PCMI lecture series: from sunflowers to thresholds

Lecturer: Shachar Lovett
TA: Anthony Ostuni

July 2025

Abstract

These are lecture notes for a mini-course given at PCMI in July 2025. The main topics are the sunflower conjecture, threshold phenomena, their connection using the notion of spreadness, and an application of these notions in complexity theory.

Contents

1 Introduction to the sunflower conjecture

Definition 1.1 (Sunflower)**.** *A family of distinct sets $S_1, \ldots, S_r$ is called an r-*sunflower *if all their pairwise intersection are the same. In other words, if we let $K = \cap S_i$, then $S_i \setminus K$ are pairwise disjoint. We call K the* kernel *and $S_i \setminus K$ the* petals *of the sunflower.*

Example 1.2. $S_1 = \{1,2,3,4\}, S_2 = \{1,2,5\}, S_3 = \{1,2,6,7,8\}$ *is a 3-sunflower with kernel* $K = \{1,2\}$ *and petals* $\{3,4\}, \{5\}, \{6,7,8\}$.

We allow sunflowers to have empty kernels; in this case, a sunflower is simply a family of pairwise disjoint sets.

Example 1.3. $S_1 = \{1,2\}, S_2 = \{3,4,5\}, S_3 = \{6\}$ *is a 3-sunflower with empty kernel.*

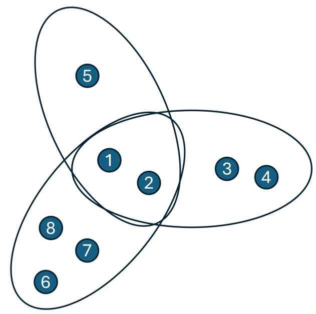

Figure 1: 3-sunflower

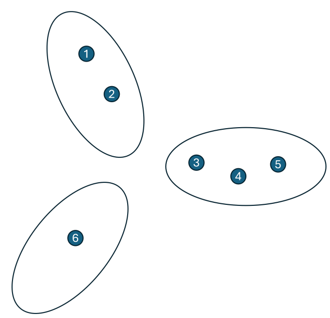

Figure 2: 3-sunflower with empty kernel

The motivating question would be to understand, how large can a family of sets be without containing a sunflower? Note that any 2 sets form a 2-sunflower, so the question is only interesting for $r \geq 3$. We will need to put some restriction on the sets - either their size is bounded, or the universe size is bounded, otherwise the family can be infinite.

Exercise 1.1. *Show that there is an infinite family of sets not containing a* 3*-sunflower.*

1.1 Erdős-Rado sunflowers

The most well-studied restriction is to bound the size of the sets in the family, a setting that was initiated by Erdős and Rado in 1960 [ER60]. An n-set is a set of size n. Denote

$$SF(n, r) = \text{largest size of a family of } n\text{-sets that does not contain an } r\text{-sunflower.}$$

Equivalently, any family of n-sets of size $SF(n, r) + 1$ must contain an r-sunflower.

We will restrict our attention to families of n-sets; with small modifications, the arguments generalize to families of sets of size at most n.

Exercise 1.2. *Define $SF^{\leq}(n, r)$ to be the largest size of a family of sets of size* at most *n that does not contain an r-sunflower. Prove that $SF^{\leq}(n, r) = SF(n, r)$.*

Claim 1.4. $SF(n, r) \geq (r - 1)^n$.

Proof. Let $R_1, \ldots, R_n$ be pairwise disjoint sets, each of size $r-1$, and let U be their union. Consider the set system

$$\mathcal{F} = \{S \subset U : |S \cap R_i| = 1 \; \forall i \in [n]\}.$$

Equivalently, one can think of U as an $n \times (r-1)$ matrix, and $\mathcal{F}$ contains all sets which have exactly one element in each row.

Note that $\mathcal{F}$ is a family of n-sets, and $|\mathcal{F}| = (r-1)^n$. We claim that $\mathcal{F}$ does not contain an r-sunflower, which implies the claimed bound. Assume towards a contradiction that $S_1, \ldots, S_r \in \mathcal{F}$ form a sunflower. Consider a set R_i. By construction, $|S_j \cap R_i| = 1$ for all $j \in [r]$. However, since there are r sets $S_1, \ldots, S_r$, and $|R_i| = r-1$, there must be $j', j'' \in [r]$ such that $S_{j'} \cap R_i = S_{j''} \cap R_i = \{x\}$. Since the sets form a sunflower, it must be that $S_j \cap R_i = \{x\}$ for all $j = 1, \ldots, r$. Since this applies for all i, we must have $S_1 = \cdots = S_r$, a contradiction, since sets in a sunflower must be distinct. □

The following lemma by Erdős-Rado [ER60] gives a bound to the maximal size of a family avoiding a sunflower.

Lemma 1.5 (Sunflower lemma)**.** *$SF(n, r) \leq n! \cdot (r-1)^n$.*

Proof. The proof is by induction on n. For $n = 1$, any family of 1-sets forms a sunflower (with empty kernel), so $SF(1, r) = r - 1$.

Let $\mathcal{F}$ be any family of n-sets which does not contain a sunflower. Let $S_1, \ldots, S_m$ be a maximal collection of pairwise disjoint sets in $\mathcal{F}$. Note that $m \leq r-1$, otherwise we found r pairwise disjoint sets which form an r-sunflower (with empty kernel).

Let $A = \cup S_i$ and note that $|A| \leq (r-1)n$. By construction, A intersects all sets in $\mathcal{F}$. Let $x \in A$ be the most common element, and define

$$\mathcal{F}' = \{S \setminus \{x\} : S \in \mathcal{F}, x \in S\}.$$

Note that $\mathcal{F}'$ is a family of $(n-1)$-sets, of size

$$|\mathcal{F}'| \geq \frac{|\mathcal{F}|}{|A|} \geq \frac{|\mathcal{F}|}{(r-1)n}.$$

Moreover, $\mathcal{F}'$ does not contain an r-sunflower. This is since if $S_1, \ldots, S_r$ are a sunflower in $\mathcal{F}'$ then $S_1 \cup \{x\}, \ldots, S_r \cup \{x\}$ are a sunflower in $\mathcal{F}$, which we assume does not exist. Thus by induction

$$|\mathcal{F}'| \leq SF(n-1, r) \leq (n-1)! \cdot (r-1)^{n-1}.$$

This then implies the claim bound:

$$|\mathcal{F}| \leq (r-1)n \cdot |\mathcal{F}'| \leq n! \cdot (r-1)^n.$$ □

Erdős and Rado conjectured that the true behavior is closer to the lower bound than the upper bound, and made the following concrete conjecture, known as the *sunflower conjecture*.

Conjecture 1.6 (Sunflower conjecture)**.** *For any $r \geq 3$, $SF(n,r) \leq c^n$ for some $c = c(r)$.*

History of the Erdős-Rado sunflower conjecture (below c denotes some global constant):

- Sunflower lemma: $SF(n,r) \leq n! \cdot (r-1)^n \approx (rn)^n$
- Kostochka [Kos96] improved bound for constant r to $n^{(1-o(1))n}$.
- Fukuyama [Fuk18] improved bound for $r = 3$ to $\approx n^{(3/4)n}$.
- Alweiss, Lovett, Wu, Zhang [ALWZ21] improved to $(cr\log(rn))^n \cdot (\log\log n)^{O(n)}$.
- Frankston, Kahn, Narayanan, Park [FKNP21] improved bound to $(cr\log(rn))^n$; used technique to prove the "fractional Kahn-Kalai conjecture".
- Proof re-cast in language of information theory by Rao [Rao20] and Tao [Tao20].
- Bell, Chueluecha, Warnke [BCW21] improved the bound to $(cr\log n)^n$.
- Park and Pham [PP24] refined the technique and proved the full Kahn-Kalai conjecture.

1.2 Erdős-Szemerédi sunflowers

Erdős and Szemerédi [ES78] considered another restriction, bounding the size of the universe instead of the size of the sets. Define

$$ES(N,r) = \text{largest size of a family of sets of } [N] \text{ which does not contain an } r\text{-sunflower.}$$

Note that $ES(N,r) \leq 2^N$ since there are 2^N subsets of $[N]$ in total.

