Combinatorics, Probability and Computing (2023), **32**, pp. 809–818
doi:10.1017/S0963548323000202

ARTICLE

A smoother notion of spread hypergraphs

Sam Spiro

Department of Mathematics, University of California San Diego, La Jolla, CA, USA
E-mail: sspiro@ucsd.edu

(Received 4 January 2022; revised 12 November 2022; accepted 13 May 2023; first published online 8 June 2023)

Abstract

Alweiss, Lovett, Wu, and Zhang introduced q-spread hypergraphs in their breakthrough work regarding the sunflower conjecture, and since then q-spread hypergraphs have been used to give short proofs of several outstanding problems in probabilistic combinatorics. A variant of q-spread hypergraphs was implicitly used by Kahn, Narayanan, and Park to determine the threshold for when a square of a Hamiltonian cycle appears in the random graph $G_{n,p}$. In this paper, we give a common generalization of the original notion of q-spread hypergraphs and the variant used by Kahn, Narayanan, and Park.

Keywords: thresholds; hypergraphs; spread

2020 MSC Codes: Primary: 05C80

1. Introduction

This paper concerns hypergraphs, and throughout we allow our hypergraphs to have repeated edges. If A is a set of vertices of a hypergraph $\mathcal{H}$, we define the *degree of* A to be the number of edges of $\mathcal{H}$ containing A, and we denote this quantity by $d_{\mathcal{H}}(A)$, or simply by $d(A)$ if $\mathcal{H}$ is understood. We say that a hypergraph $\mathcal{H}$ is *q-spread* if it is non-empty and if $d(A) \le q^{|A|}|\mathcal{H}|$ for all sets of vertices A. A hypergraph is said to be *r-bounded* if each of its edges have size at most r and it is *r-uniform* if all of its edges have size exactly r.

The notion of q-spread hypergraphs was introduced by Alweiss, Lovett, Wu, and Zhang [2] where it was a key ingredient in their groundbreaking work which significantly improved upon the bounds on the largest size of a set system that contains no sunflower. Their method was refined by Frankston, Kahn, Narayanan, and Park [4] who proved the following.

Theorem 1.1 ([4]). *There exists an absolute constant K_0 such that the following holds. Let $\mathcal{H}$ be an r-bounded q-spread hypergraph on V. If W is a set of size $K_0(\log r)q|V|$ chosen uniformly at random from V, then W contains an edge of $\mathcal{H}$ with probability tending to 1 as r tends towards infinity.*

This theorem was used in [4] to prove a number of remarkable results. In particular it resolved a conjecture of Talagrand, and it also gave a much simpler solution to Shamir's problem, which had originally been solved by Johansson, Kahn, and Vu [6].

Kahn, Narayanan, and Park [7] used a variant of the method from [4] to show that for certain q-spread hypergraphs, the conclusion of Theorem 1.1 holds for random sets W of size only $Cq|V|$.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650112.

They used this to determine the threshold for when a square of a Hamiltonian cycle appears in the random graph $G_{n,p}$, which was a long-standing open problem.

In a talk, Narayanan asked if there was a "smoother" definition of spread hypergraphs which interpolated between q-spread hypergraphs and hypergraphs like those in [7] where the $\log r$ term of Theorem 1.1 can be dropped. The aim of this paper is to provide such a definition.

Definition 1.2. Let $0 < q \le 1$ be a real number and $r_1 > \cdots > r_\ell$ positive integers. We say that a hypergraph $\mathcal{H}$ on V is $(q; r_1, \ldots, r_\ell)$-spread if $\mathcal{H}$ is non-empty, r_1-bounded, and if for all $A \subseteq V$ with $d(A) > 0$ and $r_i \ge |A| \ge r_{i+1}$ for some $1 \le i < \ell$, we have for all $j \ge r_{i+1}$ that

$$M_j(A) := |\{S \in \mathcal{H} : |A \cap S| \ge j\}| \le q^j |\mathcal{H}|.$$

Roughly speaking, this condition says that every set A of r_i vertices intersects few edges of $\mathcal{H}$ in more than r_{i+1} vertices.

As a warm-up, we show how this definition relates to the definition of being q-spread.

Proposition 1.3. *We have the following.*

(a) *If $\mathcal{H}$ is $(q; r_1, \ldots, r_\ell, 1)$-spread, then it is q-spread.*

(b) *If $\mathcal{H}$ is q-spread and r_1-bounded, then it is $(4q; r_1, \ldots, r_\ell)$-spread for any sequence of integers r_i satisfying $r_i > r_{i+1} \ge \frac{1}{2} r_i$.*

Proof. For (a), assume $\mathcal{H}$ is $(q; r_1, \ldots, r_\ell, 1)$-spread and let $r_{\ell+1} = 1$. Let A be a set of vertices of $\mathcal{H}$. If $A = \emptyset$, then $d(A) = |\mathcal{H}| = q^{|A|}|\mathcal{H}|$, so we can assume A is non-empty. If $d(A) = 0$, then trivially $d(A) \le q^{|A|}|\mathcal{H}|$, so we can assume $d(A) > 0$. This means $|A| \le r_1$ since in particular $\mathcal{H}$ is r_1-bounded. Thus, there exists an integer $1 \le i \le \ell$ such that $r_i \ge |A| \ge r_{i+1}$, so the hypothesis that $\mathcal{H}$ is $(q; r_1, \ldots, r_\ell, 1)$-spread and $d(A) > 0$ implies

$$d(A) \le M_{|A|}(A) \le q^{|A|}|\mathcal{H}|,$$

proving that $\mathcal{H}$ is q-spread.

