On extremal problems concerning the traces of sets

Simón Piga[1] and Bjarne Schülke[2]

[1] Fachbereich Mathematik, Universität Hamburg, Germany
simon.piga@uni-hamburg.de
[2] Fachbereich Mathematik, Universität Hamburg, Germany
bjarne.schuelke@uni-hamburg.de

Abstract. Given two non-negative integers n and s, define $m(n, s)$ to be the maximal number m such that every hypergraph $\mathcal{H}$ on n vertices and with at most m edges has a vertex x such that $|\mathcal{H}_x| \geqslant |E(\mathcal{H})| - s$, where $\mathcal{H}_x = \{H \backslash \{x\} : H \in E(\mathcal{H})\}$. The problem of determining the limit $m(s) = \lim_{n\to\infty} \frac{m(n,s)}{n}$ was posed by Füredi and Pach and by Frankl and Tokushige. While the first results were only for specific small values of s, Frankl determined $m(2^{d-1} - 1)$ for all $d \in \mathbb{N}$. Here we prove that $m(2^{d-1} - c) = \frac{(2^d-c)}{d}$ for every $c, d \in \mathbb{N}$ with $d \geqslant 4c$ and give an example showing that this equality does not hold anymore for $d = c$. The other line of research on this problem is to determine $m(s)$ for small values of s. In this line, our second result determines $m(2^{d-1} - c)$ for $c \in \{3, 4\}$. This solves more instances of the problem for small s and in particular solves a conjecture by Frankl and Watanabe.

Keywords: Extremal set theory, traces of sets, abstract simplicial complexes

1 Introduction

A hypergraph $\mathcal{H}$ is a pair $(V, \mathcal{F})$ where V is the set of vertices and $\mathcal{F} \subseteq 2^V$ is the set of edges. In the literature, the problems we consider in this work are often presented in the context of families rather than hypergraphs. If not necessary, it is then not distinguished between the family $\mathcal{F} \subseteq 2^V$ and the hypergraph $(V, \mathcal{F})$. We will follow this notational path.

Let V be an n-element set and let $\mathcal{F}$ be a family of subsets of V. For a subset T of V, define the *trace* of $\mathcal{F}$ on T by $\mathcal{F}_{|T} = \{F \cap T : F \in \mathcal{F}\}$. For integers n, m, a, and b, we write

$$(n, m) \to (a, b)$$

if for every family $\mathcal{F} \subseteq 2^V$ with $|\mathcal{F}| \geqslant m$ and $|V| = n$, there is an a-element set $T \subseteq V$ such that $|\mathcal{F}_{|T}| \geqslant b$ (we also say that (n, m) *arrows* (a, b)).

In this context, Füredi and Pach [5] and, more recently, Frankl and Tokushige [3] posed the following problem[3]:

[3] There have been slightly different versions in use for the arrowing notation and for what we denote by $m(n, s)$. In this work, we follow the notation in [3].

Problem 1. Given non-negative integers n and s, what is the maximum value $m(n,s)$ such that for every $m \leqslant m(n,s)$, we have

$$(n,m) \to (n-1, m-s).$$

As described in the short abstract, this problem can also be formulated as finding the maximal number $m(n,s)$ such that the following holds. In every hypergraph $\mathcal{H}$ with some n-set V as vertex set and with at most $m(n,s)$ edges, there is a vertex x such that $|\mathcal{H}_x| \geqslant |\mathcal{H}| - s$, where $\mathcal{H}_x = \mathcal{H}_{|V\setminus\{x\}} = \{H\setminus\{x\} : H \in \mathcal{H}\}$.

A family $\mathcal{F}$ is *hereditary* if for every $F' \subseteq F \in \mathcal{F}$, we have that $F' \in \mathcal{F}$. In [2], Frankl proves that among families with a fixed number of edges and vertices, the trace is minimised by hereditary families. Thus, the problems considered here, and in particular Problem 1, can be reduced to hereditary families. Note that in hereditary families, Problem 1 is asking for the maximum number of edges such that there is always a vertex of small degree (as usual, we define the degree of a vertex v as the number of edges that contain v).