Exercise 1.3. *Show that $ES(r-1,r) = 2^{r-1} - 1$.*

Claim 1.7. *Let $r \geq 3$ and N be large enough. Then $ES(N,r) \geq 2^{(1-\varepsilon)N}$ where $\varepsilon = \varepsilon(r) = O\left(\frac{\log r}{r}\right)$.*

Proof. Assume $N = m(r-1)$ and decompose $[N]$ as the disjoint union of $R_1, \ldots, R_m$, each of size $r-1$. Let $k = \lceil (r-1)/2 \rceil$. Define

$$\mathcal{F} = \{S \subset [N] : \forall i, |S \cap R_i| = k\}.$$

Note that

$$|\mathcal{F}| = \binom{r-1}{k}^{N/(r-1)} = 2^N \cdot \left(\frac{\binom{r-1}{k}}{2^{r-1}}\right)^{N/(r-1)} \approx 2^N \cdot \left(\frac{c}{\sqrt{r}}\right)^{N/(r-1)} \approx 2^{\left(1-c'\frac{\log r}{r}\right)N}.$$

We claim that $\mathcal{F}$ does not contain an r-sunflower. Assume towards a contradiction it contains a sunflower $S_1, \ldots, S_r$. Fix $i \in [m]$ and consider the sets $T_1 = S_1 \cap R_i, \ldots, T_r = S_r \cap R_i$. Note that by construction, $T_1, \ldots, T_r$ all have size k.

Consider first the case where two of these sets are the same, say $T_1 = T_2$. In this case, since $S_1, \ldots, S_r$ form a sunflower, we must have

$$S_1 \cap S_2 \cap R_i = S_1 \cap S_j \cap R_i \qquad \forall j = 3, \ldots, r,$$

which implies that

$$T_1 = T_1 \cap T_2 = T_1 \cap T_j \qquad \forall j = 3, \ldots, r.$$

Since all T_j have the same size k, can only happen if $T_1 = T_2 = \cdots = T_r$.

The other case is that $T_1, \ldots, T_r$ are all distinct. In this case, they form an r-sunflower in R_i. This however is impossible: if $K = \cap T_i$ is their kernel, then $T_1 \setminus K, \cdots, T_r \setminus K$ are r pairwise disjoint non-empty sets inside R_i, a set of size $r-1$. So the former case must hold. Since we can choose any i, we must have $S_1 = \cdots = S_r$, a contradiction. $\square$

Erdős and Szemerédi conjectured that the true answer should be closer to the lower bound.

Conjecture 1.8 (Erdős-Szemerédi sunflower conjecture)**.** *For any $r \geq 3$, $ES(N, r) \leq 2^{(1-\varepsilon)N}$ for some $\varepsilon = \varepsilon(r) > 0$.*

Lemma 1.9. *The Erdős-Rado sunflower conjecture (Conjecture 1.6) implies the Erdős-Szemerédi sunflower conjecture (Conjecture 1.8).*

Proof. Fix $r \geq 3$ and N large enough. Let $\mathcal{F}$ be a family of subsets of $[N]$ which does not contain an r-sunflower. We want to prove that $|\mathcal{F}| \leq 2^{(1-\varepsilon)N}$ for some $\varepsilon = \varepsilon(r) > 0$.

As a first step, let $s \in [N]$ be the most common size of sets in $\mathcal{F}$, and define

$$\mathcal{F}_1 := \{S \in \mathcal{F} : |S| = s\}.$$

Note that $|\mathcal{F}| \leq (N+1)|\mathcal{F}_1|$, so it will suffice to bound the size of $\mathcal{F}_1$. Moreover, we may assume $s \in [0.4N, 0.6N]$, otherwise $|\mathcal{F}_1| \leq \binom{N}{0.4N} \leq 1.9^N$ and we are done.

Now that we have reduced our analysis to a family of s-sets, it is tempting to apply the Erdős-Rado sunflower conjecture directly to $\mathcal{F}_1$. Observe, however, that this would only yield a bound of $|\mathcal{F}| \lessapprox c^{0.6N} = 2^{0.6 \log(c) \cdot N}$ for some $c = c(r)$, which is trivial if (say) $c \geq 4$. Instead, let $m < s$ to be optimized later, and let $I \subset [N]$ be a set of size $|I| = s - m$ which is contained in the largest number of sets in $\mathcal{F}_1$. Define

$$\mathcal{F}_2 := \{S \setminus I : S \in \mathcal{F}_1, I \subset S\}.$$

Note that $\mathcal{F}_2$ is a family of sets of size m which does not contain an r-sunflower. Since we assume the Erdős-Rado sunflower conjecture, there exists $c = c(r)$ such that

$$|\mathcal{F}_2| \leq c^m.$$

On the other hand, since each set $S \in \mathcal{F}_1$ contains $\binom{s}{m}$ many sets of size $s - m$, and since we pick I to be the most popular set, we have

$$|\mathcal{F}_2| \geq \frac{\binom{s}{m}}{\binom{N}{s-m}} |\mathcal{F}_1| \geq \frac{\binom{s}{m}}{2^N} |\mathcal{F}_1|.$$

Putting everything together gives

$$|\mathcal{F}| \leq (N+1)2^N \frac{c^m}{\binom{s}{m}}.$$

We now set $m = \delta s$ for $\delta = O(1/c)$. Note that

$$\binom{s}{m} = \binom{m/\delta}{m} \approx (1/\delta)^m$$

so that overall we can bound

$$|\mathcal{F}| \lessapprox 2^N (c\delta)^m = 2^{(1-\varepsilon)N}$$

for $\varepsilon = \Theta(1/c)$. □

Exercise 1.4. *Show that if in Lemma 1.9 we use the sunflower lemma (Lemma 1.5) instead of the sunflower conjecture, then it proves the weaker upper bound*

$$ES(N, r) \leq 2^{N - c\sqrt{N}}$$

for some $c = c(r)$.

History of the Erdős-Szemerédi sunflower conjecture:

- Conjectured by Erdős-Szemerédi [ES78] in 1978.
- Alon, Shpilka, Umans [ASU12] observed a connection to matrix multiplication.
- Ellenberg, Gijswijt [EG17] prove the cap-set conjecture, which implies the $r = 3$ case of the Erdős-Szemerédi conjecture via results in [ASU12].

2 Improved bounds for the sunflower conjecture via the spread lemma

We will study Erdős-Rado sunflowers from now on. That is, we are given a family of n-sets and are asking for the maximal size of such a family that does not contain an r-sunflower. Recall that the sunflower lemma gives a bound of

$$SF(n, r) \leq n! \cdot (r-1)^n \approx (rn)^n$$

whereas the Erdős-Rado sunflower conjecture speculates that

$$SF(n, r) \leq c^n$$

for some constant $c = c(r)$. Our main goal is to prove the following improved bound for the sunflower lemma, getting closer to the conjectured bound.

Theorem 2.1 (Improved bound for the sunflower lemma). *$SF(n,r) \leq (cr\log n)^n$ for some absolute constant c.*

The proof we will see is a combination of the original proof of [ALWZ21] and the follow up works refining it [FKNP21, BCW21, PP24]. In order to build motivation, let us review the proof of the original sunflower lemma from a different perspective.

Lemma 2.2 (Sunflower lemma, again). *$SF(n,r) \leq n! \cdot (r-1)^n$.*

Proof. The proof is by induction on n. Let $\mathcal{F}$ be a family of n-sets which does not contain an r-sunflower. Set $k = (r-1)n$. Assume first that there is an element x that belongs to at least a $(1/k)$-fraction of sets in $\mathcal{F}$. Define

$$\mathcal{F}_x = \{S \setminus \{x\} : S \in \mathcal{F}, x \in S\},$$

where by assumption $|\mathcal{F}_x| \geq |\mathcal{F}|/k$. Next, note that since $\mathcal{F}_x$ is a family of $(n-1)$-sets which does not contain an r-sunflower, we get by induction that

$$|\mathcal{F}_x| \leq (n-1)! \cdot (r-1)^{n-1}$$

and the claim follows.

