Set systems with union and intersection constraints

Dhruv Mubayi * Reshma Ramadurai †

October 15, 2008

Abstract

Let $2 \le d \le k$ be fixed and n be sufficiently large. Suppose that $\mathcal{G}$ is a collection of k-element subsets of an n-element set, and $|\mathcal{G}| > \binom{n-1}{k-1}$. Then $\mathcal{G}$ contains d sets with union of size at most $2k$ and empty intersection. This extends the Erdős-Ko-Rado theorem and verifies a conjecture of the first author for large n.

1 Introduction

A d-cluster of k-element sets (henceforth k-sets) is a collection of d sets with union of size at most $2k$ and empty intersection. The seminal Erdős-Ko-Rado theorem [3] states that the maximum size of a family of k-sets of $[n] = \{1, \ldots, n\}$ which contains no 2-cluster is $\binom{n-1}{k-1}$ (note that a 2-cluster comprises two disjoint sets). Katona asked the corresponding question when $d = 3$. Frankl and Füredi [4] showed that the answer is again $\binom{n-1}{k-1}$ as long as n is sufficiently large, and conjectured that this holds for all $n \ge 3k/2$. The first author [8] recently proved their conjecture, and generalized it still further by introducing the concept of a d-cluster (the word d-cluster to describe this particular set of configurations is due to Chen, Liu and Wang [1]). A star is a collection of sets that all contain a fixed element.

Conjecture 1. **([8])** *Let $2 \le d \le k$ and $n \ge kd/(d-1)$. Suppose that $\mathcal{G}$ is a collection of k-sets of $[n]$ containing no d-cluster. Then $|\mathcal{G}| \le \binom{n-1}{k-1}$. Moreover, if $d \ge 3$ and equality holds, then $\mathcal{G}$ is a star.*

The first author [7] recently proved that for fixed $2 \le d \le k$ we have $|\mathcal{G}| \le (1 + o(1))\binom{n-1}{k-1}$ as $n \to \infty$. Regarding exact results we have already observed that Conjecture 1 holds for $d = 2$ and $d = 3$. The only other known case for Conjecture 1 is when $d = k$, where it follows from an old result of Chvátal [2] (this was recently observed in [1]).

There has been further progress when one or more of the parameters is large. In [7], it is proved that Conjecture 1 holds for $d = 4$ and large n, while Keevash and the first author [6] recently proved Conjecture 1 in a different range of n, namely when k/n and $n/2 - k$ are both bounded away from zero. This includes the case $n = ck$ where c is a fixed constant greater than 2.

In this paper we provide further evidence for Conjecture 1 by proving it for all $2 \le d \le k$ as long as n is sufficiently large.

*Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, Illinois 60607, USA; research partially supported by NSF grant DMS 0653946 E-mail: mubayi@math.uic.edu

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213-3890

Theorem 1. *Fix $2 \le d \le k$ and let n be sufficiently large. Suppose that $\mathcal{G}$ is a family of k-sets of $[n]$ that contains no d-cluster. Then $|\mathcal{G}| \le \binom{n-1}{k-1}$, and equality holds if and only if $\mathcal{G}$ is a star.*

Our proof of Theorem 1 is based on the stability approach pioneered by Erdős and Simonovits (see [9]). In [7], this method is used to prove the case $d = 4$ and here we add some new ideas (see Section 3) to extend those arguments. Recently Füredi and Ozkahya [5] have also proved Theorem 1. Their proof uses the delta system method, which is a completely different approach.

2 Preliminaries

Suppose $\mathcal{G}$ is a collection of subsets of $[n]$ and $x \in [n]$. The degree $d_{\mathcal{G}}(x)$ is the number of sets of $\mathcal{G}$ that contain x. The sets $A \subset [n] - \{x\}$ with $A \cup \{x\} \in \mathcal{G}$ fall into two families: $L_x(\mathcal{G})$ consists of those A for which there is some $y \ne x$ for which $A \cup \{y\}$ is also in $\mathcal{G}$; $S_x(\mathcal{G})$ consists of those A for which $A \cup \{y\} \in \mathcal{G}$ implies that $y = x$. Note that $d_{\mathcal{G}}(x) = |L_x(\mathcal{G})| + |S_x(\mathcal{G})|$.