The investigation of this problem started with Bondy [1] and Bollobás [7] determining $m(n,0)$ and $m(n,1)$, respectively. Later Frankl [2] and Frankl and Watanabe [4] proved the following identities

$$m(n, 2^{d-1}-1) = \frac{n}{d}(2^d - 1) \quad \text{and} \quad m(n, 2^{d-1}-2) = \frac{n}{d}(2^d-2) \tag{1}$$

for $d, n \in \mathbb{N}$ and $d \mid n$.

Consider a family consisting of a set of size d and all possible subsets, and take n/d vertex disjoint copies of it. The resulting family has minimum degree 2^{d-1} and $\frac{n}{d}(2^d-1)+1$ edges. Thus, this family is an extremal construction for the first identity of (1). By taking out all sets of size d, we obtain an extremal construction for the second one.

More generally, for an integer $c \geqslant 1$, if we arbitrarily take out $(c-1)$ sets in each of those d-sets, then the minimum degree is at least $2^{d-1}-c+1$ and the number of edges is $\frac{n}{d}(2^d-c)+1$. More precisely for an arbitrary family $\mathcal{A} \subseteq 2^{[d]}$ of size $(c-1)$ we consider the family

$$\mathcal{F}_c(\mathcal{A}) = \left\{F + (i-1)d : F \in 2^{[d]}\setminus\mathcal{A} \text{ and } i \in \left[\frac{n}{d}\right]\right\} \subseteq 2^{[n]}.$$

These families show that $m(n, 2^{d-1}-c) \leqslant \frac{n}{d}(2^d-c)$. The following theorem says that in fact we have equality as long as $c \leqslant \frac{d}{4}$.

Theorem 1 (Main theorem). *Let $d, c, n \in \mathbb{N}$ with $d \geqslant 4c$ and $d|n$. Then*

$$m(n, 2^{d-1}-c) = \frac{n}{d}(2^d - c).$$

Remark 1. In fact, our proof yields that for $d \geqslant 4c$ and $m \leqslant \frac{n}{d}(2^d-c)$, we have $(n,m) \to (n-1, m-(2^{d-1}-c))$ without any divisibility conditions on n. The assumption $d|n$ is only necessary for the extremal constructions showing the maximality of $\frac{n}{d}(2^d-c)$. Analogous remarks hold for the identities in (1) above and Theorem 2 below.

One might also try to solve Problem 1 for small values of s. Apart from the aforementioned results by Bondy and Bollobás, progress was made by Frankl [2], Watanabe [11, 12], and by Frankl and Watanabe [4]. In [4], they conjectured that $m(n, 12) = (28/5 + o(1))n$. Theorem 1 does not consider cases for which d is very small in terms of c. The following results extend the identities in (1) for $c = 3$ and $c = 4$ and every $d \geqslant 5$ (for smaller d the respective $m(n, s)$ is not defined or has been determined previously). In particular, it proves the conjecture of Frankl and Watanabe for $s = 12$ in a strong sense.

Theorem 2. *Let $d, n \in \mathbb{N}$ with $d \geqslant 5$ and $d|n$. Then*

1. $m(n, 2^{d-1} - 3) = \frac{n}{d}(2^d - 3)$ *and*
2. $m(n, 2^{d-1} - 4) = \frac{n}{d}(2^d - 4)$. *In particular,* $m(n, 12) = \frac{28}{5}n$.

Note that for larger d, this theorem is of course a special case of Theorem 1.

2 Idea of the proof

Here we present a sketch of the proof. For the complete proof we refer the reader to [8].

We need to show that for every hereditary hypergraph $\mathcal{F}$ on n vertices with minimum degree at least $2^{d-1} - c + 1$, we have that

$$|\mathcal{F}| \geqslant \frac{n}{d}(2^d - c) + 1.$$

In the proofs of the identities in (1) in [2,4], they observe that by double counting we have $|\mathcal{F}\backslash\{\varnothing\}| = \sum_{v \in V} \sum_{H \in L_v} \frac{1}{|H|+1}$, where $L_v = \{A \subseteq V : A \cup \{v\} \in \mathcal{F}\}$ is the *link of the vertex* v. Subsequently, they used a generalised form of the Kruskal-Katona Theorem to obtain a lower bound for $\sum_{H \in L_v} \frac{1}{|H|+1}$ which is the same for every vertex v. Due to the aforementioned double counting this in turn yields the lower bound on the number of edges.

