Discrete Mathematics 344 (2021) 112367

Contents lists available at ScienceDirect

Discrete Mathematics

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

Note

Note on sunflowers

Tolson Bell [a,1], Suchakree Chueluecha [b,2], Lutz Warnke [a,3]

[a] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA
[b] Department of Mathematics, Lehigh University, Bethlehem, PA 18015, USA

ARTICLE INFO

Article history:
Received 17 October 2020
Accepted 23 February 2021
Available online xxxx

Keywords:
Sunflower
Set system
Sunflower conjecture
Sunflower problem
Intersection theorem

ABSTRACT

A sunflower with p petals consists of p sets whose pairwise intersections are identical. The goal of the sunflower problem is to find the smallest $r = r(p, k)$ such that any family of r^k distinct k-element sets contains a sunflower with p petals. Building upon a breakthrough of Alweiss, Lovett, Wu and Zhang from 2019, Rao proved that $r = O(p \log(pk))$ suffices; this bound was reproved by Tao in 2020. In this short note we record that $r = O(p \log k)$ suffices, by using a minor variant of the probabilistic part of these recent proofs.

1. Introduction

A sunflower with p petals is a family of p sets whose pairwise intersections are identical (the intersections may be empty). Let $\mathrm{Sun}(p, k)$ denote the smallest natural number s with the property that any family of at least s distinct k-element sets contains a sunflower with p petals. In 1960, Erdős and Rado [4] proved that $(p-1)^k < \mathrm{Sun}(p, k) \le (p-1)^k k! + 1 = O((pk)^k)$, and conjectured that for any $p \ge 2$ there is a constant $C_p > 0$ such that $\mathrm{Sun}(p, k) \le C_p^k$ for all $k \ge 2$. This famous conjecture in extremal combinatorics was one of Erdős' favorite problems [2], for which he offered a $1000 reward [3]; it remains open despite considerable attention [7].

In 2019, there was a breakthrough on the sunflower conjecture: using iterative encoding arguments, Alweiss, Lovett, Wu and Zhang [1] proved that $\mathrm{Sun}(p, k) \le (Cp^3 \log k \log\log k)^k$ for some constant $C > 0$, opening the floodgates for further improvements. Using Shannon's noiseless coding theorem, Rao [8] subsequently simplified their proof and obtained a slightly better bound. Soon thereafter, Frankston, Kahn, Narayanan and Park [5] refined some key counting arguments from [1]. Their ideas were then utilized by Rao [9] to improve the best-known sunflower bound to $\mathrm{Sun}(p, k) \le (Cp \log(pk))^k$ for some constant $C > 0$, which in 2020 was reproved by Tao in his blog [10] using Shannon entropy arguments.

The aim of this short note is to record, for the convenience of other researchers, that a minor variant of (the probabilistic part of) the arguments from [9,10] gives $\mathrm{Sun}(p, k) \le (Cp \log k)^k$ for some constant $C > 0$.

Theorem 1. *There is a constant $C \ge 4$ such that $\mathrm{Sun}(p, k) \le (Cp \log k)^k$ for all integers $p, k \ge 2$.*

Setting $r(p, k) = Cp \log k + \mathbb{1}_{\{k=1\}} p$, we shall in fact prove $\mathrm{Sun}(p, k) \le r(p, k)^k$ for all integers $p \ge 2$ and $k \ge 1$. Similarly to the strategy of [1,9,10], this upper bound follows easily by induction on $k \ge 1$ from Lemma 2, where a family $\mathcal{S}$

E-mail addresses: tbell37@gatech.edu (T. Bell), suc221@lehigh.edu (S. Chueluecha), warnke@math.gatech.edu (L. Warnke).

1 Research supported by National Science Foundation, USA Grant DMS-1851843.
2 Research supported by Georgia Institute of Technology, College of Sciences, USA.
3 Research supported by National Science Foundation, USA Grant DMS-1703516, National Science Foundation CAREER, USA grant DMS-1945481 and a Sloan Research Fellowship, USA.

https://doi.org/10.1016/j.disc.2021.112367
0012-365X/

of k-element sets is called r-spread if there are at most $r^{k-|T|}$ sets of $\mathcal{S}$ that contain any non-empty set T. (Indeed, the base case $k = 1$ is trivial due to $r(p, 1) = p$, and the induction step $k \geq 2$ uses a simple case distinction: if $\mathcal{S}$ is $r(p, k)$-spread, then Lemma 2 guarantees a sunflower with p petals; otherwise there is a non-empty set T such that more than $r(p, k)^{k-|T|} \geq r(p, k-|T|)^{k-|T|}$ sets of $\mathcal{S}$ contain T, and among this family of sets we easily find a sunflower with p petals using induction.)

