Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in DISCRETE MATHEMATICS

www.elsevier.com/locate/endm

A simple removal lemma for large nearly-intersecting families

Tuan Tran [1] Shagnik Das [2]

Institut für Mathematik, Freie Universität Berlin, Germany

Abstract

A k-uniform family of subsets of $[n]$ is *intersecting* if it does not contain a disjoint pair of sets. The study of intersecting families is central to extremal set theory, dating back to the seminal Erdős–Ko–Rado theorem of 1961 that bounds the largest such families. A recent trend has been to investigate the structure of set families with few disjoint pairs.

Friedgut and Regev proved a general removal lemma, showing that when $\gamma n \leq k \leq (\frac{1}{2} - \gamma)n$, a set family with few disjoint pairs can be made intersecting by removing few sets. Our main contribution in this paper is to provide a simple proof of a special case of this theorem, when the family has size close to $\binom{n-1}{k-1}$. However, our theorem holds for all $2 \leq k < \frac{1}{2}n$ and provides sharp quantitative estimates.

We then use this removal lemma to settle a question of Bollobás, Narayanan and Raigorodskii regarding the independence number of random subgraphs of the Kneser graph $K(n,k)$. The Erdős–Ko–Rado theorem shows $\alpha(K(n,k)) = \binom{n-1}{k-1}$. For some constant $c > 0$ and $k \leq cn$, we determine the sharp threshold for when this equality holds for random subgraphs of $K(n,k)$, and provide strong bounds on the critical probability for $k \leq \frac{1}{2}(n-3)$.

Keywords: Erdős-Ko-Rado theorem, intersecting families, removal lemma

[1] Email: `tuan@math.fu-berlin.de`

[2] Email: `shagnik@mi.fu-berlin.de`

1 Introduction

Extremal set theory, with its many connections and applications to other areas, has experienced remarkable growth in recent decades. Inspired by one of the cornerstones of the field, the celebrated Erdős–Ko–Rado theorem of 1961, a great deal of research concerns intersecting families. We say a family of sets is *intersecting* if it does not contain a pair of disjoint sets. In this paper we derive a stability result for large families that are nearly intersecting, and apply it to obtain a sparse extension of the Erdős–Ko–Rado theorem. We begin with a brief survey of relevant results.

1.1 Intersecting families and stability

We restrict our attention to k-uniform families of subsets of $[n]$. The natural extremal question is how large such a family can be if it is intersecting. When $n < 2k$, there are no two disjoint sets, and hence $\binom{[n]}{k}$ itself is intersecting. For $n \geq 2k$, a natural construction is to take all sets containing some fixed element $i \in [n]$. This family, called the *star with centre* i, contains $\binom{n-1}{k-1}$ sets, and Erdős, Ko and Rado [9] showed this is best possible.

Given the extremal result, great efforts have been made to better understand the general structure of large intersecting families. Hilton and Milner [17] determined the size of the largest intersecting family that is not a subset of a star, before Frankl [11] extended this to determine the size of the largest intersecting family not containing too large a star.

In the years since these initial papers appeared, a series of stability results have been obtained. Friedgut [12] and Dinur and Friedgut [8] used spectral techniques to show, provided $k \leq (\frac{1}{2} - \gamma)n$ for some $\gamma > 0$, any intersecting family of size close to $\binom{n-1}{k-1}$ is almost entirely contained in a star. Keevash and Mubayi [20] and Keevash [19] combined these methods with combinatorial arguments to provide similar results when k is close to $\frac{1}{2}n$.

However, a recent trend in extremal set theory is to go beyond the Erdős–Ko–Rado threshold and study set families that may not be intersecting, but contain few disjoint pairs. Das, Gan and Sudakov [5] studied the supersaturation problem, determining the minimum number of disjoint pairs appearing in sufficiently sparse k-uniform families. Furthermore, a probabilistic variant of this supersaturation problem was studied in [6], [18], [22] and [23].

Another direction that has been pursued has been the transferral of the Erdős–Ko–Rado theorem to the sparse random setting. This study was initiated by Balogh, Bohman and Mubayi [1], who asked when the largest intersecting subfamily of a random k-uniform hypergraph is the largest star. Progress on this problem has been made in subsequent papers by Gauy, Hàn and Oliveira [14], Balogh, Das, Delcourt, Liu and Sharifzadeh [3] and Hamm and Kahn [15,16]. An alternative version of a sparse Erdős–Ko–Rado theorem, which we shall discuss in greater detail in Section 1.3, was introduced by Bollobás, Narayanan and Raigorodskii [4].