For $c \geqslant 3$, there are extremal families which show that a general bound on $\sum_{H \in L_v} \frac{1}{|H|+1}$ for every vertex v is not sufficient to provide the desired bound on the number of edges. To overcome this difficulty, first observe that the double counting argument can be generalised by interpreting $\sum_{H \in L_v} \frac{1}{|H|+1}$ as the weight $w_{\mathcal{F}}(v)$ of a vertex v. We will refer to this weight as *uniform weight* since it can be imagined as uniformly distributing the unit weight of an edge to each of its vertices. In contrast, to prove Theorem 1 and Theorem 2, we will use non-uniform weights. Moreover, instead of bounding the weight of single vertices we will bound the weight of sets of vertices.

To this end, take a maximal set $\mathcal{L}$ of "light" vertices with neighbourhoods[4] of size at most $d - 1$ such that the neighbourhoods of all vertices in $\mathcal{L}$ are pairwise disjoint. For all $v \in \mathcal{L}$, we call the set $V_v = N(v) \cup \{v\}$ *cluster*. Observe that if the size of the neighbourhood of a vertex is at most $d - 1$, then it has to intersect

[4] For $v \in V$, the neighbourhood of v is $N(v) = \{w \in V\backslash\{v\} : \exists A \in \mathcal{F} : \{v, w\} \subseteq A\}$

one of the clusters. For vertices whose neighbourhood does not intersect any cluster (and which therefore have a neighbourhood of size at least d), we use the uniform weight. To bound these uniform weights, we introduce a “local” lemma which is a close relative to a general form of the Kruskal-Katona theorem. Given a vertex of fixed degree, it provides a lower bound on the uniform weight and furthermore the minimum weight surplus if its link deviates enough from the minimising link. Since the link of every vertex whose neighbourhood does not intersect any cluster indeed deviates enough from the minimising link (because their neighbourhood contains at least d vertices), the lemma then gives that these vertices will have a large uniform weight.

The next step is to bound the weight of vertices in the clusters. The difficulty is that the weights of different vertices in a cluster might vary. Here, the first key idea is used. Instead of bounding the weight of each single vertex, we bound the average weight of the vertices in a cluster. Even if the number of edges inside a cluster is not large enough, $\mathcal{F}$ being hereditary and the minimum degree of $\mathcal{F}$ still provide some lower bound for the number of edges in each cluster. Then a second local lemma yields that there are several vertices within that cluster whose degree with respect to the cluster is not the minimum degree in $\mathcal{F}$. Therefore, there exist several crossing edges, i.e., edges containing vertices from both the inside and the outside of the cluster. If we use the uniform weight, these crossing edges will contribute enough to the weight of the cluster, even more than needed.

At this point, we still need to bound the weight of vertices with neighbourhoods of size at most $d-1$ lying outside of any cluster. As mentioned above, the neighbourhood of every such vertex intersects some cluster, meaning every such vertex is contained in a crossing edge. Recall that in fact, a uniform weight on crossing edges would contribute more weight than needed for the inside of a cluster. Now the second idea comes into play: the unit weight of these edges will be distributed non-uniformly among its vertices. Hence, when splitting the unit weight of such a crossing edge according to the aforementioned imbalance, both sides will get a share that is big enough.

We remark that this strategy is in some sense compatible with the extremal constructions in so far as that those are composed of disjoint copies of almost complete families on d vertices (corresponding to the clusters in the proof).

3 Further Remarks and Open Problems

As in the abstract, consider $m(s)$ to be the following limit

$$m(s) := \lim_{n\to\infty} \frac{m(n,s)}{n}\,.$$

It is not difficult to check that $m(s)$ is well-defined (see [4]). Rephrased by means of this definition, Theorem 1 implies that for $c \leqslant d/4$, we have that $m(2^{d-1}-c) = \frac{2^d-c}{d}$. Further, given $d \geqslant 1$, define $c_\star(d)$ to be the maximum integer such that for every $c \leqslant c_\star(d)$,

$$m(2^{d-1} - c) = \frac{2^d - c}{d}\,. \tag{2}$$

In view of Theorem 1 we have that $c_\star(d) \geqslant \lfloor \frac{d}{4} \rfloor$. The following construction shows that for $d \geqslant 5$, $c_\star(d) < d$.