Lemma 2. *There is a constant $C \geq 4$ such that, setting $r(p, k) = Cp \log k$, the following holds for all integers $p, k \geq 2$. If a family $\mathcal{S}$ with $|\mathcal{S}| \geq r(p, k)^k$ sets of size k is $r(p, k)$-spread, then $\mathcal{S}$ contains p disjoint sets.*

Inspired by [1], in [9,10] probabilistic arguments are used to deduce Lemma 2 with $r(p, k) = \Theta(p \log(pk))$ from Theorem 3, where X_δ denotes the random subset of X in which each element is included independently with probability δ.

Theorem 3 (*Main Technical Estimate of [9,10]*). *There is a constant $B \geq 1$ such that the following holds for any integer $k \geq 2$, any reals $0 < \delta, \epsilon \leq 1/2$, $r \geq B\delta^{-1} \log(k/\epsilon)$, and any family $\mathcal{S}$ of k-element subsets of a finite set X. If $\mathcal{S}$ is r-spread with $|\mathcal{S}| \geq r^k$, then $\mathbb{P}(\exists S \in \mathcal{S} : S \subseteq X_\delta) > 1 - \epsilon$.* □

The core idea of [1,9,10] is to randomly partition the set X into $V_1 \cup \cdots \cup V_p$, by independently placing each element $x \in X$ into a randomly chosen V_i. Note that the marginal distribution of each V_i equals the distribution of X_δ with $\delta = 1/p$. Invoking Theorem 3 with $\epsilon = 1/p$ and $r = B\delta^{-1} \log(k/\epsilon)$, a standard union bound argument implies that, with non-zero probability, all of the random partition-classes V_i contain a set from $\mathcal{S}$. Hence p disjoint sets $S_1, \ldots, S_p \in \mathcal{S}$ must exist, which proves Lemma 2 with $r(p, k) = Bp \log(pk)$.

We prove Lemma 2 with $r(p, k) = \Theta(p \log k)$ using a minor twist: by randomly partitioning the vertex-set into more than p classes V_i, and then using linearity of expectation (instead of a union bound).

Proof of Lemma 2. Set $C = 4B$. We randomly partition the set X into $V_1 \cup \cdots \cup V_{2p}$, by independently placing each element $x \in X$ into a randomly chosen V_i. Let I_i be the indicator random variable for the event that V_i contains a set from $\mathcal{S}$. Since V_i has the same distribution as X_δ with $\delta = 1/(2p)$, by invoking Theorem 3 with $\epsilon = 1/2$ and $r = r(p, k) = 2Bp \log(k^2) \geq B\delta^{-1} \log(k/\epsilon)$, we obtain $\mathbb{E}I_i > 1/2$. Using linearity of expectation, the expected number of partition-classes V_i with $I_i = 1$ is thus at least p. Hence there must be a partition where at least p of the V_i contain a set from $\mathcal{S}$, which gives the desired p disjoint sets $S_1, \ldots, S_p \in \mathcal{S}$. □

Generalizing this idea, Theorem 3 gives $p > \lfloor 1/\delta \rfloor (1 - \epsilon)$ disjoint sets $S_1, \ldots, S_p \in \mathcal{S}$, which in the special case $\lfloor 1/\delta \rfloor \epsilon \leq 1$ (used in [1,9,10] with $\delta = \epsilon = 1/p$) simplifies to $p \geq \lfloor 1/\delta \rfloor$.

2. Remarks

Our proof of Lemma 2 only invokes Theorem 3 with $\epsilon = 1/2$, i.e., it does not exploit the fact that Theorem 3 has an essentially optimal dependence on ϵ (see Lemma 4). In particular, this implies that we could alternatively also prove Lemma 2 and thus the $\mathrm{Sun}(p, k) \leq (Cp \log k)^k$ bound of Theorem 1 using the combinatorial arguments of Frankston, Kahn, Narayanan and Park [5] (we have verified that the proof of [5, Theorem 1.7] can be extended to yield Theorem 3 under the stronger assumption $r \geq B\delta^{-1} \max\{\log k, \log^2(1/\epsilon)\}$, say).

We close by recording that Theorem 3 is essentially best possible with respect to the r-spread assumption, which follows from the construction in [1, Section 2] that in turn builds upon [4, Theorem II].