For (b), assume $\mathcal{H}$ is q-spread and r_1-bounded. If A is any set of vertices of $\mathcal{H}$, then for all $j \ge \frac{1}{2}|A|$ we have

$$M_j(A) \le \sum_{B \subseteq A : |B| = j} d(B) \le 2^{|A|} \cdot q^j |\mathcal{H}| \le (4q)^j |\mathcal{H}|.$$

In particular, if $r_i \ge |A| \ge r_{i+1}$, then this bound holds for any $j \ge r_{i+1}$ since $r_{i+1} \ge \frac{1}{2} r_i \ge \frac{1}{2}|A|$. We conclude that $\mathcal{H}$ is $(4q; r_1, \ldots, r_\ell)$-spread. □

We now state our main result for uniform hypergraphs, which says that a random set of size $C\ell q|V|$ will contain an edge of an r_1-uniform $(q; r_1, \ldots, r_\ell, 1)$-spread hypergraph with high probability as $C\ell$ tends towards infinity. An analogous result can be proven for non-uniform hypergraphs, but for ease of presentation we defer this result to Section 3.

Theorem 1.4. *There exists an absolute constant K_0 such that the following holds. Let $\mathcal{H}$ be an r_1-uniform $(q; r_1, \ldots, r_\ell, 1)$-spread hypergraph on V. If W is a set of size $C\ell q|V|$ chosen uniformly at random from V with $C \ge K_0$, then*

$$\mathbb{P}[W \text{ contains an edge of } \mathcal{H}] \ge 1 - \frac{K_0}{C\ell}.$$

We note that Theorem 1.4 with $\ell = \Theta(\log r)$ together with Proposition 1.3(b) implies Theorem 1.1 for uniform $\mathcal{H}$. In [7], it is implicitly proven that the hypergraph $\mathcal{H}$ encoding squares of Hamiltonian cycles is a $(2n)$-uniform $(Cn^{-1/2}; 2n, C_0 n^{1/2}, 1)$-spread hypergraph for some appropriate constants C, C_0, so the $\ell = 2$ case of Theorem 1.4 suffices to prove the main result of [7]. Thus, at least in the uniform case, Theorem 1.4 provides an interpolation between

https://doi.org/10.1017/S0963548323000202 Published online by Cambridge University Press

the results of [4, 7]. Theorem 1.4 can also be used to recover results from very recent work of Espuny Díaz and Person [3] who extended the results of [7] to other spanning subgraphs[1] of $G_{n,p}$.

2. Proof of Theorem 1.4

Our approach borrows heavily from Kahn, Narayanan, and Park [7]. We break our proof into three parts: the main reduction lemma, auxiliary lemmas to deal with some special cases, and a final subsection proving the theorem.

2.1. The main lemma

We briefly sketch our approach for proving Theorem 1.4. Let $\mathcal{H}$ be a hypergraph with vertex set V. We first choose a random set $W_1 \subseteq V$ of size roughly $q|V|$. If W_1 contains an edge of $\mathcal{H}$ then we would be done, but most likely we will need to try and add in an additional random set W_2 of size $q|V|$ and repeat the process. In total then we are interested in finding the smallest I such that $W_1 \cup \cdots \cup W_I$ contains an edge of $\mathcal{H}$ with relatively high probability. One way to guarantee that I is small would be if we had $|S \setminus W_1|$ small for most $S \in \mathcal{H}$ (i.e., most vertices of most edges $S \in \mathcal{H}$ are covered by W_1), and then that W_2 covered most of the vertices of most $S \setminus W_1$, and so on.

The condition that, say, $|S \setminus W_1|$ is small for most $S \in \mathcal{H}$ turns out to be too strong a condition to impose. However, if $\mathcal{H}$ is sufficiently spread, then we can guarantee a weaker result: for most $S \in \mathcal{H}$, there is an $S' \subseteq S \cup W_1$ such that $|S' \setminus W_1|$ is small. We can then discard S and focus only on S', and by iterating this repeatedly we obtain the desired result.

To be more precise, given a hypergraph $\mathcal{H}$, we say that a pair of sets (S, W) is *k-good* if there exists $S' \in \mathcal{H}$ such that $S' \subseteq S \cup W$ and $|S' \setminus W| \le k$, and we say that the pair is *k-bad* otherwise. We note for later that if (S, W) is k-bad with $S \cup W = Z$, then $(S, Z \setminus S)$ is also k-bad.

The next lemma shows that $(q; r, k)$-spread hypergraphs have few k-bad pairs with $S \in \mathcal{H}$ and W a set of size roughly $q|V|$. In the lemma statement we adopt the notation that $\binom{V}{m}$ is the set of subsets of V of size m.

Lemma 2.1. *Let $\mathcal{H}$ be an r-uniform n-vertex hypergraph on V which is $(q; r, k)$-spread. Let $C \ge 4$ and define $p = Cq$. If $pn \ge 2r$ and $p \le \frac{1}{2}$, then*

$$\left|\left\{(S, W) : S \in \mathcal{H},\ W \in \binom{V}{pn},\ (S, W) \text{ is } k\text{-bad}\right\}\right| \le 3(C/2)^{-k/2}|\mathcal{H}|\binom{n}{pn}.$$

Proof. Throughout this lemma, we make frequent use of the identity

$$\binom{a-c}{b-c} \Big/ \binom{a}{b} = \binom{b}{c} \Big/ \binom{a}{c},$$

which follows from the simple combinatorial identity $\binom{a}{b}\binom{b}{c} = \binom{a}{c}\binom{a-c}{b-c}$.