The other case is that every element belongs to strictly less than a $(1/k)$-fraction of sets in $\mathcal{F}$. In this case, we claim that $\mathcal{F}$ contains r pairwise disjoint sets, which form an r-sunflower with empty kernel. To see that, let $S_1, \ldots, S_m$ be a maximal collection of pairwise disjoint sets in $\mathcal{F}$, and assume towards a contradiction that $m \leq r-1$. Let $A = \cup S_i$ and note that $|A| = mn \leq k$. Since each element $x \in A$ intersects strictly less than a $(1/k)$-fraction of sets in $\mathcal{F}$, we can lower bound the number of sets in $\mathcal{F}$ disjoint from A by

$$|\{S \in \mathcal{F} : S \cap A = \emptyset\}| \geq |\mathcal{F}| - \sum_{x \in A} |\{S \in \mathcal{F} : x \in S\}| > |\mathcal{F}|(1 - |A|/k) \geq 0.$$

This contradicts the maximality of m. $\square$

Note that in the proof we only used the "structured" case where a single element belongs to many sets in $\mathcal{F}$. But we also could have used two elements, or three elements, or any number of elements. This motivates the following definitions.

Definition 2.3 (Link). *Let $\mathcal{F}$ be a family of sets, and T be a set. The* link *of $\mathcal{F}$ at T is*

$$\mathcal{F}_T = \{S \setminus T : S \in \mathcal{F}, T \subset S\}.$$

Example 2.4. *Let $\mathcal{F} = \{\{1,2,3\},\{2,3,4\},\{4,5\}\}$ and $T = \{2,3\}$. Then $\mathcal{F}_T = \{\{1\},\{4\}\}$.*

Definition 2.5 (Spread family). *Let $\mathcal{F}$ be a family of sets, and let $k > 1$. We say that $\mathcal{F}$ is* k-spread *if for every set T,*

$$|\mathcal{F}_T| \leq |\mathcal{F}|/k^{|T|}.$$

In other words, a family of sets is k-spread if every element is in at most a $1/k$-fraction of sets; every pair of elements is in at most a $1/k^2$-fraction of sets; and so on.

The motivation for this definition is that it allows for a more general induction, which reduces the sunflower question from general families to spread families.

Lemma 2.6 (Reduction to spread families)**.** *Fix n, r. Assume that for all $n' \le n$, every family of n'-sets which is k-spread contains an r-sunflower. Then $SF(n,r) < k^n$.*

Proof. The proof is by induction on n. The base case of $n = 1$ follows since any family of k 1-sets is k-spread. Let $\mathcal{F}$ be a family of n-sets of size $|\mathcal{F}| \ge k^n$. We will prove that $\mathcal{F}$ contains an r-sunflower.

If $\mathcal{F}$ is k-spread then by assumption it contains an r-sunflower. Otherwise, assume $\mathcal{F}$ is not k-spread. By definition, there exists a non-empty set T of some size $|T| = t$ such that

$$|\mathcal{F}_T| > |\mathcal{F}|/k^t \ge k^{n-t}.$$

Note that $\mathcal{F}_T$ is a family of $(n-t)$-sets, so we can apply the induction hypothesis to conclude that $\mathcal{F}_T$ contains an r-sunflower $S_1, \ldots, S_r$. But then $S_1 \cup T, \ldots, S_r \cup T$ is an r-sunflower in $\mathcal{F}$. $\square$

Thus, we are left by analyzing sunflowers in spread families of sets. We will show that spread enough families must contain many disjoint sets. The main engine behind it is the "spread lemma", which shows that for any spread family, a random set contains one of its sets with high probability.

To formally define it, we first need the following definition.

Definition 2.7 (p-biased set)**.** *Let U be a finite set and $p \in (0,1)$. We denote by $Bin(U,p)$ the distribution over subsets $W \subset U$ obtained by including each element $x \in U$ in W independently with probability p.*

Lemma 2.8 (Spread lemma)**.** *Let $\mathcal{F}$ be a family of n-sets defined over a universe U. Let $p \in (0,1)$ and let $W \sim Bin(U,p)$. Assume that $\mathcal{F}$ is k-spread, where $k = cp^{-1} \log n$ for a large enough absolute constant c. Then*

$$\Pr_W[\exists S \in \mathcal{F}, S \subset W] \ge 1/2.$$

We defer the proof of the spread lemma to the next section, and first describe how it allows us to complete the proof of Theorem 2.1. The following lemma shows that any spread enough family contains many pairwise disjoint sets, which form a sunflower with empty kernel.

Lemma 2.9 (Spread families contain many disjoint sets)**.** *Let $\mathcal{F}$ be a family of n-sets which is k-spread for $k = cr \log n$, for a large enough absolute constant c. Then $\mathcal{F}$ contains r pairwise disjoint sets.*

Proof. Let U denote the universe set over which $\mathcal{F}$ is defined. Consider a random partition of U into $2r$ sets $W_1, \ldots, W_{2r}$, where each element $x \in U$ is placed in one of the sets uniformly and independently. Define the events

$$E_i = [\exists S \in \mathcal{F}, S \subset W_i], \qquad i = 1, \ldots, 2r.$$

Observe that the marginal distribution of each W_i is $\mathrm{Bin}(U, p)$ for $p = 1/2r$, so we can apply the spread lemma and get

$$\Pr[E_i] = \Pr_{W_i}[\exists S \in \mathcal{F}, S \subset W_i] \geq 1/2.$$

Let X be a random variable counting the number of successful events. Then

$$\mathbb{E}[X] = \sum \Pr[E_i] \geq (1/2) \cdot 2r = r.$$

This means that there exists a partition $W_1, \ldots, W_{2r}$ with at least r successful events. That is, there are r sets $W_{j_1}, \ldots, W_{j_r}$ in the partition, such that for each W_{j_i} there is a set $S_i \in \mathcal{F}$ contained in W_{j_i}. But this implies that $S_1, \ldots, S_r$ are pairwise disjoint, which concludes the proof. $\square$

2.1 Robust sunflowers

Robust sunflowers (also called quasi-sunflowers in some works) were introduced by Rossman [Ros14] towards proving monotone circuit lower bounds. They were further developed in [LLZ18, ALWZ21] and found applications in monotone circuit lower bounds [Ros14, CKR22] and communication complexity [LMM+22].

Definition 2.10 (Robust sunflower)**.** *Let $\mathcal{F}$ be a family of sets over a universe U. Let $K = \cap_{S \in \mathcal{F}} S$ be their common intersection. We say that $\mathcal{F}$ is a (p, ε)*-robust sunflower *(with kernel K) if $K \notin \mathcal{F}$ and*

$$\Pr_{W \sim Bin(U,p)}[\exists S \in \mathcal{F}, S \subset K \cup W] \geq 1 - \varepsilon.$$

Robust sunflowers contain sunflowers.

Claim 2.11. *Let $\mathcal{F}$ be a $(1/2r, 1/2)$-robust sunflower. Then $\mathcal{F}$ contains an r-sunflower.*

Exercise 2.1. *Prove Claim 2.11; hint: consult the proof of Lemma 2.9.*

The improved bounds for the sunflower lemma (Theorem 2.1) can be viewed as a direct corollary of the following bound on the size of families without robust sunflowers.

Theorem 2.12 (Robust sunflower lemma)**.** *Let $\mathcal{F}$ be a family of n-sets without a (p, ε)-robust sunflower. Then for some constant c,*

$$|\mathcal{F}| \leq (cp^{-1} \log(n/\varepsilon))^n.$$

The first step towards proving Theorem 2.12, is to boost the success probability in the spread lemma (Lemma 2.8) from 1/2 to $1-\varepsilon$.