We need the following lemma proved in [7] (see also [6]). We will present the short proof for completeness.

Lemma 1. *Suppose $n > k \ge d \ge 2$, $\mathcal{G}$ is a collection of k-sets of $[n]$ and $x \in [n]$. If $L_x(\mathcal{G})$ contains a $(d-1)$-cluster then $\mathcal{G}$ contains a d-cluster.*

Proof. Suppose that $L_x(\mathcal{G})$ contains the $(d-1)$-cluster $A_1, \cdots, A_{d-1}$. There exists $y \ne x$ such that $B_d = A_1 \cup \{y\} \in \mathcal{G}$. Let $B_i = A_i \cup \{x\}$ for $i \in [d-1]$. Since $A_1, \ldots, A_{d-1}$ forms a $(d-1)$-cluster, $\cap_{i=1}^{d-1} A_i = \emptyset$, and so $\cap_{i=1}^{d-1} B_i = \{x\}$. As $x \notin B_d$, we conclude that $\cap_{i=1}^{d} B_i = \emptyset$. Also, $| \cup_{i=1}^{d} B_i| \le | \cup_{i=1}^{d-1} A_i| + |\{x, y\}| \le 2(k-1) + 2 = 2k$. Consequently, $B_1, \cdots, B_d$ is a d-cluster in $\mathcal{G}$. □

The other crucial tool is the following stability result proved in [7].

Theorem 2. (Stability) *Fix $2 \le d \le k$. For every $\epsilon > 0$, there exists $\delta > 0$ and n_0 such that the following holds for all $n > n_0$: Suppose that $\mathcal{G}$ is a collection of k-sets of $[n]$ containing no d-cluster. If $|\mathcal{G}| \ge (1-\delta)\binom{n-1}{k-1}$, then there exists an $x \in [n]$ such that the number of k-sets omitting x is at most $\epsilon\binom{n-1}{k-1}$. In particular, this implies that $|\mathcal{G}| \le 2\binom{n-1}{k-1}$ for sufficiently large n.*

3 A bound for bipartite families

In order to prove the main result in the next section, we need some estimates on various subfamilies with a certain bipartite structure. The crucial lemma below provides this. Although we need it only for $p \ge 3$, we will prove it for $p \ge 1$ in order to facilitate an induction argument. This was pointed out to us by a referee.

Lemma 2. *Fix $2 \le d \le k$, $1 \le p \le k$, and $k < b \le n/2$ with n sufficiently large. Suppose that $[n]$ has partition $B \cup C$, $b = |B|, c = |C|$ and $\mathcal{F}$ is a collection of k-sets of $[n]$ such that $|A \cap B| = p$ for every $A \in \mathcal{F}$. If $\mathcal{F}$ contains no d-cluster, then $|\mathcal{F}| \le kb^{p-1}c^{k-p}$.*

Proof. We proceed by induction on d. First suppose that $d = 2$, so $\mathcal{F}$ is an intersecting family. Let $S \in \mathcal{F}$. Then every set in $\mathcal{F}$ has a point in $S \cap B$ or a point in $S \cap C$. Consequently, $|\mathcal{F}| \le pb^{p-1}c^{k-p} + (k-p)b^p c^{k-p-1}$ if $p < k$ and $|\mathcal{F}| \le kb^{k-1}$ if $p = k$. Since $b \le c$, in either case we obtain $|\mathcal{F}| \le kb^{p-1}c^{k-p}$ as desired.