For $t \le r$, define

$$\mathcal{B}_t = \{(S, W) : S \in \mathcal{H},\ W \in \binom{v}{pn},\ (S, W) \text{ is } k\text{-bad},\ |S \cap W| = t\}.$$

Observe that the quantity we wish to bound is $\sum_t |\mathcal{B}_t|$, so it suffices to bound each term of this sum. From now on, we fix some t and define

$$w = pn - t.$$

[1]Somewhat more precisely, let $\mathcal{H}$ be the hypergraph whose hyperedges consist of copies of F in K_n. If F has r edges and maximum degree d, and if $\mathcal{H}$ is (q, α, δ)-superspread as defined in [3], then one can show that $\mathcal{H}$ is $(Cq; r, C_1 r^{1-\alpha}, C_2 r^{1-2\alpha}, \ldots, C_{\lfloor 1/\alpha \rfloor} r^{1-\lfloor 1/\alpha \rfloor \alpha}, 1)$-spread for some constants C, C_i which depend on d, δ. Indeed, when verifying Definition 1.2 for $j \ge \delta k$, one can use a similar argument as in Proposition 1.3(b) and the fact that $\mathcal{H}$ is q-spread. If $j < \delta k$, then the superspread condition together with Lemma 2.3 of [3] can be used to give the result.

At this point, we need to count the number of elements in $\mathcal{B}_t$, and there are several natural approaches that could be used. One way would be to first pick any $S \in \mathcal{H}$ and then count how many W satisfy $(S, W) \in \mathcal{B}_t$. Another approach would be to pick any set Z of size $r + w$ (which will be the size of $S \cup W$ since $|S \cap W| = t$) and then bound how many $S, W \subseteq Z$ have $(S, W) \in \mathcal{B}_t$. For some pairs, the first approach is more efficient, and for others the second is. In particular, the second approach will be more effective whenever $Z = S \cup W$ contains few elements of $\mathcal{B}_t$.

With this in mind, we say that a set Z is *pathological* if

$$|\{S \in \mathcal{H} : S \subseteq Z,\ (S, Z \setminus S) \text{ is } k\text{-bad}\}| > N,$$

where

$$N := (C/2)^{-k/2}|\mathcal{H}| \binom{n-r}{w} / \binom{n}{r+w} = (C/2)^{-k/2}|\mathcal{H}| \binom{r+w}{r} / \binom{n}{r},$$

and we say that Z is *non-pathological* otherwise. We say that a pair (S, W) is *pathological* if the set $S \cup W$ is pathological and that (S, W) is *non-pathological* otherwise. □

Claim 2.2. *The number of* $(S, W) \in \mathcal{B}_t$ *which are non-pathological is at most*

$$\binom{n}{r+w} N \binom{r}{t} = (C/2)^{-k/2}|\mathcal{H}| \binom{r}{t} \binom{n-r}{w}.$$

Proof. We identify each of the non-pathological pairs (S, W) by specifying $S \cup W$, then S, then $S \cap W$.

Observe that $S \cup W$ is a non-pathological set of size $r + w$, and in particular there are at most $\binom{n}{r+w}$ ways to make this first choice. Fix such a non-pathological set Z of size $r + w$. As noted before the statement of Lemma 2.1, if (S, W) is k-bad with $S \cup W = Z$, then $(S, Z \setminus S)$ is also k-bad. Because Z is non-pathological, there are at most N choices for S such that $(S, Z \setminus S)$ is k-bad. Given S, there are at most $\binom{r}{t}$ choices for $S \cap W$. Multiplying the number of choices at each step gives the stated result. □

Claim 2.3. *The number of* $(S, W) \in \mathcal{B}_t$ *which are pathological is at most*

$$2(C/2)^{-k/2}|\mathcal{H}| \binom{r}{t} \binom{n-r}{w}$$

Proof. We identify these pairs by first specifying $S \in \mathcal{H}$, then $S \cap W$, then $W \setminus S$.

Note that S and $S \cap W$ can be specified in at most $|\mathcal{H}| \cdot \binom{r}{t}$ ways, and from now on we fix such a choice of S and $S \cap W$. It remains to specify $W \setminus S$, which will be some element of $\binom{V \setminus S}{w}$. Thus it suffices to count the number of $W' \in \binom{V \setminus S}{w}$ such that (S, W') is both k-bad and pathological.

For $W' \in \binom{V \setminus S}{w}$, define

$$\mathcal{S}(W') = |\{S' \in \mathcal{H} : S' \subseteq S \cup W',\ |S' \cap S| \geq k\}|.$$

Observe that if (S, W') is k-bad, then every edge $S' \subseteq S \cup W'$ has $|S' \cap S| \geq k$ (since $|S' \cap S| \geq |S' \setminus W'|$), so the W' we wish to count satisfy

$$\mathcal{S}(W') = |\{S' \in \mathcal{H} : S' \subseteq S \cup W'\}|.$$

If (S, W') is pathological, then this latter set has size at least N. In total, if $\mathbf{W}'$ is chosen uniformly at random from $\binom{V \setminus S}{w}$, then

$$\mathbb{P}[(S, \mathbf{W}') \text{ is } k\text{-bad and pathological}] \leq \mathbb{P}[\mathcal{S}(\mathbf{W}') \geq N] \leq \frac{\mathbb{E}[\mathcal{S}(\mathbf{W}')]}{N}, \tag{1}$$

where this last step used Markov's inequality. It remains to upper bound $\mathbb{E}[\mathcal{S}(\mathbf{W}')]$.