Lemma 2.13 (Spread lemma, high success probability)**.** *Let $\mathcal{F}$ be a family of n-sets defined over a universe U. Let $p \in (0,1)$ and let $W \sim Bin(U,p)$. Assume that $\mathcal{F}$ is k-spread, where $k = cp^{-1}\log(n/\varepsilon)$ for a large enough absolute constant c. Then*

$$\Pr_W[\exists S \in \mathcal{F}, S \subset W] \geq 1-\varepsilon.$$

Proof. We will prove a slightly weaker bound, with $k = cp^{-1}\log(n)\log(1/\varepsilon)$. Let $t = \log(1/\varepsilon)$. Sample $W_1,\ldots,W_t \sim \mathrm{Bin}(U,p/t)$. Applying the spread lemma (Lemma 2.8) to each W_i gives

$$\Pr_{W_i}[\exists S \in \mathcal{F}, S \subset W_i] \geq 1/2.$$

The claim follows for $W = \cup W_i$. □

Proof of Theorem 2.12. Let $\mathcal{F}$ be a family of n-sets of size $|\mathcal{F}| > k^n$ for $k = cp^{-1}\log(n/\varepsilon)$. We will prove that $\mathcal{F}$ contains a (p,ε)-robust sunflower. Let T be maximal so that

$$|\mathcal{F}_T| \geq |\mathcal{F}|k^{-|T|}.$$

Note that such a T exists because we can choose $T=\emptyset$. Note also that $|T|<n$ by the assumption on $|\mathcal{F}|$, and hence $T \notin \mathcal{F}$. Observe that $\mathcal{F}_T$ is k-spread, as otherwise there exists a nonempty set R disjoint from T, for which $|\mathcal{F}_{R\cup T}| \geq |\mathcal{F}_T|/k^{|R|} \geq |\mathcal{F}|/k^{|R|+|T|}$ which violates the maximality of T. Lemma 2.13 then gives

$$\Pr_{W\sim \mathrm{Bin}(U,p)}[\exists S \in \mathcal{F}_T, S \subset W] \geq 1-\varepsilon.$$

This implies that

$$\mathcal{F}' = \{S \in \mathcal{F} : T \subset S\}$$

is a (p,ε)-robust sunflower with kernel T. □

The bound in Theorem 2.12 turns out to be sharp. Thus, any hope to improve the bound for the sunflower lemma would have to go through a different route.

Exercise 2.2. *Prove that the bound in Theorem 2.12 is sharp, by considering the following example. Let $R_1,\ldots,R_n$ be pairwise disjoint sets, each of size $k=\log(n/c)$ for some $1 \leq c \ll n$ to be chosen later, and let U be their union. Consider the set system*

$$\mathcal{F} = \{S \subset U : |S \cap R_i| = 1 \ \forall i \in [n]\}.$$

Note that $\mathcal{F}$ is a family of n-sets of size $|\mathcal{F}| = k^n$. Let $W \sim Bin(U,1/2)$.

1. *Show that $\mathcal{F}$* does *contain a $(1/2,1/2)$-robust sunflower.*

It turns out that by refining $\mathcal{F}$, we can remove all such robust sunflowers. Set $\varepsilon = 1/c$, and define $\mathcal{F}' \subset \mathcal{F}$ to satisfy

$$|S \cap S'| \leq (1-\varepsilon)n \quad \text{for all } S, S' \in \mathcal{F}', S \neq S'.$$

By a greedy approach, one can find such an $\mathcal{F}'$ of size $|\mathcal{F}'| \geq 2^{-n} k^{(1-\varepsilon)n}$. (You may assume this without proof.) Assume by contradiction $\mathcal{F}'$ does contain a $(1/2, 1/2)$-robust sunflower.

2. *Show that $\Pr_W[W$ is disjoint from some set $R_i] > 1/2$.*

3. *Use this to argue that $\mathcal{F}'$ does not contain a $(1/2, 1/2)$-robust sunflower.*

Observe that by setting $c = \sqrt{n}$, we have $|\mathcal{F}'| \geq (\log n)^{(1-o(1))n}$.

3 Proof of the spread lemma

We prove the spread lemma (Lemma 2.8) in this section. It will be technically more convenient to prove a version where the random set has a fixed size.

Lemma 3.1 (Spread lemma; fixed set size)**.** *Let $\mathcal{F}$ be a family of n-sets defined over a universe U. Let $p \in (0,1)$ and let $W \subset U$ be uniformly sampled of size $|W| = p|U|$. Assume that $\mathcal{F}$ is k-spread, where $k = cp^{-1} \log n$ for a large enough absolute constant c. Then*

$$\Pr_W[\exists S \in \mathcal{F}, S \subset W] \geq 1/2.$$

We first show how Lemma 2.8 is derived from Lemma 3.1.

Proof of Lemma 2.8 given Lemma 3.1. Let $\mathcal{F}$ be a family of n-sets over a universe U. Let U' be a superset of U, and apply Lemma 3.1 to $\mathcal{F}$ and U'. It implies that for randomly chosen $W' \subset U'$ of size $|W'| = p|U'|$, W' contains some set in $\mathcal{F}$ with probability at least $1/2$. Since $\mathcal{F}$ is supported on U, the same holds also for $W = W' \cap U$. Take now U' of growing size; the limiting distribution of W converges to $\mathrm{Bin}(U, p)$. $\square$

We prove Lemma 3.1 in the remainder of this subsection. The main idea in the proof is to show that by removing a small random set, sets in the family can be "compressed" to a smaller size, and then iterate the process. The definition of *minimal fragments* introduced by [PP24] makes the proof streamlined.

Definition 3.2 (Minimal fragments)**.** *Let $\mathcal{F}$ be a family of sets, fix a set $S \in \mathcal{F}$, and let V be an arbitrary set. Consider the family of sets*

$$\mathcal{M}_{\mathcal{F}}(S, V) = \{S' \setminus V : S' \in \mathcal{F}, S' \subset S \cup V\}.$$

The minimal fragment *of S, V (with respect to $\mathcal{F}$), denoted $M_{\mathcal{F}}(S, V)$, is an element in $\mathcal{M}_{\mathcal{F}}(S, V)$ of minimal size (breaking ties arbitrarily). When the family $\mathcal{F}$ is clear from context, we will write it as $M(S, V)$.*

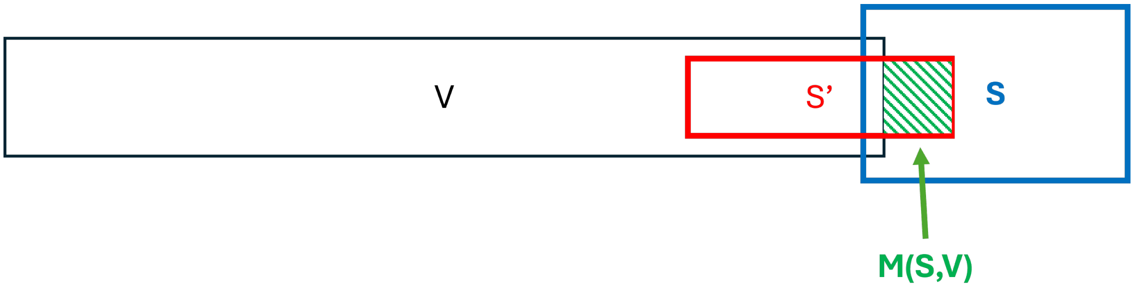

Figure 3: Minimal fragment

Here are some useful observations about minimal fragments:

1. $M(S,V) \subset S$.
2. $M(S,V)$ is disjoint from V.
3. $M(S,V) = \emptyset$ iff there exists $S' \in \mathcal{F}$ with $S' \subset V$.

The next claim establishes an important but less obvious property of minimal fragments.

Claim 3.3. *Let $\mathcal{F}$ be a family of sets, $S \in \mathcal{F}$, and fix a set V. Define $Z = V \cup M(S,V)$ and $\mathcal{F}' = \{S' \in \mathcal{F} : S' \subset Z\}$. Then:*

1. *$\mathcal{F}'$ is not empty.*
2. *Any $S' \in \mathcal{F}'$ satisfies $S' \setminus V = M(S,V)$. In particular, $M(S,V) \subset S'$.*

Proof. By definition, $M(S,V) = S^* \setminus V$ for some set $S^* \in \mathcal{F}$ that satisfies $S^* \subset S \cup V$. This implies that $S^* \subset Z$ which implies $S^* \in \mathcal{F}'$. Next, fix any $S' \in \mathcal{F}'$. Note that $S' \subset Z \subset S \cup V$, so S' was a candidate when defining the minimal fragment $M(S,V)$. The minimality of $M(S,V)$ implies that $|S' \setminus V| \geq |M(S,V)|$. On the other hand, since $S' \subset Z$ we have $S' \setminus V \subset M(S,V)$. Overall, we must have $S' \setminus V = M(S,V)$. □

The next claim analyzes the typical size of fragments. It shows that for spread enough families, the minimal fragments are typically small.