1.2 Removal lemmas for disjoint pairs

As these new problems go beyond the Erdős–Ko–Rado threshold, we require more robust forms of stability that apply not only to intersecting families, but also to families with few disjoint pairs. This motivated the search for a *removal lemma* that would show one can remove few sets from any family with a small number of disjoint pairs to obtain an intersecting family. Friedgut and Regev [13] proved the first such removal lemma, stated below.

Theorem 1.1 (Friedgut–Regev) *Let $\gamma > 0$, and let k and n be positive integers satisfying $\gamma n \leq k \leq (\frac{1}{2} - \gamma)n$. Then for every $\varepsilon > 0$ there is a $\delta > 0$ such that any family $\mathcal{F} \subset \binom{[n]}{k}$ with at most $\delta |\mathcal{F}| \binom{n-k}{k}$ disjoint pairs can be made intersecting by removing at most $\varepsilon \binom{n-1}{k-1}$ sets from $\mathcal{F}$.*

Our main contribution is a simple proof of the following removal lemma for disjoint pairs.

Theorem 1.2 *There is an absolute constant $C > 0$ such that if k and n are positive integers satisfying $2 \leq k < \frac{1}{2}n$, and $\mathcal{F} \subset \binom{[n]}{k}$ is a family of size $|\mathcal{F}| = (1-\alpha)\binom{n-1}{k-1}$ with at most $\beta \binom{n-1}{k-1}\binom{n-k-1}{k-1}$ disjoint pairs, where $\max(|\alpha|, \beta) \leq \frac{n-2k}{(20C)^2 n}$, then there is some star $\mathcal{S}$ with $|\mathcal{F}\Delta\mathcal{S}| \leq C(\alpha + 2\beta)\frac{n}{n-2k}\binom{n-1}{k-1}$.*

Let us compare Theorems 1.1 and 1.2. The original theorem of Friedgut and Regev requires k to be linear in n but bounded away from $\frac{1}{2}n$. However, for such k and n, the theorem provides stability for families of all possible sizes. In particular, their very general theorem applies even when the closest intersecting family is not a star. In contrast, Theorem 1.2 only applies in the special case when $\mathcal{F}$ will be close to a star, which requires $|\mathcal{F}|$ to be close to $\binom{n-1}{k-1}$. However, our theorem holds for a wider range of k.

Theorem 1.2 also gives quantitative control over how close to a star $\mathcal{F}$ must be in terms of its size (parametrised by α), the number of disjoint pairs (parametrised by β), and how close k is to $\frac{1}{2}n$. By taking $\beta = 0$, we obtain a stability result for intersecting families, and the bounds sharpen those given by Keevash and Mubayi [20] and Keevash [19].

These bounds are sharp up to the constant. If k is bounded away from $\frac{1}{2}n$, then one may take a star and add $\alpha\binom{n-1}{k-1}$ sets from another star to obtain a family of size $(1+\alpha)\binom{n-1}{k-1}$ with $\alpha\binom{n-1}{k-1}\binom{n-k-1}{k-1}$ disjoint pairs that is $\alpha\binom{n-1}{k-1}$-far from a star. On the other hand, if $t = n - 2k = o(n)$, consider the anti-star $\binom{[n-1]}{k}$. This has size $\left(1+\frac{t}{k}\right)\binom{n-1}{k-1}$, contains approximately $\frac{t}{n}\binom{n-1}{k-1}\binom{n-k-1}{k-1}$ disjoint pairs, and yet is approximately $\binom{n-1}{k-1}$-far from a star.

1.3 Erdős–Ko–Rado for sparse Kneser subgraphs

We shall apply Theorem 1.2 to a problem of Bollobás, Narayanan and Raigorodskii [4] regarding an extension of the Erdős–Ko–Rado theorem to the sparse random setting. To define the problem, we first need to introduce the Kneser graph and its connection to the Erdős–Ko–Rado theorem.

Given integers $1 \leq k \leq \frac{1}{2}n$, the Kneser graph $K(n,k)$ is defined on the vertex set $V = \binom{[n]}{k}$, with two k-sets $F, G \in \binom{[n]}{k}$ adjacent in $K(n,k)$ if and only if $F \cap G = \emptyset$. Since edges of $K(n,k)$ denote disjoint pairs in $\binom{[n]}{k}$, it follows that independent sets of $K(n,k)$ correspond directly to intersecting families in $\binom{[n]}{k}$. Thus the Erdős–Ko–Rado theorem, viewed from the perspective of the Kneser graph, shows $\alpha(K(n,k)) = \binom{n-1}{k-1}$ when $n \geq 2k$.

Bollobás, Narayanan and Raigorodskii [4] transferred the Erdős–Ko–Rado theorem to the random setting by considering not the entire Kneser graph $K(n,k)$, but rather random subgraphs of it. Given some probability $0 \leq p \leq 1$, let $K_p(n,k)$ denote the subgraph of $K(n,k)$ where every edge is retained independently with probability p. As $K_p(n,k) \subseteq K(n,k)$, we clearly have $\alpha(K_p(n,k)) \geq \alpha(K(n,k)) = \binom{n-1}{k-1}$. They then asked for which p we have equality.