For the induction step, assume that $d \ge 3$. Suppose for a contradiction, that $|\mathcal{F}| > kb^{p-1}c^{k-p}$. Then

$$kb^{p-1}c^{k-p} < |\mathcal{F}| \le \sum_{x \in B} d_{\mathcal{F}}(x) = \sum_{x \in B} |L_x(\mathcal{F})| + \sum_{x \in B} |S_x(\mathcal{F})|.$$

A typical set in $S_x(\mathcal{F})$ has $p-1$ points in B and $k-p$ points in C, and is not counted by any other $S_y(\mathcal{F})$ with $y \ne x$. Therefore $\sum_{x \in B} |S_x(\mathcal{F})| \le b^{p-1}c^{k-p}$ and we have

$$\sum_{x \in B} |L_x(\mathcal{F})| > (k-1)b^{p-1}c^{k-p}.$$

First suppose that $p = 1$. Then we have $\sum_{x \in B} |L_x(\mathcal{F})| > (k-1)c^{k-1}$ and so there exists $w \in B$ for which

$$|L_w(\mathcal{F})| > (k-1)\frac{c^{k-1}}{b} \ge (k-1)c^{k-2} > 2\binom{c-1}{k-2}.$$

Now Theorem 2 (applied with $2 \le d-1 \le k-1$) implies that $L_w(\mathcal{F})$ contains a $(d-1)$-cluster, since we have assumed that $c \ge n/2$ is sufficiently large. Lemma 1 then gives that $\mathcal{F}$ contains a d-cluster which is a contradiction.

Now suppose that $p \ge 2$. Again we find $w \in B$ for which $|L_w(\mathcal{F})| > (k-1)b^{p-2}c^{k-p}$. By Lemma 1, $L_w(\mathcal{F})$ contains no $(d-1)$-cluster, so by the induction hypothesis (replacing d with $d-1$, p with $p-1$, k with $k-1$, b with $b-1$, and n with $n-1$), we obtain the contradiction $|L_w(\mathcal{F})| \le (k-1)b^{p-2}c^{k-p}$. □

4 Proof of Theorem 1

In this section we complete the proof of Theorem 1. At one point the argument is identical to that in [7], and we refer the reader there for the details.

Proof of Theorem 1. We assume that $k \ge d \ge 3$ since $d = 2$ follows from the Erdős-Ko-Rado Theorem. Choose n sufficiently large that all statements in the following proof requiring this hold.

Suppose that $\mathcal{G}$ is a collection of k-sets of $[n]$ containing no d-cluster with $|\mathcal{G}| = \binom{n-1}{k-1}$. We will show that $\mathcal{G}$ is a star. Since a star is a maximal family with no d-cluster, this proves the required bound on $|\mathcal{G}|$, with the characterization of equality as well. Let $\mathcal{G} - x$ be the collection of sets in $\mathcal{G}$ that omit x. By Theorem 2, there exists $x \in [n]$ such that $m := |\mathcal{G} - x| < \epsilon\binom{n-1}{k-1}$ with $\epsilon < (12k^2)^{-k}$. If $m = 0$, then $\mathcal{G}$ is a star and we are done, hence we may assume that $m > 0$. Let

$$\mathcal{G}_x = \{E \subset [n] : |E| = k-1 \text{ and } E \cup \{x\} \in \mathcal{G}\}.$$

Claim 1. There are pairwise disjoint $(k-2)$-sets $S_1, S_2, S_3 \subset [n] - \{x\}$ such that for each i

$$d_{\mathcal{G}_x}(S_i) = |\{y \in [n] : S_i \cup \{x, y\} \in \mathcal{G}\}| \ge n - k + 1 - \frac{2km}{\binom{n-1}{k-2}}.$$

Proof. See the corresponding claim in [7]. $\square$

By Claim 1, for each i,

$$|\{y \in [n] : S_i \cup \{x, y\} \notin \mathcal{G}\}| < k + \frac{2km}{\binom{n-1}{k-2}}.$$

Let

$$B = \{y \in [n] : S_i \cup \{x, y\} \notin \mathcal{G} \text{ for some } i\}.$$

Then $|B| < 3k + 6km/\binom{n-1}{k-2}$. By adding points arbitrarily to B, we may assume that $|B| = 3k + \lfloor 6km/\binom{n-1}{k-2} \rfloor$. Since $m \geq 1$, we may suppose that there exists $S \in \mathcal{G} - x$. For each choice of a $(k-2)$-set $S' \subset [n] - \{x\} - S$ one of the k-sets $S' \cup \{x, y\}$ where $y \in S$ must be absent from $\mathcal{G}$, otherwise we obtain a d-cluster using $d-1$ of these sets and S. This immediately yields $m \geq \binom{n-k-1}{k-2} > \frac{1}{2}\binom{n-1}{k-2}$. Consequently,

$$|B| = 3k + \left\lfloor \frac{6km}{\binom{n-1}{k-2}} \right\rfloor < \frac{12km}{\binom{n-1}{k-2}} < \frac{12k\epsilon\binom{n-1}{k-1}}{\binom{n-1}{k-2}} < 12k\epsilon n < \frac{n-1}{2}.$$