Claim 3.4. *Let $\mathcal{F}$ be a family of sets of size at most n, defined over a universe U. Let $q \in (0,1)$, and sample $V \subset U$ uniformly of size $|V| = q|U|$. Assume that $\mathcal{F}$ is k-spread for $k \geq cq^{-1}$, for a large enough constant c. Then*

$$\Pr_{S \in \mathcal{F}, V}[|M(S,V)| \geq n/2] \leq 100^{-n}.$$

Proof. Let $N = |U|$. Define the family of "bad" sets:

$$\mathcal{B} := \{(S,V) : S \in \mathcal{F}; V \subset U, |V| = qN; |M(S,V)| \geq n/2\}$$

and observe that

$$\Pr_{S,V}[|M(S,V)| \geq n/2] = \frac{|\mathcal{B}|}{|\mathcal{F}|\binom{N}{qN}}.$$

Thus, our goal is to upper bound the size of $\mathcal{B}$. In order to do so, we define an *encoding scheme*. Given $(S,V) \in \mathcal{B}$ define:

1. $Z = V \cup M(S,V)$.
2. $\mathcal{F}' = \{S' \in \mathcal{F} : S' \subset Z\}$. By Claim 3.3 (1), $\mathcal{F}'$ is not empty.
3. Fix $S' \in \mathcal{F}'$ arbitrarily (say, first in some pre-defined order).
4. By Claim 3.3 (2), $M(S,V) \subset S'$. Specify it as a subset $M \subset S'$.
5. Given $M = M(S,V)$, we have $M \subset S$ and hence $S \setminus M \in \mathcal{F}_M$.

We define the encoding of (S,V) as

$$\phi(S,V) = (Z, S', M, S \setminus M).$$

Note that we can decode (S,V) given $\phi(S,V)$ since $S = M \cup (S \setminus M)$ and $V = Z \setminus M$. Thus, in order to bound the size of $\mathcal{B}$, it suffices to bound the number of possible encodings.

Let us bound the number of encodings with a fixed minimal fragment size $|M| = m$. These can be bound by:

1. The number of choices for Z is $\binom{N}{qN+m} \leq \binom{N}{qN} q^{-m}$.
2. The set S' is uniquely defined given Z.
3. Given S', there are at most $\binom{n}{m} \leq 2^n$ options for $M = M(S,V) \subset S'$.
4. Given M, there are $|\mathcal{F}_M| \leq |\mathcal{F}| k^{-m}$ choices for $S \setminus M$ (here is where we are using the assumption that $\mathcal{F}$ is k-spread!)

Overall, the number of choices is bound by:

$$\binom{N}{qN} q^{-m} \cdot 2^n \cdot |\mathcal{F}| k^{-m} = \binom{N}{qN} |\mathcal{F}| \cdot 2^n (kq)^{-m}.$$

Summing over $n/2 \leq m \leq n$ gives

$$\Pr_{S \in \mathcal{F}, V}[|M(S,V)| \geq n/2] = \frac{|\mathcal{B}|}{|\mathcal{F}|\binom{N}{qN}} \leq \sum_{m=n/2}^{n} 2^n (kq)^{-m} \leq \sum_{m=n/2}^{n} (4/kq)^m.$$

Taking $k = cq^{-1}$ for large enough c, this is at most 100^{-n}. $\square$

In order to conclude the proof of Lemma 3.1, we apply Claim 3.4 iteratively. In the proof, it will be convenient to allow families to have repeated sets; it can be verified that everything we did so far extends to this case (with the same proofs). The next claim establishes the inductive step, going from families of sets of size $\leq n$ to sets of size $\leq n/2$, with only a minimal loss in spreadness.

Claim 3.5. *Let $\mathcal{F}$ be a family of sets of size at most n, defined over a universe U. Let $q \in (0,1)$, and sample $V \subset U$ uniformly of size $|V| = q|U|$. Assume that $\mathcal{F}$ is k-spread for $k \geq cq^{-1}$, for a large enough constant c. Define*

$$\mathcal{F}' = \{M(S,V) : S \in \mathcal{F}, |M(S,V)| \leq n/2\}.$$

Then with probability at least $1 - 10^{-n}$ over the choice of V, the following hold:

1. *$\mathcal{F}'$ is a family of sets of size $\leq n/2$.*
2. *$|\mathcal{F}'| \geq (1 - 10^{-n})|\mathcal{F}|$.*
3. *$\mathcal{F}'$ is k'-spread, where $k' = (1 - 10^{-n})k$.*

Proof. Apply Claim 3.4, and note that by the conclusion,

$$\mathbb{E}[|\mathcal{F}'|] \geq (1 - 100^{-n})|\mathcal{F}|.$$

By Markov's inequality, with probability at least $1 - 10^{-n}$ over the choice of V, we have $|\mathcal{F}'| \geq (1 - 10^{-n})|\mathcal{F}|$. The last item follows by the following claim, left as an exercise. $\square$

Exercise 3.1. *Let $\mathcal{F} = \{S_1, \ldots, S_m\}$ be a family of sets (possibly with repetitions) that is k-spread. Prove the following.*

1. *Let $\mathcal{F}' = \{S'_1, \ldots, S'_m\}$ be a family of sets with $S'_i \subset S_i$. Then $\mathcal{F}'$ is also k-spread.*
2. *Let $\mathcal{F}' \subset \mathcal{F}$ of size $|\mathcal{F}'| \geq (1-\varepsilon)|\mathcal{F}|$. Then $\mathcal{F}'$ is $((1-\varepsilon)k)$-spread.*

The next claim shows that if we compute minimal fragments with respect to several sets V_i, and obtain an empty set as a result, then the union $\cup V_i$ contains a set in the family.

Claim 3.6. *Let $\mathcal{F}_0$ be a family, and let $V_1, \ldots, V_t$ be sets. For $i = 1, \ldots, t$ define*

$$\mathcal{F}_i = \{M_{\mathcal{F}_{i-1}}(S, V_i) : S \in \mathcal{F}_{i-1}\}.$$

Assume that $\mathcal{F}_t$ contains the empty set. Then $\cup V_i$ contains some set in $\mathcal{F}_0$.

Proof. We prove this by induction on t. For $t = 1$, if $\mathcal{F}_1$ contains the empty set, then $M_{\mathcal{F}_0}(S, V_1) = \emptyset$ for some $S \in \mathcal{F}_0$. By definition, $M_{\mathcal{F}_0}(S, V_1) = S_0 \setminus V_1$ for some $S_0 \in \mathcal{F}_0$ (and moreover, $S_0 \subset S \cup V_1$, but we won't use this). We then have $S_0 \subset V_1$.

For general t, by applying the claim to $\mathcal{F}_1$ and $V_2, \ldots, V_t$, we get that some $S_1 \in \mathcal{F}_1$ satisfies $S_1 \subset V_2 \cup \ldots \cup V_t$. But by definition $S_1 = S_0 \setminus V_1$ for some $S_0 \in \mathcal{F}_0$, which means that $S_0 \subset \cup V_i$. $\square$

The proof of Lemma 3.1 follows by applying Claim 3.5 iteratively, and then Claim 3.6.

Proof of Lemma 3.1. We assume for simplicity that $n = 2^m$ is a power of two. It will be convenient to define $e(t) = 1 - 10^{-n/2^t}$. We note for later use that

$$\prod_{i=0}^{m-1} e(i) \geq 1 - \sum_{i=0}^{m-1} 10^{-n/2^i} \geq 0.8.$$

Next, let $q = p/\log n$. Our choice of k satisfies $k \geq cq^{-1}$ so we can apply Claim 3.5. Let $\mathcal{F}_0$ be a family of sets of size $\leq n$ which is k-spread. Applying Claim 3.5 gives that for a randomly sampled $V_1 \subset U$ of size $|V_1| = q|U|$, with probability at least $e(0)$, we have that

$$\mathcal{F}_1 = \{M_{\mathcal{F}_0}(S, V_1) : S \in \mathcal{F}_0, |M(S, V_1)| \leq n/2\}$$

satisfies:

1. It is a family of sets of size $\leq n/2$.
2. Its size is $|\mathcal{F}_1| \geq e(0)|\mathcal{F}_0|$.
3. It is k'-spread for $k' = e(0)k \geq 0.8k$.