Construction 1 Let k be a positive integer and set $n = 2dk$. Take V to be a set of n vertices. Consider $U_1, \ldots, U_{2k}$ to be a partition of V into sets of size d, and for every set U_i, arbitrarily pick a vertex $x_i \in U_i$. Define

$$\begin{aligned}
\mathcal{G} &= \{S \subseteq V \colon \exists i \in [2k] \text{ with } S \subseteq U_i \text{ and } |S| \leqslant d-2\}\,,\\
\mathcal{H} &= \{U_i \backslash \{x_i\} \colon i \in [2k]\}\,, \text{ and}\\
\mathcal{I} &= \{\{x_i, x_{i+1}\} \colon i \in \{1, 3, 5, \ldots, 2k-1\}\}\,.
\end{aligned}$$

One can check that the number of edges of the family $\mathcal{F} = \mathcal{G} \cup \mathcal{H} \cup \mathcal{I}$ is given by

$$|\mathcal{G}| + |\mathcal{H}| + |\mathcal{I}| = \frac{2^d - d - 2}{d} n + 1 + \frac{n}{d} + \frac{n}{2d} = \frac{2^d - d - \frac{1}{2}}{d} n + 1.$$

Finally, since every vertex in V has degree $s = 2^{d-1} - d + 1$, we obtain

$$m(n, 2^{d-1} - d) \leqslant \frac{n}{d}\left(2^d - d - \frac{1}{2}\right) < \frac{n}{d}\left(2^d - d\right),$$

and so $c_\star(d) < d$ follows.

It would be interesting to understand the behaviour of $m(2^{d-1} - c)$ for $c > d$. To this end, we suggest the following three problems.

Problem 2. Given $\varepsilon > 0$ sufficiently small, determine $m(2^{d-1} - c)$ for all $d \in \mathbb{N}$ and $c \in \mathbb{N}$ with $d < c \leqslant (1+\varepsilon)d$.

Problem 3. Given $\varepsilon > 0$ sufficiently small, determine $m(2^{d-1} - c)$ for all $d \in \mathbb{N}$ and $c \in \mathbb{N}$ with $d < c \leqslant d^{1+\varepsilon}$.

The following problem seems very difficult, and even estimates might be interesting.

Problem 4. Given $\varepsilon > 0$ sufficiently small, determine $m(\lfloor (1-\varepsilon)2^{d-1} \rfloor)$ for all $d \in \mathbb{N}$.

Bibliography

[1] J. A. Bondy, *Induced subsets*, Journal of Combinatorial Theory, Series B **12** (1972), no. 2, 201–202. ↑2

[2] P. Frankl, *On the trace of finite sets*, Journal of Combinatorial Theory, Series A **34** (1983), no. 1, 41–45. ↑2, 3

[3] P. Frankl and N. Tokushige, *Extremal problems for finite sets*, Vol. 86, American Mathematical Soc., 2018. ↑1

[4] P. Frankl and M. Watanabe, *Some best possible bounds concerning the traces of finite sets*, Graphs and Combinatorics **10** (1994), no. 2-4, 283–292. ↑2, 3, 4

[5] Z. Füredi and J. Pach, *Traces of finite sets: extremal problems and geometric applications*, Extremal problems for finite sets **3** (1991), 255–282. ↑1

[6] G. O. H. Katona, *Optimization for order ideals under a weight assignment*, Problèmes Comb, et Théorie des Graphes (1978), 257–258. ↑

[7] L. Lovász, *Combinatorial Problems and Exercises. 1979*, North-Holland, Amsterdam. ↑2

[8] S. Piga and B. Schülke, *On extremal problems concerning the traces of sets*, Journal of Combinatorial Theory, Series A **182** (2021), 105447. ↑3

[9] S. Shelah, *A combinatorial problem; stability and order for models and theories in infinitary languages*, Pacific Journal of Mathematics **41** (1972), no. 1, 247–261. ↑

[10] V. N. Vapnik and A. Y. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Measures of complexity, 2015, pp. 11–30. ↑

[11] M. Watanabe, *Some best possible bounds concerning the traces of finite sets II*, Graphs and Combinatorics **11** (1995), no. 3, 293–303. ↑3

[12] ———, *Arrow relations on families of finite sets*, Discrete mathematics **94** (1991), no. 1, 53–64. ↑3