In their paper, they showed the Erdős–Ko–Rado theorem is surprisingly robust when $k = o(n^{1/3})$. In other words, we almost surely have $\alpha(K_p(n,k)) = \binom{n-1}{k-1}$ even for very small probabilities p (and thus very sparse subgraphs of $K(n,k)$). Furthermore, they exhibited a sharp threshold for when this sparse Erdős–Ko–Rado theorem holds. They conjectured that a similar result should hold for $k = o(n^{1/2})$, and possibly even for $k = O(n^{1-\delta})$ for any fixed $\delta > 0$. For these larger values of k, partial progress was made by Balogh, Bollobás and Narayanan [2].

By applying Theorem 1.2, we obtain sharper results for large k, as given in the theorem below.

Theorem 1.3 *There is an absolute constant $C > 0$ such that the following holds. Let k and n be integers with $1 \ll k \leq \frac{1}{2}(n-3)$, let $\varepsilon = \omega(k^{-1})$,* *and set* $p_c = \frac{\log\left(n\binom{n-1}{k}\right)}{\binom{n-k-1}{k-1}}$. *Then, as* $n \to \infty$, $\mathbb{P}\left(\alpha(K_p(n,k)) = \binom{n-1}{k-1}\right) \to 0$ *if* $p \le (1-\varepsilon)p_c$.

For $k \le \frac{n}{6C}$, $\mathbb{P}\left(\alpha(K_p(n,k)) = \binom{n-1}{k-1}\right) \to 1$ *if* $p \ge (1+\varepsilon)p_c$, *and the stars are the only maximum independent sets. For* $k \le \frac{1}{2}(n-3)$, *the same conclusion holds for* $p \ge \frac{2Cn}{n-2k}p_c$.

Theorem 1.3 exhibits a sharp threshold for $k \le \frac{n}{6C}$, thus extending the result of Bollobás et. al. [2] to k as large as linear in n. Furthermore, when $k \le (\frac{1}{2} - \gamma)n$, as considered in [2], $\frac{n}{n-2k} \le (2\gamma)^{-1}$, and so Theorem 1.3 determines the critical probability up to a constant factor. Finally, when k is close to $\frac{1}{2}n$, we find that the sparse version of the Erdős–Ko–Rado theorem still holds for very small edge probabilities; when $k = \frac{1}{2}(n-3)$, we almost surely have $\alpha(K_p(n,k)) = \binom{n-1}{k-1}$ even for $p = \Omega(n^{-1})$.

2 The removal lemma

In this section we shall outline a proof of our version of the removal lemma, Theorem 1.2. We first introduce the necessary terminology.

Given a family $\mathcal{F} \subset \binom{[n]}{k}$, the *characteristic function* $f : \binom{[n]}{k} \to \{0,1\}$ of $\mathcal{F}$ is a Boolean function with $f(F) = 1$ if and only if $F \in \mathcal{F}$. We may embed $\binom{[n]}{k} \subset \{0,1\}^n$ into the n-dimensional hypercube, and thus think of f as a function on the k-uniform *slice* of the cube $\{(x_1, \ldots, x_n) \in \{0,1\}^n : \sum_i x_i = k\}$. A function f is *affine* if $f(x_1, x_2, \ldots, x_n) = a_0 + \sum_{i \in [n]} a_i x_i$ for some constants a_i, $0 \le i \le n$. We equip this space with the L_2-norm with respect to the uniform measure on $\binom{[n]}{k}$, defining $\|f - g\|^2 = \frac{1}{\binom{n}{k}} \sum_{F \in \binom{[n]}{k}} |f(F) - g(F)|^2$, and say f and g are ε-*close* if $\|f-g\|^2 \le \varepsilon$.

The first step of our proof is the following lemma, which transfers the problem into the analytic framework set up above. Its proof requires information on the spectrum of the Kneser graph provided by Lovász [21].

Lemma 2.1 *Let* k *and* n *be positive integers satisfying* $2 \le k < \frac{1}{2}n$, *and let* $\mathcal{F} \subset \binom{[n]}{k}$ *be a family of size* $|\mathcal{F}| = (1-\alpha)\binom{n-1}{k-1}$ *with at most* $\beta\binom{n-1}{k-1}\binom{n-k-1}{k-1}$ *disjoint pairs. If* $f : \binom{[n]}{k} \to \{0,1\}$ *is the characteristic function of* $\mathcal{F}$, *then there is some affine function* $g : \binom{[n]}{k} \to \mathbb{R}$ *with* $\|f-g\|^2 \le (\alpha + 2\beta)\frac{k}{n-2k}$.