We continue to apply Claim 3.5 iteratively to $\mathcal{F}_1$ and $U_1 = U \setminus V_1$. For $t \geq 1$, sample $V_t \subset U_t$ of size $|V_t| = q|U_t|$, and define the family $\mathcal{F}_t$ as

$$\mathcal{F}_t = \{M_{\mathcal{F}_{t-1}}(S, V_t) : S \in \mathcal{F}_{t-1}, |M(S, V_t)| \leq n/2^t\}.$$

With probability $e(t-1)$ over the choice of V_t, it satisfies:

1. It is a family of sets of size $\leq n/2^t$.
2. Its size is $|\mathcal{F}_t| \geq e(t-1)|\mathcal{F}_{t-1}|$.
3. It is k'-spread for $k' = \prod_{i=0}^{t-1} e(i)k \geq 0.8k$.

Overall, we repeat this $m = \log n$ times, and get the following: we sample $V_1, \ldots, V_m \subset U$ pairwise disjoint, each of size $|V_i| = q|U_i| = q(1-q)^{i-1}|U| \leq q|U|$; with probability at least $\prod_{i=0}^{m} e(i) \geq 0.8$, all the families $\mathcal{F}_t$ have size $\geq 0.8|\mathcal{F}_0|$. In particular, the last set $\mathcal{F}_m$ is non-empty, and hence contains the empty set. The lemma now follows from Claim 3.6. $\square$

Exercise 3.2. *Prove Lemma 2.13 with the improved bound of* $k = cp^{-1}\log(n/\varepsilon)$.

4 Threshold phenomena, and a proof of the Kahn-Kalai conjecture

Consider the following problem in random graph theory. Let $G(n,p)$ be a random graph on n nodes, where each edge is chosen independently with probability p. For what p should we expect $G(n,p)$ to contain a Hamiltonian cycle with high probability?

We are going to use this as a running example for estimating the *critical probability* of events (formally, of increasing events, which are precisely defined as containing some structure).

Definition 4.1 (Critical probability). *Let $\mathcal{F}$ be a family of sets over a universe U. Given $p \in (0,1)$ define*

$$\mu_p(\mathcal{F}) \coloneqq \Pr_{W \sim Bin(U,p)}[\exists S \in \mathcal{F}, S \subset W].$$

The critical probability *$p_C(\mathcal{F})$ is the minimal p such that $\mu_p(\mathcal{F}) \geq 1/2$.*

Example 4.2 (Hamiltonian cycles). *In the example of Hamiltonian cycles, we have $U = \binom{[n]}{2}$ all edges of an n-vertex complete graph. Given a permutation $\pi \in S_n$ define*

$$S_\pi \coloneqq \{\{i, \pi(i)\} : i \in [n]\}.$$

The family of Hamiltonian cycles is

$$\mathcal{F}_{Ham} \coloneqq \{S_\pi : \pi \text{ is a cycle}\}.$$

Then $\mu_p(\mathcal{F}_{Ham})$ is precisely the probability that $G(n,p)$ contains a Hamiltonian cycle.

One way to bound the critical probability is by counting the expected number of $S \in \mathcal{F}$ that appear in W.

Definition 4.3 ("Naive" expectation threshold). *Let $\mathcal{F}$ be a family of sets. The "naive" expectation threshold, denoted $p_E^*(\mathcal{F})$, is the value of p for which*

$$\sum_{S \in \mathcal{F}} p^{|S|} = 1/2.$$

Claim 4.4. *$p_C(\mathcal{F}) \geq p_E^*(\mathcal{F})$.*

Proof. Let $\mathcal{F}$ be defined over universe U. Let $p = p_E^*(\mathcal{F})$ and sample $W \sim \mathrm{Bin}(U,p)$. Then

$$\mu_p(\mathcal{F}) = \Pr_W[\exists S \in \mathcal{F}, S \subset W] \leq \sum_{S \in \mathcal{F}} \Pr_W[S \subset W] = \sum_{S \in \mathcal{F}} p^{|S|} = 1/2$$

so we must have $p_C(\mathcal{F}) \geq p$. $\square$

Going again to the example of Hamiltonian cycles, we have

$$\sum_{S \in \mathcal{F}_{\mathrm{Ham}}} p^{|S|} = (n-1)! \cdot p^n \approx (pn/e)^n$$

which shows that $p_E^*(\mathcal{F}_{\mathrm{Ham}}) = c/n$ for some constant c, and hence

$$p_C(\mathcal{F}_{\mathrm{Ham}}) \geq \frac{c}{n}.$$

How close is this to the true value? As we shall see, there are two obstacles: isolated nodes, and covers.

In $G(n,p)$, as long as $p \leq \frac{c \log n}{n}$, then with high probability there is an isolated node. So in particular, there cannot be a Hamiltonian cycle. But this is merely a $\log n$ factor larger than the lower bound given by the naive expectation threshold. Can this always be the case? Not always, as the following example shows.

Example 4.5 (Unions of Hamiltonian cycles and 4-cliques)**.** *Given $T \subset [n]$ define the clique on T as*

$$K_T = \{\{i,j\} : i,j \in T, i \neq j\}.$$

Define the family of 4-cliques as

$$\mathcal{F}_{K_4} = \{K_T : |T| = 4\}.$$

Define the family of unions of Hamiltonian cycles and 4-cliques as

$$\mathcal{F}_{Ham,K_4} = \{S \cup S' : S \in \mathcal{F}_{Ham}, S' \in \mathcal{F}_{K_4}\}.$$

The naive expectation thresholds for the families are

$$p_E^*(\mathcal{F}_{K_4}) \approx n^{-2/3}, \quad p_E^*(\mathcal{F}_{Ham,K_4}) \approx 1/n.$$

Clearly for $G(n,p)$ to contain the union of a Hamiltonian cycle and a 4-clique, it must contain a 4-clique, which means that $p_C(\mathcal{F}_{Ham,K_4}) \geq cn^{-2/3}$; this means that the "naive" bound given by $p_E^*(\mathcal{F}_{Ham,K_4}) \approx 1/n$ is far from the truth.

Exercise 4.1. *Verify the calculations of $p_E^*(\mathcal{F}_{K_4})$ and $p_E^*(\mathcal{F}_{Ham,K_4})$.*

We see that the "naive" expectation threshold can be far from the truth if there is a "hard core" that controls the threshold behavior. This motivates the following definitions.

Definition 4.6 (Span)**.** *Let $\mathcal{G}$ be a family of sets over a universe U. Its* span*, denoted $\langle \mathcal{G} \rangle$ is the family of all supersets of some set in $\mathcal{G}$,*

$$\langle \mathcal{G} \rangle \coloneqq \{S : \exists T \in \mathcal{G}, T \subset S\}.$$

Definition 4.7 (Cover)**.** *Let $\mathcal{F}, \mathcal{G}$ be families of sets over a universe U. We say that $\mathcal{G}$* covers *$\mathcal{F}$ if $\mathcal{F} \subset \langle \mathcal{G} \rangle$.*

Claim 4.8. *Let $\mathcal{G}$ be a cover of $\mathcal{F}$. Then $p_C(\mathcal{F}) \geq p_C(\mathcal{G})$.*

Proof. Let $W \sim \text{Bin}(U, p)$. Because every $S \in \mathcal{F}$ contains some $T \in \mathcal{G}$ we have

$$\mu_p(\mathcal{F}) = \Pr_W[\exists S \in \mathcal{F} : S \subset W] \leq \Pr_W[\exists T \in \mathcal{G} : T \subset W] = \mu_p(\mathcal{G})$$

and hence $p_C(\mathcal{F}) \geq p_C(\mathcal{G})$. $\square$

The (non-naive) expectation threshold of $\mathcal{F}$ is the maximum over all naive expectation thresholds of covers of $\mathcal{F}$.