Lemma 4. *For any reals $0 < \delta, \epsilon \leq 1/2$ and any integers $k \geq 1$, $1 \leq r \leq 0.25\delta^{-1} \log(k/\epsilon)$, there exists an r-spread family $\mathcal{S}$ of k-element subsets of $X = \{1, \ldots, rk\}$ with $|\mathcal{S}| = r^k$ and $\mathbb{P}(\exists S \in \mathcal{S} : S \subseteq X_\delta) < 1 - \epsilon$.*

Proof. We fix a partition $V_1 \cup \cdots \cup V_k$ of X into sets of equal size $|V_i| = r$, and define $\mathcal{S}$ as the family of all k-element sets containing exactly one element from each V_i. It is easy to check that $\mathcal{S}$ is r-spread, with $|\mathcal{S}| = r^k$. Focusing on the necessary event that X_δ contains at least one element from each V_i, we obtain

$$\mathbb{P}(\exists S \in \mathcal{S} : S \subseteq X_\delta) \leq \left(1 - (1-\delta)^r\right)^k \leq e^{-(1-\delta)^r k} < e^{-e^{-2\delta r} k} \leq e^{-\sqrt{\epsilon k}} < 1 - \epsilon$$

by elementary considerations (since $e^{-\sqrt{\epsilon}} < 1 - \epsilon$ due to $0 < \epsilon \leq 1/2$). □

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was conducted as part of the 2020 REU program at Georgia Institute of Technology.

Appendix. Theorem 3

Theorem 3 follows from Tao's proof of Proposition 5 in [10] (noting that any r-spread family $\mathcal{S}$ with $|\mathcal{S}| \geq r^k$ sets of size k is also r-spread in the sense of [10]). We now record that Theorem 3 also follows from Rao's proof of Lemma 4 in [9] (where the random subset of X is formally chosen in a slightly different way).

Proof of Theorem 3 based on [9]. Set $\gamma = \delta/2$ and $m = \lceil \gamma |X| \rceil$. Let X_i denote a set chosen uniformly at random from all i-element subsets of X. Since X_δ conditioned on containing exactly i elements has the same distribution as X_i, by the law of total probability and monotonicity it routinely follows that $\mathbb{P}(\exists S \in \mathcal{S} : S \subseteq X_\delta)$ is at least $\mathbb{P}(\exists S \in \mathcal{S} : S \subseteq X_m) \cdot \mathbb{P}(|X_\delta| \geq m)$. The proof of Lemma 4 in [9] shows that $\mathbb{P}(\exists S \in \mathcal{S} : S \subseteq X_m) > 1 - \epsilon^2$ whenever $r \geq \alpha\gamma^{-1} \log(k/\epsilon)$, where $\alpha > 0$ is a sufficiently large constant. Noting $|\mathcal{S}| \leq |X|^k$ we see that $|\mathcal{S}| \geq r^k$ enforces $|X| \geq r$, so standard Chernoff bounds (such as [6, Theorem 2.1]) imply that $\mathbb{P}(|X_\delta| < m) \leq \mathbb{P}(|X_\delta| \leq |X|\delta/2)$ is at most $e^{-|X|\delta/8} \leq e^{-r\delta/8} \leq \epsilon^2$ whenever $r \geq 16\delta^{-1} \log(1/\epsilon)$. This completes the proof with $B = \max\{2\alpha, 16\}$, say (since $(1-\epsilon^2)^2 \geq 1 - \epsilon$ due to $0 < \epsilon \leq 1/2$). □

References

[1] R. Alweiss, S. Lovett, K. Wu, J. Zhang, Improved bounds for the Sunflower lemma, in: Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, ACM, 2020, pp. 624–630. Extended preprint available at arXiv:1908.08483v2.
[2] P. Erdős, On the combinatorial problems which I would most like to see solved, Combinatorica 1 (1981) 25–42.
[3] P. Erdős, Some of my old and new combinatorial problems, in: Paths, Flows, and VLSI-layout (Bonn 1988), Algorithms & Combinatorics, vol. 9, Springer, Berlin, 1990, pp. 35–45.
[4] P. Erdős, R. Rado, Intersection theorems for systems of sets, J. Lond. Math. Soc. 35 (1960) 85–90.
[5] K. Frankston, J. Kahn, B. Narayanan, J. Park, Thresholds versus fractional Expectation-thresholds, 2019, arXiv:1910.13433v2, Preprint.
[6] S. Janson, T. Łuczak, A. Ruciński, Random Graphs, Wiley-Interscience, 2000.
[7] A. Kostochka, Extremal problems on Δ-systems, in: Numbers, Information and Complexity (Bielefeld 1998), Kluwer Academic Publishers, Boston, 2000, pp. 143–150.
[8] A. Rao, Coding for Sunflowers, 2019, arXiv:1909.04774v1, Preprint.
[9] A. Rao, Coding for Sunflowers, Discrete Anal. 2 (2020) 8, arXiv:1909.04774v2.
[10] T. Tao, The Sunflower lemma via Shannon entropy, 2020. Blog post. Available at https://terrytao.wordpress.com/2020/07/20/the-sunflower-lemma-via-shannon-entropy/.