https://doi.org/10.1017/S0963548323000202 Published online by Cambridge University Press

Let

$$m_j(S) = |\{S' \in \mathcal{H} : |S \cap S'| = j\}|,$$

and observe that for any S' with $|S \cap S'| = j$, the number of $W' \in \binom{V \setminus S}{w}$ with $S' \subseteq S \cup W'$ is exactly $\binom{n-2r+j}{w-r+j}$. With this we see that

$$\mathbb{E}[\mathcal{S}(\mathbf{W}')] = \sum_{j \geq k} m_j(S) \frac{\binom{n-2r+j}{w-r+j}}{\binom{n-r}{w}} = \sum_{j \geq k} m_j(S) \frac{\binom{w}{r-j}}{\binom{n-r}{r-j}} = \frac{\binom{r+w}{r}}{\binom{n}{r}} \sum_{j \geq k} m_j(S) \frac{\binom{w}{r-j}}{\binom{n-r}{r-j}} \cdot \frac{\binom{n}{r+w}}{\binom{n-r}{w}}. \tag{2}$$

Because $\mathcal{H}$ is $(q; r, k)$-spread, we have for each $j \geq k$ in the sum that

$$m_j(S) \leq M_j(S) \leq q^j |\mathcal{H}|. \tag{3}$$

For integers x, y, define the falling factorial $(x)_y := x(x-1) \cdots (x-y+1)$. With this we have

$$\frac{\binom{w}{r-j}}{\binom{n-r}{r-j}} \cdot \frac{\binom{n}{r+w}}{\binom{n-r}{w}} = \frac{(w)_{r-j}}{(n-r)_{r-j}} \cdot \frac{(n)_r}{(r+w)_r} \leq \left(\frac{w}{n-r}\right)^{r-j} \cdot \left(\frac{n-r}{w}\right)^r = \left(\frac{w}{n-r}\right)^{-j} \leq (Cq/2)^{-j}, \tag{4}$$

where the first inequality used $w \leq pn \leq \frac{1}{2}n \leq n-r$, and the second inequality used

$$w = pn - t \geq pn - r \geq pn/2 = Cqn/2.$$

Combining (2), (3), and (4) shows that

$$\mathbb{E}[\mathcal{S}(\mathbf{W}')] \leq \frac{\binom{r+w}{r}}{\binom{n}{r}} |\mathcal{H}| (C/2)^{-k} \cdot \sum_{j \geq k} (C/2)^{k-j} \leq \frac{\binom{r+w}{r}}{\binom{n}{r}} |\mathcal{H}| (C/2)^{-k} \cdot 2,$$

where this last step used $C \geq 4$. Plugging this into (1) shows that the number of $W' \in \binom{V \setminus S}{w}$ such that (S, W') is k-bad and pathological is at most

$$2(C/2)^{-k} |\mathcal{H}| \frac{\binom{r+w}{r}}{\binom{n}{r} N} \cdot \binom{n-r}{w} = 2(C/2)^{-k/2} \cdot \binom{n-r}{w}.$$

Combining this with the fact that there were $|\mathcal{H}| \cdot \binom{r}{t}$ ways of choosing S and $S \cap W$ gives the claim. □

In total, $|\mathcal{B}_t|$ is at most the sum of the bounds from these two claims. Using this and $w = pn - t$ implies

$$\sum_{t \leq r} |\mathcal{B}_t| \leq \sum_{t \leq r} 3(C/2)^{-k/2} |\mathcal{H}| \binom{r}{t} \binom{n-r}{pn-t}$$
$$= 3(C/2)^{-k/2} |\mathcal{H}| \binom{n}{pn},$$

giving the desired result.

https://doi.org/10.1017/S0963548323000202 Published online by Cambridge University Press

2.2. Auxiliary lemmas

To prove Theorem 1.4, we need to consider two special cases. The first is when $\mathcal{H}$ is r-uniform with r relatively small. In this case, the following lemma gives effective bounds.

Lemma 2.4 ([4] Corollary 4.2). *Let $\mathcal{H}$ be a q-spread r-bounded hypergraph on V and let $\alpha \in (0,1)$ satisfy $\alpha \geq 2rq$. If W is a set of size $\alpha|V|$ chosen uniformly at random from V, then the probability that W does not contain an element of $\mathcal{H}$ is at most*

$$2e^{-\alpha/(2rq)}.$$

The other special case we consider is the following.

Lemma 2.5. *Let $\mathcal{H}$ be an r-uniform $(q; r, 1)$-spread hypergraph on V and $\alpha \in (0,1)$ such that $\alpha \geq 4q$. If W is a set of size $\alpha|V|$ chosen uniformly at random from V, then the probability that W does not contain an edge of $\mathcal{H}$ is at most*

$$4q\alpha^{-1} + 2e^{-\alpha|V|/4}.$$

Proof. Let W' be a random set of V obtained by including each vertex independently and with probability $\alpha/2$. Let $X = |\{S \in \mathcal{H} : S \subseteq W'\}|$ and define $m_j(S)$ to be the number of $S' \in \mathcal{H}$ with $|S \cap S'| = j$. Note that $\mathbb{E}[X] = (\alpha/2)^r|\mathcal{H}|$ and that