Definition 4.9 (Expectation threshold)**.** *The* expectation threshold *of a family $\mathcal{F}$, denoted $p_E(\mathcal{F})$, is*

$$p_E(\mathcal{F}) \coloneqq \max\{p_E^*(\mathcal{G}) : \mathcal{G} \text{ is a cover of } \mathcal{F}\}.$$

Claim 4.10 (Expectation threshold lower bound)**.** $p_C(\mathcal{F}) \geq p_E(\mathcal{F})$.

Proof. For all covers $\mathcal{G}$ of $\mathcal{F}$ we have $p_C(\mathcal{F}) \geq p_C(\mathcal{G}) \geq p_E^*(\mathcal{G})$. $\square$

In applications, computing $p_E^*(\mathcal{F})$ is straightforward, but computing $p_E(\mathcal{F})$ seems daunting - how can you go over all possible covers? The following useful claim shows how to upper bound it using spreadness. In most applications, this upper bound is close to the lower bound given by $p_E^*(\mathcal{F})$, which allows one to estimate $p_E(\mathcal{F})$ with good accuracy.

Claim 4.11. *Let $\mathcal{F}$ be a family of sets which is k-spread. Then $p_E(\mathcal{F}) \leq 1/k$.*

Proof. Let $\mathcal{G}$ be a cover of $\mathcal{F}$. This implies that

$$\sum_{T \in \mathcal{G}} |\mathcal{F}_T| \geq |\mathcal{F}|.$$

As we assume $\mathcal{F}$ is k-spread, we have $|\mathcal{F}_T| \leq k^{-|T|}|\mathcal{F}|$. Hence

$$\sum_{T \in \mathcal{G}} k^{-|T|} \geq 1$$

which implies that $p_E^*(\mathcal{G}) \leq 1/k$. $\square$

Exercise 4.2. *Prove that $p_E(\mathcal{F}_{Ham}) = \Theta(1/n)$.*

Kahn and Kalai [KK07] made a bold conjecture - the expectation threshold controls the critical probability, up to the $\log n$ gap needed to handle isolated nodes.

Conjecture 4.12 (Kahn-Kalai [KK07])**.** *Let $\mathcal{F}$ be a family of sets of size $\leq n$. Then $p_C(\mathcal{F}) \leq c \cdot p_E(\mathcal{F}) \cdot \log n$ for some absolute constant c.*

Amazingly, this conjecture is true, as proved by Park and Pham [PP24].

Theorem 4.13 (Park-Pham [PP24])**.** *The Kahn-Kalai conjecture is true.*

In some cases, the extra $\log n$ factor is not needed. It is an active line of research to figure out exactly when it can be replaced by a smaller factor. In the case of Hamiltonian cycles, however, it cannot be reduced.

Exercise 4.3. *Prove that $p_C(\mathcal{F}_{Ham}) = \Theta\left(\frac{\log n}{n}\right)$.*

4.1 Proof of the Kahn-Kalai conjecture

The proof of the Kahn-Kalai conjecture builds on and refines the proof of the spread lemma. In fact, the spread lemma can be proven as a corollary of it.

Exercise 4.4. *Prove the spread lemma (Lemma 2.8) from Theorem 4.13.*

Recall the definition of minimal fragments, restated below.

Definition 3.2 (Minimal fragments)**.** *Let $\mathcal{F}$ be a family of sets, fix a set $S \in \mathcal{F}$, and let V be an arbitrary set. Consider the family of sets*

$$\mathcal{M}_{\mathcal{F}}(S, V) = \{S' \setminus V : S' \in \mathcal{F}, S' \subset S \cup V\}.$$

The minimal fragment *of S, V (with respect to $\mathcal{F}$), denoted $M_{\mathcal{F}}(S, V)$, is an element in $\mathcal{M}_{\mathcal{F}}(S, V)$ of minimal size (breaking ties arbitrarily). When the family $\mathcal{F}$ is clear from context, we will write it as $M(S, V)$.*

Given a family $\mathcal{F}$ of sets of size $\leq n$ and a set V, define two new families:

$$\mathcal{F}^{large}(V) := \{M(S, V) : S \in \mathcal{F}, |M(S, V)| > n/2\}$$
$$\mathcal{F}^{small}(V) := \{M(S, V) : S \in \mathcal{F}, |M(S, V)| \leq n/2\}$$

Note that $\mathcal{F}^{large}(V) \cup \mathcal{F}^{small}(V)$ is a cover of $\mathcal{F}$, since $M(S, V) \subset S$ for any V. The following lemma is the main engine behind the proof of Theorem 4.13

Lemma 4.14. *Let $p \in (0, 1)$, $q = cp$ for a large enough constant c. Let $\mathcal{F}$ be a family of sets of size $\leq n$ defined over a universe U. Let $V \sim Bin(U, q)$. Then*

$$\mathbb{E}_V \left[\sum_{T \in \mathcal{F}^{large}(V)} p^{|T|} \right] \leq 100^{-n}.$$

Proof sketch. First, it suffices to prove the lemma when we sample $V \subset U$ uniformly of size $|V| = q|U|$, and then transfer to the case of $V \sim \text{Bin}(U, q)$, similarly to how we did it in the proof of the spread lemma.

Given a uniformly sampled V of size $|V| = q|U|$, the proof is essentially identical to the proof of Claim 3.4. The main observation is that in its proof, for any V we in fact construct a cover $\mathcal{G}$ for $\mathcal{F}$, where each set $T \in \mathcal{G}$ is weighed by $|\mathcal{F}_T|/|\mathcal{F}|$, which by assumption there is at most $k^{-|T|}$. Here, we explicitly weigh it by $p^{|T|}$, which removes the need to make any assumption on $\mathcal{F}$. □

Exercise 4.5. *Give a complete proof of Lemma 4.14.*

Building on Lemma 4.14, we prove the Kahn-Kalai conjecture.

Proof of Theorem 4.13. Let $p = p_E(\mathcal{F})$, $q = cp$ for a large enough constant c. Let $t = \log n$, sample $V_1, \ldots, V_t \sim \mathrm{Bin}(U, q)$ independently, and define $V = \cup V_i$. We will show that

$$\Pr[\exists S \in \mathcal{F}, S \subset V] \geq 1/2$$

which will show that $p_C(\mathcal{F}) \leq qt = cp \log n$ as claimed.

Let $\mathcal{F}_0 = \mathcal{F}$ and for $i = 1, \ldots, t$ define

$$\mathcal{F}_i \coloneqq \mathcal{F}_{i-1}^{small}(V_i), \quad \mathcal{G}_i \coloneqq \mathcal{F}_{i-1}^{large}(V_i).$$

Note that $\mathcal{F}_i \cup \mathcal{G}_i$ is a cover of $\mathcal{F}_{i-1}$, and that all sets in $\mathcal{F}_i$ have size $\leq n/2^i$. Define $\mathcal{G} = \cup \mathcal{G}_i$. Applying Lemma 4.14 gives

$$\mathbb{E} \sum_{T \in \mathcal{G}} p^{|T|} \leq \mathbb{E} \sum_i \sum_{T \in \mathcal{G}_i} p^{|T|} \leq \sum_i 100^{-n/2^i} \leq 0.1.$$

Next, consider the event

$$E = [\mathcal{F}_t \text{ is empty}].$$

Assume first that $\Pr[E] \geq 1/2$. Observe that for each choice of $V_1, \ldots, V_t$ for which E holds, we obtain that $\mathcal{G}$ is a cover of $\mathcal{F}$. Moreover,

$$\mathbb{E}\left[\sum_{T \in \mathcal{G}} p^{|T|} \middle| E \right] \leq \frac{\mathbb{E} \sum_{T \in \mathcal{G}} p^{|T|}}{\Pr[E]} \leq 0.2$$

which means by averaging, that for some choice of $V_1, \ldots, V_t$, we obtain a cover $\mathcal{G}$ of $\mathcal{F}$ with $p_E^*(\mathcal{G}) > p$. This violates the choice of $p = p_E(\mathcal{F})$.