The above lemma shows that if a set family $\mathcal{F}$ is as in the statement of Theorem 1.2 then its characteristic function can be approximated well by an affine function, from which we shall deduce that $\mathcal{F}$ itself is close to a star. Note that the characteristic function h of the star with centre i is $h(x_1, \ldots, x_n) = x_i$, and is thus determined by a single coordinate. A result of Filmus [10] states that if a Boolean function $f : \binom{[n]}{k} \to \{0,1\}$ is close to an affine function, then it is close to a function determined by at most one coordinate. Putting everything together we obtain a proof of Theorem 1.2.

Theorem 2.2 (Filmus) *For some constant $C > 1$ the following holds. Suppose $2 \le k \le \frac{1}{2}n$ and $\varepsilon < \frac{k}{128n}$. For every Boolean function $f : \binom{[n]}{k} \to \{0,1\}$ that is ε-close to an affine function, there is some set $S \subset [n]$ of size $|S| \le \max\left(1, \frac{Cn\sqrt{\varepsilon}}{k}\right)$ such that either f or $1-f$ is $(C\varepsilon)$-close to $\max_{i\in S} x_i$.*

For the full details we refer the reader to our paper [7].

References

[1] Balogh, J., T. Bohman and D. Mubayi, *Erdős–Ko–Rado in random hypergraphs*, Combin. Probab. Comput. **18.5** (2009), 629–646.

[2] Balogh, J., B. Bollobás and B. P. Narayanan, *Transference for the Erdős–Ko–Rado theorem*, preprint.

[3] Balogh, J., S. Das, M. Delcourt, H. Liu and M. Sharifzadeh, *Intersecting families of discrete structures are typically trivial*, J. Combin. Theory Ser. A **132** (2015), 224–245.

[4] Bollobás, B., B. P. Narayanan and A. M. Raigorodskii, *On the stability of the Erdős–Ko–Rado theorem*, arXiv:1408.1288.

[5] Das, S., W. Gan and B. Sudakov, *The minimum number of disjoint pairs in set systems and related problems*, Combinatorica, to appear.

[6] Das, S., and B. Sudakov, *Most probably intersecting hypergraphs*, Electron. J. Comb. **22** (2015), P1.80.

[7] Das, S., and T. Tran, *A simple removal lemma for large nearly-intersecting families*, arXiv:1412.7885.

[8] Dinur, I., and E. Friedgut, *Intersecting families are essentially contained in juntas*, Combin. Probab. Comput. **18.1-2** (2009), 107–122.

[9] Erdős, P., C. Ko and R. Rado, *Intersection theorems for systems of finite sets*, Q. J. Math. **12.1** (1961), 313–320.

[10] Filmus, Y., *Friedgut–Kalai–Naor theorem for slices of the Boolean cube*, arXiv:1410.7834.

[11] Frankl, P., *Erdős–Ko–Rado theorem with conditions on the maximal degree*, J. Combin. Theory Ser. A **46** (1987), 252–263.

[12] Friedgut, E., *On the measure of intersecting families, uniqueness and stability*, Combinatorica **28.5** (2008), 503–528.

[13] Friedgut, E., and O. Regev, *personal communication.*

[14] Gauy, M. M, H. Hàn and I. C. Oliveira, *Erdős–Ko–Rado for random hypergraphs: asymptotics and stability*, arXiv:1409.3634.

[15] Hamm, A., and J. Kahn, *On Erdős–Ko–Rado for random hypergraphs I*, arXiv:1412.5085.

[16] Hamm, A., and J. Kahn, *On Erdős–Ko–Rado for random hypergraphs II*, arXiv:1406.5793.

[17] Hilton, A. J. W., and E. C. Milner, *Some intersection theorems for systems of finite sets*, Q. J. Math. **18.1** (1967), 369–384.

[18] Katona, G. O. H., G. Y. Katona and Z. Katona, *Most probably intersecting families of subsets*, Combin. Probab. Comput. **21.1-2** (2012), 219–227.

[19] Keevash, P., *Shadows and intersections: stability and new proofs*, Adv. Math. **218** (2008), 1685–1703.

[20] Keevash, P., and D. Mubayi, *Set systems without a simplex or a cluster*, Combinatorica **30.2** (2010), 175–200.

[21] Lovász, L., *On the Shannon capacity of a graph*, IEEE T. Inform. Theory **25.1** (1979), 1–7.

[22] Russell, P. A. *Compressions and probably intersecting families*, Combin. Probab. Comput. **21.1-2** (2012), 301–313.

[23] Russell, P. A., and M. Walters, *Probably intersecting families are not nested*, Combin. Probab. Comput. **22.1** (2013), 146–160.