$$\begin{aligned}
\mathrm{Var}(X) &\leq (\alpha/2)^{2r} \sum_{S,S' \in \mathcal{H},\, S\cap S' \neq \emptyset} (\alpha/2)^{-|S\cap S'|} = (\alpha/2)^{2r} \sum_{S\in\mathcal{H}} \sum_{j=1}^{r} (\alpha/2)^{-j} \cdot m_j(S) \\
&\leq (\alpha/2)^{2r} \sum_{S\in\mathcal{H}} \sum_{j=1}^{r} (\alpha/2)^{-j} \cdot q^j|\mathcal{H}| = (\alpha/2)^{2r} \sum_{j=1}^{r} (\alpha/2q)^{-j}|\mathcal{H}|^2 \\
&= \mathbb{E}[X]^2(\alpha/2q)^{-1} \sum_{j=1}^{r} (\alpha/2q)^{1-j} \leq 4\mathbb{E}[X]^2 q\alpha^{-1},
\end{aligned}$$

where the second inequality used that $\mathcal{H}$ being $(q; r, 1)$-spread implies $m_j(S) \leq q^j|\mathcal{H}|$ for all $S \in \mathcal{H}$ and $j \geq 1$, and the last inequality used $\alpha/2q \geq 2$. By Chebyshev's inequality we have

$$\mathbb{P}[X = 0] \leq \mathrm{Var}(X)/\mathbb{E}[X]^2 \leq 4q\alpha^{-1}.$$

Lastly, observe that

$$\begin{aligned}
\mathbb{P}[W \text{ contains an edge of } \mathcal{H}] &\geq \mathbb{P}[W' \text{ contains an edge of } \mathcal{H} \,\big|\, |W'| \leq \alpha|V|] \\
&\geq \mathbb{P}[W' \text{ contains an edge of } \mathcal{H}] - \mathbb{P}[W' > \alpha|V|].
\end{aligned}$$

By the Chernoff bound (see for example [1]) we have $\mathbb{P}[|W'| > \alpha|V|] \leq 2e^{-\alpha|V|/4}$. Note that W' contains an edge of $\mathcal{H}$ precisely when $X > 0$, so the result follows from our analysis above. □

We conclude this subsection with a small observation.

Lemma 2.6. *If $\mathcal{H}$ is an r_1-uniform $(q; r_1, \ldots, r_\ell)$-spread hypergraph on V, then $r_1 \leq eq|V|$.*

Proof. Let $m = \max_{S\in\mathcal{H}} d(S)$, i.e. this is the maximum multiplicity of any edge in $\mathcal{H}$. Then for any $S \in \mathcal{H}$ with $d(S) = m$, we have

$$m = M_{r_1}(S) \leq q^{r_1}|\mathcal{H}| \leq q^{r_1} \cdot m\binom{|V|}{r_1} \leq m(eq|V|/r_1)^{r_1},$$

proving the result. □

2.3. Putting the pieces together

We now prove a technical version of Theorem 1.4 with more explicit quantitative bounds. Theorem 1.4 will follow shortly (but not immediately) after proving this.

Theorem 2.7. *Let $\mathcal{H}$ be an r_1-uniform $(q; r_1, \ldots, r_\ell, 1)$-spread hypergraph on V and let $C \geq 8$. If W is a set of size $2C\ell q|V|$ chosen uniformly at random from V, then*

$$\mathbb{P}[W \text{ contains an edge of } \mathcal{H}] \geq 1 - 6\ell^2(C/4)^{-r_\ell/2} - 40(C\ell)^{-1}, \tag{5}$$

and for any i with $4r_i \leq C\ell$ we have

$$\mathbb{P}[W \text{ contains an edge of } \mathcal{H}] \geq 1 - 6\ell^2(C/4)^{-r_i/2} - 2e^{-C\ell/4r_i}. \tag{6}$$

Proof. Define $p := Cq$ and $n := |V|$. We can assume $p \leq \frac{1}{2}$, as otherwise the result is trivial (since the set W in the hypothesis of the theorem has size at least $|V|$). Let $W_1, \ldots W_{\ell-1}$ be chosen independently and uniformly at random from $\binom{V}{pn}$. Throughout this proof we let $r_{\ell+1} = 1$.

Let $\mathcal{H}_1 = \mathcal{H}$ and let $\phi_1 : \mathcal{H}_1 \to \mathcal{H}$ be the identity map. Inductively assume we have defined $\mathcal{H}_i$ and $\phi_i : \mathcal{H}_i \to \mathcal{H}$ for some $1 \leq i < \ell$. Let $\mathcal{H}'_i \subseteq \mathcal{H}_i$ be all the edges $S \in \mathcal{H}_i$ such that (S, W_i) is r_{i+1}-good with respect to $\mathcal{H}_i$. Thus for each $S \in \mathcal{H}'_i$, there exists an $S' \in \mathcal{H}_i$ such that $S' \subseteq S \cup W_i$ and $|S' \setminus W_i| \leq r_{i+1}$. Choose such an S' for each $S \in \mathcal{H}'_i$ and let A_S be any subset of S of size exactly r_{i+1} that contains $S' \setminus W_i$ (noting that $S' \setminus W_i \subseteq S$ since $S' \subseteq S \cup W_i$). Finally, define $\mathcal{H}_{i+1} = \{A_S : S \in \mathcal{H}'_i\}$ and $\phi_{i+1} : \mathcal{H}_{i+1} \to \mathcal{H}$ by $\phi_{i+1}(A_S) = \phi_i(S)$.

