Contents lists available at ScienceDirect

Discrete Applied Mathematics

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

Note

Two results about the hypercube

József Balogh [a], Tamás Mészáros [b], Adam Zsolt Wagner [c,*]

[a] Department of Mathematical Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[b] Freie Universität Berlin, 14195 Berlin, Germany
[c] University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 23 October 2017
Received in revised form 19 March 2018
Accepted 28 March 2018
Available online xxxx

Keywords:
VC-dimension
Integrity
Hypercube
Network vulnerability

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 maximum families of VC dimension k. Alon, Moran and Yehudayoff showed that

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

We close the gap by showing that $\log(m(n,k)) = (1+o(1))\binom{n}{k}\log n$. We also show how a bound on 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 $c\frac{2^n}{\sqrt{n}} \le I(Q_n) \le C\frac{2^n}{\sqrt{n}}\log n$ and suspected that their upper bound is the right value. We prove that the truth lies below the upper bound by showing that $I(Q_n) \le C\frac{2^n}{\sqrt{n}}\sqrt{\log n}$.

1. Introduction

Throughout the paper we will use standard notation. For an integer $n \ge 1$ we will write $[n]$ for the set $\{1, 2, \dots, n\}$, $\mathcal{P}(n)$ 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 $\binom{[n]}{\le k}$ (resp. $\binom{n}{\le k}$).

For an integer $n \ge 1$ 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 by surveying two problems related to the structure of subsets of the hypercube.

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 $\mathrm{Sh}(\mathcal{F}) := \{S \subseteq [n] : \mathcal{F} \text{ shatters } S\}$. The *Vapnik–Chervonenkis dimension*, *VC-dimension* for short, of a family $\mathcal{F} \subseteq \mathcal{P}(n)$ is

* Corresponding author.
E-mail addresses: jobal@math.uiuc.edu (J. Balogh), tamas.meszaros@fu-berlin.de (T. Mészáros), zawagne2@illinois.edu (A.Z. Wagner).

https://doi.org/10.1016/j.dam.2018.03.086

defined as

$$\mathrm{VC}(\mathcal{F}) = \max\{|S| : \mathcal{F} \text{ shatters } S\}.$$

Pajor's version [14] of the Sauer–Shelah lemma states that we always have $|\mathrm{Sh}(\mathcal{F})| \geq |\mathcal{F}|$. We say that a family $\mathcal{F} \subseteq \mathcal{P}(n)$ is *(shattering-)extremal* if $|\mathrm{Sh}(\mathcal{F})| = |\mathcal{F}|$. For example, if $\mathcal{F}$ is a down-set (meaning that whenever $A \in \mathcal{F}$ and $B \subset A$, we have $B \in \mathcal{F}$) then it is extremal, simply because in this case $\mathrm{Sh}(\mathcal{F}) = \mathcal{F}$. For an integer $k \geq 0$ let $\mathrm{ExVC}(n, k)$ 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]. Their work was then continued, among others, by Bollobás and Radcliffe [6], by Frankl [8] and more recently by Kozma and Moran [11] and by Mészáros and Rónyai [12,13,15].

The Sauer–Shelah lemma [16] states that for any family $\mathcal{F} \subseteq \mathcal{P}(n)$ we have $|\mathcal{F}| \leq \binom{n}{\leq \mathrm{VC}(\mathcal{F})}$. A family is called *maximum* if $|\mathcal{F}| = \binom{n}{\leq \mathrm{VC}(\mathcal{F})}$. Maximum families serve as important examples in the theory of machine learning. They have several nice properties, among others, every maximum family is clearly extremal. Frankl [7] raised the question of estimating $m(n, k)$, the number of maximum families in $\mathcal{P}(n)$ of VC dimension k, and showed that

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

Alon, Moran and Yehudayoff [1] showed that for constant $k \geq 2$ we have, as $n \to \infty$, that

$$n^{(1+o(1))\frac{1}{k+1}\binom{n}{k}} \leq m(n, k) \leq n^{(1+o(1))\binom{n}{k}}. \tag{1.1}$$

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

A *matching* in a graph $G = (V, E)$ is a set of edges $M \subseteq E$ without any common vertices. An *induced matching* is a matching such that no endpoints of two edges of M are joined by an edge of G. For an integer $k \geq 0$ let $\binom{[n]}{k}$ be the collection of those vertices of Q_n which contain precisely k ones. We will refer to these collections for different values of k as the layers of the hypercube Q_n. Let further $\mathrm{IndMat}(n, k)$ be the number of induced matchings in Q_n between the layers $\binom{[n]}{k}$ and $\binom{[n]}{k+1}$. Our next result concerns the quantities $\mathrm{IndMat}(n, k)$ and $\mathrm{ExVC}(n, k)$. We show that the asymptotics of the logarithm of these two quantities and of $m(n, k)$ have the same value, and that these three functions follow a hierarchy:

Theorem 1.2. *Let $k = n^{o(1)}$. Then*

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

1.2. The integrity of Q_n

Next we consider the problem of finding a small set $S \subseteq \{0, 1\}^n$ such that all connected components of $Q_n \setminus S$ 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

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

In [9] 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 exist 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. They picked $\log_2 n$ points $x_1, x_2, \ldots, x_{\log_2 n} \in \{0, 1\}^n$ at pairwise distance $n/2$ and defined the set S to consist of all points that have distance exactly $n/2$ to at least one x_i. They suspected this natural construction to be of the correct order of magnitude. We show that the true value of $I(Q_n)$ lies below their upper bound. Instead of orthogonal cuts given by balls of radius $n/2$ around some carefully chosen points, we consider balls of radius slightly smaller than $n/2$ around a randomly chosen collection of points.

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

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

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.

Please cite this article in press as: J. Balogh, et al., Two results about the hypercube, Discrete Applied Mathematics (2018), https://doi.org/10.1016/j.dam.2018.03.086.