alpha} \left( \frac{n}{r_0 - j \sqrt{n}/\alpha} \right) \left( \frac{n}{r_0 - j \sqrt{n}/\alpha} \right) \leq \frac{\sqrt{n}}{\alpha} \left( \frac{n}{r_0} \right) \sum_{j=1}^{r_0 \sqrt{n}/\alpha} \left( \frac{1}{1.1} \right) = O \left( \frac{n}{r_0} \frac{\sqrt{n}}{\log n} \right).
\]

These inequalities together give
\[
\left( \frac{n}{r_0} \right) = \Theta \left( \frac{n}{r_0} \frac{\sqrt{n}}{\log n} \right).
\]

On the other hand we have
\[
\left( \frac{n}{r_0} \right) = \left( \frac{n}{n/2} \right) \prod_{i=r_0}^{n/2-1} \left( \frac{n}{i+1} \right) \left( \frac{n}{i+1} \right) \left( \frac{n}{n/2+i-1} = \left( \frac{n}{n/2} \right) ^{\alpha \sqrt{n}/n/2+i+1} = \left( \frac{n}{n/2} \right) ^{\alpha \sqrt{n}/n/2+i+1},
\]
and hence, using the inequality \( 1 - x \leq e^{-x^2} \), we get
\[
\left( \frac{n}{r_0} \right) \leq \left( \frac{n}{n/2} \right) \prod_{i=1}^{n/2} e^{-\frac{2i-1}{n/2+i+1}} = \left( \frac{n}{n/2} \right) e^{-\sum_{i=1}^{n/2} \frac{2i-1}{n/2+i+1}} = O \left( e^{-2\alpha^2 \left( \frac{n}{n/2} \right)} \right),
\]
and similarly, using the inequality \( 1 - x \geq e^{-\frac{2x}{1-x}} \) for \( 0 < x < 1 \), we get
\[
\left( \frac{n}{r_0} \right) \leq \left( \frac{n}{n/2} \right) \prod_{i=1}^{n/2} e^{-\frac{2i-1}{n/2+i+1}} = \left( \frac{n}{n/2} \right) e^{-\sum_{i=1}^{n/2} \frac{2i-1}{n/2+1+i+1}} = \Omega \left( e^{-2\alpha^2 \left( \frac{n}{n/2} \right)} \right).
\]
Together this gives
\[
\left( \frac{n}{r_0} \right) = \Theta \left( e^{-2\alpha^2 \left( \frac{n}{n/2} \right)} \right) = \Theta \left( \frac{\log n}{\sqrt{n}} \cdot \frac{2^n}{\sqrt{n}} \right) = \Theta \left( \frac{2^n \log n}{n} \right),
\]
and the claim follows. \(\square\)

**Acknowledgement:** We are grateful to Shagnik Das for his many insightful comments regarding the manuscript.
References

[1] N. Alon, S. Moran and A. Yehudayoff, Sign rank versus VC dimension, Electronic Colloquium on Computational Complexity (ECCC), Vol. 21. No. 135. (2014).

[2] C. A. Barefoot, R. Entringer and H. Swart, Vulnerability in graphs – a comparative survey, Journal of Combinatorial Mathematics and Combinatorial Computing Vol. 1: 13–22 (1987).

[3] L. W. Beineke, W. Goddard, P. Hamburger, D. J. Kleitman, M. J. Lipman and R. E. Pippert, The integrity of the cube is small, Journal of Combinatorial Mathematics and Combinatorial Computing Vol. 9: 191–193 (1991).

[4] B. Bollobás, The art of mathematics: Coffee time in Memphis, Cambridge University Press (2006).

[5] B. Bollobás, I. Leader, A.J. Radcliffe, Reverse Kleitman Inequalities, Proceedings of the London Mathematical Society, Vol. 3.1: 153–168 (1989).

[6] P. Frankl, Traces of antichains, Graphs and Combinatorics Vol. 5.1: 295–299 (1989).

[7] W. Goddard and H. C. Swart, On the integrity of combinations of graphs, Journal of Combinatorial Mathematics and Combinatorial Computing Vol. 4: 3–18 (1988).

[8] G. Greco, Embeddings and trace of finite sets, Information Processing Letters Vol. 67:199–203 (1998).

[9] A. Pajor, Sous-espaces $l_1^n$ des espaces de Banach, No. 16. Hermann, 1985.

[10] N. Sauer, On the density of families of sets, Journal of Combinatorial Theory, Series A 13.1 (1972): 145-147.