Intuitively, $\phi_i(A)$ is meant to correspond to the "original" edge $S \in \mathcal{H}$ which generated A. More precisely, we have the following. □

Claim 2.8. *For $i \leq \ell$, the maps ϕ_i are injective and $A \subseteq \phi_i(A)$ for all $A \in \mathcal{H}_i$.*

Proof. This claim trivially holds for $i = 1$. Inductively assume the result has been proved for $1, \ldots, i$. Observe that in the process of generating $\mathcal{H}_{i+1}$, we have implicitly defined a bijection $\psi : \mathcal{H}'_i \to \mathcal{H}_{i+1}$ through the correspondence $\psi(S) = A_S$.

By construction of ϕ_{i+1}, we have $\phi_{i+1}(A) = \phi_i(\psi^{-1}(A))$, so ϕ_{i+1} is injective since ϕ_i was inductively assumed to be injective and ψ is a bijection. Also by construction we have $A \subseteq \psi^{-1}(A)$, and by the inductive hypothesis we have $\psi^{-1}(A) \subseteq \phi_i(\psi^{-1}(A)) = \phi_{i+1}(A)$. This completes the proof. □

For $i < \ell$, we say that W_i is *successful* if $|\mathcal{H}_{i+1}| \geq (1 - \frac{1}{2\ell})|\mathcal{H}_i|$. Note that $|\mathcal{H}_{i+1}| = |\mathcal{H}_i{}'|$, so this is equivalent to saying that the number of r_{i+1}-bad pairs (S, W_i) with $S \in \mathcal{H}_i$ is at most $\frac{1}{2\ell}|\mathcal{H}_i|$.

Claim 2.9. *For $i \leq \ell$, if $W_1, \ldots, W_{i-1}$ are successful, then $\mathcal{H}_i$ is $(2q; r_i, \ldots, r_\ell, 1)$-spread.*

Proof. For a hypergraph $\mathcal{H}'$, we let $M_j(A; \mathcal{H}')$ denote the number of edges of $\mathcal{H}'$ intersecting A in at least j vertices. By Claim 2.8, if $\{A_1, \ldots, A_t\}$ are the set of edges of $\mathcal{H}_i$ which intersect some set A in at least j vertices, then $\{\phi_i(A_1), \ldots, \phi_i(A_t)\}$ is a set of t distinct edges of $\mathcal{H}$ intersecting A in at least j vertices. Thus for all sets A and integers j we have $M_j(A; \mathcal{H}_i) \leq M_j(A; \mathcal{H})$.

If A is contained in an edge A' of $\mathcal{H}_i$, then by Claim 2.8 A is contained in the edge $\phi_i(A')$ of $\mathcal{H}$. Thus $d_{\mathcal{H}_i}(A) > 0$ implies $d_{\mathcal{H}}(A) > 0$. By assumption of $\mathcal{H}$ being $(q; r_1, \ldots, r_\ell, 1)$-spread, if A is a set with $r_{i'} \geq |A| \geq r_{i'+1}$ for some integer i' such that $d_{\mathcal{H}_i}(A) > 0$, and if j is an integer satisfying $j \geq r_{i'+1}$, then our previous observations imply

$$M_j(A; \mathcal{H}_i) \leq M_j(A; \mathcal{H}) \leq q^j|\mathcal{H}|. \tag{7}$$

Because each of $W_1, \ldots, W_{i-1}$ were successful, we have

$$|\mathcal{H}_i| \geq \left(1 - \frac{1}{2\ell}\right)^i |\mathcal{H}| \geq \left(1 - \frac{1}{2\ell}\right)^\ell |\mathcal{H}| \geq \frac{1}{2}|\mathcal{H}|,$$

where in this last step we used that $(1 - 1/(2x))^x$ is an increasing function for $x \geq 1$. Plugging $|\mathcal{H}| \leq 2|\mathcal{H}_i|$ into (7) shows that $\mathcal{H}_i$ is $(2q; r_i, \ldots, r_\ell, 1)$-spread as desired. □

Claim 2.10. *For $i < \ell$,*

$$\mathbb{P}[W_i \text{ is not successful} \mid W_1, \dots, W_{i-1} \text{ are successful}] \le 6\ell(C/4)^{-r_{i+1}/2}.$$

Proof. By construction $\mathcal{H}_i$ is r_i-uniform. Conditional on $W_1, \dots, W_{i-1}$ being successful, Claim 2.9 implies that $\mathcal{H}_i$ is in particular $(2q; r_i, r_{i+1})$-spread. By hypothesis we have $p \le \frac{1}{2}$ and $C/2 \ge 4$, and by Lemma 2.6 applied to $\mathcal{H}$ we have $2r_i \le pn$ since $C \ge 2e$. Thus we can apply Lemma 2.1 to $\mathcal{H}_i$ (using $C/2$ instead of C), which shows that the expected number of r_{i+1}-bad pairs (S, W_i) is at most $3(C/4)^{-r_{i+1}/2}|\mathcal{H}_i|$. By Markov's inequality, the probability of there being more than $\frac{1}{2\ell}|\mathcal{H}_i|$ total r_{i+1}-bad pairs is at most $6\ell(C/4)^{-r_{i+1}/2}$, giving the result. □

We are now ready to prove the result. Let W and W' be sets of size $2\ell pn$ and ℓpn chosen uniformly at random from V. Observe that for any $1 \le i \le \ell$, the probability of W containing an edge of $\mathcal{H}$ is at least the probability of $W_1 \cup \cdots \cup W_{i-1} \cup W'$ containing an edge of $\mathcal{H}$, and this is at least the probability that W' contains an edge of $\mathcal{H}_i$ (since every edge of $\mathcal{H}_i$ is an edge of $\mathcal{H}$ after removing vertices that are in $W_1 \cup \cdots \cup W_{i-1}$), so it suffices to show that this latter probability is large for some i.