Now define, for each $i \in \{0, \ldots, k\}$,

$$\mathcal{T}_i = \{T \in \mathcal{G} - x : |T \cap B| = i\}.$$

Note that $\mathcal{T}_0 \cup \cdots \cup \mathcal{T}_k$ is a partition of $\mathcal{G} - x$. First we show that $\mathcal{T}_0 = \mathcal{T}_1 = \mathcal{T}_2 = \emptyset$. Observe that $S_i \subset B$ for $i = 1, 2, 3$. If $S \in \mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2$, then there is an i for which $S_i \cap S = \emptyset$. Choose $d - 2 \leq k - 2$ elements $y_1, \ldots, y_{d-2} \in S - B$ and $y \in [n] - (B \cup \{x\} \cup S)$. Now the $d-2$ sets $S_i \cup \{x, y_j\}$ (for all j) together with S and $S_i \cup \{x, y\}$ form a d-cluster in $\mathcal{G}$, which is a contradiction. Therefore, $\mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2 = \emptyset$.

Since $\mathcal{G} - x = \cup_{i=3}^{k} \mathcal{T}_i$, we may assume that $|\mathcal{T}_p| \geq m/(k-2)$ for some $3 \leq p \leq k$. Applying Lemma 2 with $C = [n] - \{x\} - B$ (and $b = |B|, c = |C|$; noting that $k < b < (n-1)/2$), we obtain

$$\frac{m}{k-2} \leq |\mathcal{T}_p| \leq kb^{p-1}c^{k-p} < k\left(\frac{12km}{\binom{n-1}{k-2}}\right)^{p-1} n^{k-p}.$$

Simplifying, we obtain

$$m^{p-2} > \binom{n-1}{k-2}^{p-1} \frac{1}{n^{k-p}k(k-2)(12k)^{p-1}}.$$

Then, since $m < \epsilon\binom{n-1}{k-1} < \epsilon n\binom{n-1}{k-2}$, we have

$$\epsilon \geq \epsilon^{p-2} > \frac{\binom{n-1}{k-2}}{n^{k-2}k(k-2)(12k)^{p-1}} \geq \frac{((n-k)/n)^{k-2}}{(k-2)!k(k-2)(12k)^{p-1}} > (12k^2)^{-k}$$

for sufficiently large n. This contradicts the choice of ϵ and completes the proof. $\square$

5 Acknowledgments

The authors thank the referees for their careful reading of the paper and for pointing out an error in the original proof of Lemma 2.

References

[1] W. Chen, J. Liu, L. Wang, Families of Sets with Intersecting Clusters, submitted

[2] V. Chvátal, An extremal set-intersection theorem, *J. London Math. Soc.* **9** (1974/75), 355–359.

[3] P. Erdős, H. Ko, R. Rado, Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford Ser.* **12** (1961) 313–320.

[4] P. Frankl, Z. Füredi, A new generalization of the Erdős-Ko-Rado theorem, *Combinatorica* **3** (1983), 341–349.

[5] Z. Füredi, L. Ozkahya, An intersection theorem with small unions, submitted

[6] P. Keevash, D. Mubayi, Set systems without a simplex or a cluster, *Combinatorica*, to appear

[7] D. Mubayi, An intersection theorem for four sets, *Adv. Math.* 215 (2007), no. 2, 601–615.

[8] D. Mubayi, Erdős-Ko-Rado for three sets, *J. Combin. Theory Ser. A*, **113** (2006), 547–550.

[9] M. Simonovits, A method for solving extremal problems in graph theory, stability problems. 1968 Theory of Graphs (Proc. Colloq., Tihany, 1966) pp. 279–319, Academic Press, New York.