By Proposition 1.3(a) and Claim 2.9, the hypergraph $\mathcal{H}_i$ will be $(2q)$-spread if $W_1, \dots, W_{i-1}$ are all successful. If i is such that $C\ell \ge 4r_i$, then by Claim 2.10 and Lemma 2.4 the probability that $W_1, \dots, W_{i-1}$ are all successful and W' contains an edge of $\mathcal{H}_i$ is at least

$$1 - 6\ell^2(C/4)^{-r_i/2} - 2e^{-C\ell/4r_i},$$

giving (6).

Alternatively, the probability that W' contains an edge of $\mathcal{H}_\ell$ can be computed using Lemma 2.5, which gives that the probability of success is at least

$$1 - 6\ell^2(C/4)^{-r_\ell/2} - 16(C\ell)^{-1} - 2e^{-C\ell qn/4}.$$

Using $qn \ge e^{-1}r_1 \ge 1/3$ from Lemma 2.6 together with $e^{-x} \le x^{-1}$ gives (5) as desired.

We now use this to prove Theorem 1.4.

Proof of Theorem 1.4. There exists a large constant K' such that if[2] $r_\ell \ge K' \log(\ell+1)$, then the result follows from (5). If this does not hold and if $r_1 > K' \log(\ell+1)$, then there exists some $I \ge 2$ such that $r_{I-1} > K' \log(\ell+1) \ge r_I$. If $r_I = K' \log(\ell+1)$, then the result follows from (6) with $i = I$ provided C is sufficiently large in terms of K'. Otherwise we define a new sequence of integers $r'_1, \dots, r'_{\ell+1}$ with $r'_i = r_i$ for $i < I$, $r'_I = K' \log(\ell+1)$, and $r'_i = r_{i-1}$ for $i > I$. It is not hard to see that $\mathcal{H}$ is $(q; r_1', \dots, r'_{\ell+1}, 1)$-spread, so the result follows[3] from (6) with $i = I$.

It remains to deal with the case $r_1 \le K' \log(\ell+1)$. Because $\ell \le r_1$, this can only hold if $r_1 \le K''$ for some large constant K''. In this case we can apply Lemma 2.4 to give the desired result by choosing K_0 sufficiently large in terms of K''. □

3. Concluding remarks

With a very similar proof one can prove the following non-uniform analog of Theorem 1.4.

Theorem 3.1. *Let $\mathcal{H}$ be a $(q; r_1, \dots, r_\ell, 1)$-spread hypergraph on V and define $s = \min_{S \in \mathcal{H}} |S|$. Assume that there exists a K such that $r_1 \le Kq|V|$, and such that for all i with $r_i > s$ we have $\log r_i \le Kr_{i+1}$. Then there exists a constant K_0 depending only on K such that if $r_\ell \le \max\{s, K_0 \log(\ell+1)\}$ and $C \ge K_0$, then a set W of size $C\ell q|V|$ chosen uniformly at random from V satisfies*

$$\mathbb{P}[W \text{ contains an edge of } \mathcal{H}] \ge 1 - \frac{K_0}{C\ell}.$$

[2] We consider $\log(\ell+1)$ as opposed to $\log(\ell)$ to guarantee that this is a positive number for all $\ell \ge 1$.

[3] The bound of (6) now uses $\ell+1$ instead of ℓ throughout because we are working with the r'_i sequence, but this does not affect the final result.

https://doi.org/10.1017/S0963548323000202 Published online by Cambridge University Press

Observe that if $\mathcal{H}$ is r_1-uniform then this reduces to Theorem 1.4 with the additional constraint that $r_1 \le Kq|V|$ for some K. By Lemma 2.6, this extra condition is always satisfied for uniform hypergraphs with $K = e$. We note that Theorem 3.1 together with Proposition 1.3(b) implies Theorem 1.1. We briefly describe the details on how to prove this.

Sketch of Proof. We first adjust the statement and proof of Lemma 2.1 to allow $\mathcal{H}$ to be r-bounded. To do this, we partition $\mathcal{H}$ into the uniform hypergraphs $\mathcal{H}_{r'} = \{S \in \mathcal{H} : |S| = r'\}$, and word for word the exact same proof[4] as before shows that the number of k-bad pairs using $S \in \mathcal{H}_{r'}$ is at most $3(C/2)^{-k/2}|\mathcal{H}|\binom{n}{pn}$. We then add these bounds over all r' to get the same bound as in Lemma 2.1 multiplied by an extra factor of r. With regards to the other lemmas, one no longer needs Lemma 2.6 due to the $r_1 \le Kq|V|$ hypothesis, and Lemmas 2.4 and 2.5 are fine as is (in particular, Lemma 2.5 still requires $\mathcal{H}$ to be uniform).

For the main part of the proof, instead of choosing A_S to be a subset of S of size exactly r_i, we choose it to have size at most r_i and at least $\min\{r_i, s\}$. With this $\mathcal{H}_i$ will be uniform if $r_i \le s$, and otherwise when we apply the non-uniform version of Lemma 2.1 our error term will have an extra factor of $r_i \le e^{Kr_{i+1}}$, with this inequality holding by our hypothesis for $r_i > s$. This term will be insignificant compared to $(C/2)^{-r_{i+1}/2}$ provided C is large in terms of K.

If $r_\ell \le K' \log(\ell + 1)$ for some large K' depending on K, then as in the proof of Theorem 1.4 we can assume $r_I = K' \log(\ell + 1)$ for some I and conclude the result as before. Otherwise $r_\ell \le s$ by hypothesis, so $\mathcal{H}_\ell$ will be uniform and we can apply Lemma 2.5 to conclude the result. □

Another extension can be made by not requiring the same "level of spreadness" throughout $\mathcal{H}$.

Definition 3.2. Let $0 < q_1, \ldots, q_{\ell-1} \le 1$ be real numbers and $r_1 > \cdots > r_\ell$ positive integers. We say that a hypergraph $\mathcal{H}$ on V is $(q_1, \ldots, q_{\ell-1}; r_1, \ldots, r_\ell)$-spread if $\mathcal{H}$ is non-empty, r_1-bounded, and if for all $A \subseteq V$ with $d(A) > 0$ and $r_i \ge |A| \ge r_{i+1}$ for some $1 \le i < \ell$, we have for all $j \ge r_{i+1}$ that

$$M_j(A) := |\{S \in \mathcal{H} : |A \cap S| \ge j\}| \le q_i^j |\mathcal{H}|.$$

Different levels of spread was also considered in [2]. Here, one can prove the following.

Theorem 3.3. *Let $\mathcal{H}$ be a $(q_1, \ldots, q_\ell; r_1, \ldots, r_\ell, 1)$-spread hypergraph on V and define $s = \min_{S \in \mathcal{H}} |S|$. Assume that there exists a K such that for all i we have $r_i \le Kq_i|V|$, and that for all i with $r_i > s$ we have $\log r_i \le Kr_{i+1}$. Then there exists a constant K_0 depending only on K such that if $r_\ell \le \max\{s, K_0 \log(\ell + 1)\}$ and if $C \ge K_0$, then a set W of size $C \sum q_i|V|$ chosen uniformly at random from V satisfies*

$$\mathbb{P}[W \text{ contains an edge of } \mathcal{H}] \ge 1 - \frac{K_0 \log(\ell + 1)}{CL},$$

where $L := \sum_i q_i / \max_i q_i$.

Note that $\sum q_i \le \ell \max q_i$, so we have $L \le \ell$ with equality if $q_i = q_j$ for all i, j.

Sketch of Proof. We now choose our random sets W_i to have sizes $Cq_i|V|$ and W' to have size $C \sum q_i|V| = C(L \cdot \max q_i)|V|$. With this any of the $\mathcal{H}_i$ could be at worst $(2 \max q_i)$-spread if each $\mathcal{H}_{i'}$ was successful, so in this case when we apply Lemma 2.4 with W' we end up getting a probability of roughly $1 - e^{-CL/r_i}$ of containing an edge. From this quantity, we should subtract roughly $\ell^2 C^{-r_i}$, since this is the probability that some $\mathcal{H}_{i'}$ is unsuccessful. If $r_i = K' \log(\ell + 1)$ for some large constant K' then this gives the desired bound. Otherwise by using the same logic as in the

[4]The $\mathcal{H}_{r'}$ hypergraphs may not be spread, but they still have the property that $m_j(S) \le q^j|\mathcal{H}|$ for all $S \in \mathcal{H}_{r'} \subseteq \mathcal{H}$, and this is the only point in the proof where we used that $\mathcal{H}$ is spread.

https://doi.org/10.1017/S0963548323000202 Published online by Cambridge University Press

proof of Theorem 1.4 we can assume $r_\ell > K' \log(\ell + 1)$ and apply Lemma 2.5 to $\mathcal{H}_\ell$ to get a probability of roughly $1 - (CL)^{-1}$, which also gives the result after subtracting $\ell^2 C^{-r_\ell}$ to account for some $\mathcal{H}_{i'}$ being unsuccessful. □

Lastly, we note that Frieze and Marbach [5] recently developed a variant of Theorem 1.1 for rainbow structures in hypergraphs. We suspect that straightforward generalizations of our proofs and those of [5] should give an analog of Theorem 1.4 (as well as Theorems 3.1 and 3.3) for the rainbow setting.

Acknowledgments

We thank Bhargav Narayanan for looking over an earlier draft.

References

[1] Alon, N. and Spencer, J. H. (2004) *The Probabilistic Method*. John Wiley & Sons.
[2] Alweiss, R., Lovett, S., Wu, K. and Zhang, J. (2020) Improved bounds for the sunflower lemma. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 624–630.
[3] Espuny Díaz, A. and Person, Y. (2021) Spanning *F*-cyles in random graphs. *arXiv preprint arXiv:* 2106.10023.
[4] Frankston, K., Kahn, J., Narayanan, B. and Park, J. (2021) Thresholds versus fractional expectation-thresholds. *Ann. Math.* **194**(2) 475–495.
[5] Frieze, A. and Marbach, T. G. (2021) Rainbow thresholds. *arXiv preprint arXiv:* 2104.05629.
[6] Johansson, A., Kahn, J. and Vu, V. (2008) Factors in random graphs. *Random Struct. Algorithms* **33**(1) 1–28.
[7] Kahn, J., Narayanan, B. and Park, J. (2021) The threshold for the square of a hamilton cycle. *Proc. Am. Math. Soc.* **149**(8) 3201–3208.

Cite this article: Spiro S (2023). A smoother notion of spread hypergraphs. *Combinatorics, Probability and Computing* **32**, 809–818. https://doi.org/10.1017/S0963548323000202