Thus, we must have $\Pr[E] < 1/2$. Note that whenever E does not hold, by definition $\mathcal{F}_t$ is not empty. This means that it contains the empty set, which can only happen when V contains some set $S \in \mathcal{F}$. We thus conclude the proof since

$$\Pr[\exists S \in \mathcal{F}, S \subset V] \geq \Pr[E^c] \geq 1/2.$$

□

5 Application: lower bounds for monotone circuits

We describe a surprising application of the spread lemma in complexity theory, where it is used to prove strong lower bounds for monotone circuits that distinguish between graphs containing k-cliques from random graphs unlikely to contain a k-clique.

5.1 Monotone functions and circuits

Partial order. Given $x, y \in \{0,1\}^n$, we say $x \leq y$ if $x_i \leq y_i$ for all $i \in [n]$; and $x < y$ if $x \leq y$ and $x \neq y$.

Monotone functions. Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. The function is called *monotone* if it satisfies: $x \le y \Rightarrow f(x) \le f(y)$. Given two monotone functions f, g, we say that $f \le g$ if $f(x) \le g(x)$ for all x.

A *min-term* of a monotone function is minimal 1-input; namely, x is a min-term of f if $f(x) = 1$ and $f(y) = 0$ for all $y < x$. We denote by $\mathcal{M}(f)$ to be the set of min-terms of f. Then:

$$f(x) = \bigvee_{M \in \mathcal{M}(f)} \bigwedge_{i \in M} x_i.$$

Example 5.1 (Majority)**.** *The majority function on* 3 *bits is monotone:*

$$Maj_3(x_1, x_2, x_3) = \mathbf{1}[x_1 + x_2 + x_3 \ge 2].$$

Its min-terms are

$$\mathcal{M}(Maj_3) = \{110, 101, 011\}.$$

It can be computed as

$$Maj_3(x_1, x_2, x_3) = (x_1 \wedge x_2) \vee (x_1 \wedge x_3) \vee (x_2 \wedge x_3).$$

Monotone circuits. A natural model of computation for monotone functions are *monotone circuits*. A monotone circuit $\mathcal{C}$ is a directed acyclic graph defining a monotone computation:

- Inputs (nodes with in-degree 0) compute the inputs $x_1, \dots, x_n$.
- Intermediate nodes have in-degree 2, and compute either the AND or the OR of their inputs.
- There is a single output node (node with out-degree 0).

We identify each node $v \in \mathcal{C}$ in the circuit with the monotone function computed by this node $v : \{0,1\}^n \to \{0,1\}$; the function computed by the circuit $\mathcal{C} : \{0,1\}^n \to \{0,1\}$ is the function computed by the single output gate. The *size* of a circuit is the number of gates in it.

5.2 Lower bounds for k-Clique

Let $k \le m$, and set $n = \binom{m}{2}$. We identify inputs $x \in \{0,1\}^n$ as describing possible edges of an m-vertex graph. Our main object of study is the k-clique function:

$$\text{Clique}_{m,k}(x) = \mathbf{1}[\text{graph given by } x \text{ contains a } k\text{-clique}]$$

Note that $\text{Clique}_{m,k}$ is a monotone function. The trivial way to check if a graph contains a k-clique is to check all possible k-cliques, the number of which is $\binom{m}{k}$. This gives a monotone circuit of size $\approx m^k$. In a breakthrough work, Razborov [Raz85] proved that any monotone

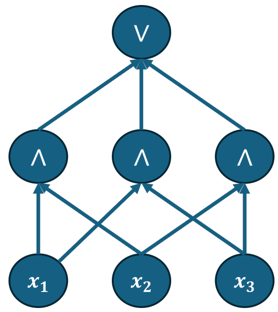

Figure 4: Monotone circuit for Majority on 3 bits

circuit that computes $\text{Clique}_{m,k}$ must have exponential size; concretely, $m^{\Omega(\sqrt{k})}$. This bound was later improved and sharpened by follow up works to $m^{\Omega(k)}$. Proving it for general (non-monotone) computations would imply $P \neq NP$, currently very much out of reach.

We will focus on a weaker goal, where as we will see the spread lemma plays a prominent role. Let $G(m,p)$ denote the Erdős-Renyi random graph model, where each edge is present independently with probability p. The probability $p_c = \Theta(m^{-\frac{2}{k-1}})$ is the critical probability for $G(m,p_c)$ to contain a k-clique. We will choose p slightly below it: $p = m^{-\frac{2}{k-1}-\delta}$ for some $\delta = \delta(k)$ small enough to be defined later. Consider the following "promise" problem: we want to distinguish, with high probability, between two distributions on graphs:

1. No distribution $\mathcal{N}$: graphs sampled from $G(m,p)$ (which with high probability *do not* contain a k-clique)

2. Yes distribution $\mathcal{Y}$: random k-clique

Rossman [Ros14] proved that monotone circuits require exponential size to solve this.

Theorem 5.2 ([Ros14])**.** *Let $\mathcal{C}$ be a monotone circuit that satisfies*

$$\Pr[\mathcal{C}(\mathcal{N}) = 0] \geq 0.9, \qquad \Pr[\mathcal{C}(\mathcal{Y}) = 1] \geq 0.9.$$

Then $\mathcal{C}$ has size $\Omega(m^{k/4})$.

Note that *non-monotone* circuits can easily solve this problem; for example, by counting the number of edges, and returning 1 only if there are exactly $\binom{k}{2}$ edges. This gives circuits of size $\text{poly}(m)$, independent of k.

5.3 Approximating monotone functions

We consider monotone functions defined over $n = \binom{m}{2}$ bits. We identify the input space $U = [n]$ with $\binom{[m]}{2}$. Note that the distribution $\mathcal{N}$, which we defined as $G(m,p)$, is equivalent to $\text{Bin}(U,p)$.

In this section, fix a family $\mathcal{F}$ of subsets of $[n]$ that is downward closed (namely, if $S \in \mathcal{F}$ then all subsets of S are in $\mathcal{F}$), and an error parameter $\varepsilon \ll 1/|\mathcal{F}|$.

Definition 5.3 (Closed monotone functions)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a monotone function. We say that f is* $(\mathcal{F}, \varepsilon)$-closed *if*

$$\Pr[f(x \vee \mathcal{N}) = 1] \geq 1 - \varepsilon \Rightarrow f(x) = 1 \qquad \forall x \in \mathcal{F}.$$

As $\mathcal{F}, \varepsilon$ are fixed throughout, we simply say that f is closed.

We first show that for any monotone function f, there is a closed monotone function $g \geq f$ which approximates it over inputs sampled from $\mathcal{N}$.

Lemma 5.4. *Let $f : \{0,1\}^n \to \{0,1\}$ be a monotone function. There exists a closed monotone function g satisfying $g \geq f$ and*

$$\Pr[g(\mathcal{N}) = 1] \leq \Pr[f(\mathcal{N}) = 1] + \varepsilon|\mathcal{F}|.$$

Proof. We define a sequence of functions $f_0, f_1, \ldots$ as follows. Let $f_0 = f$. If f_i is not closed, then by definition there is $x_i \in \mathcal{F}$ such that

$$f_i(x_i) = 0, \quad \Pr[f_i(x_i \vee \mathcal{N}) = 1] \geq 1 - \varepsilon.$$

Define $f_{i+1}(x) = f_i(x) \vee \mathbf{1}[x_i \leq x]$; it is a monotone function and satisfies $f_{i+1} \geq f_i$. This process must end after at most $\mathcal{F}$ steps; we let g be the final function. Next, note that

$$\begin{aligned}
\Pr[f_{i+1}(\mathcal{N}) = 1] - \Pr[f_i(\mathcal{N}) = 1] &= \Pr_{x \sim \mathcal{N}}[f_i(x) = 0 \wedge f_{i+1}(x) = 1] \\
&= \Pr_{x \sim \mathcal{N}}[f_i(x) = 0 \wedge x_i \leq x] \\
&\leq \Pr_{x \sim \mathcal{N}}[f_i(x_i \vee x) = 0] \\
&= \Pr[f_i(x_i \vee \mathcal{N}) = 0] \leq \varepsilon.
\end{aligned}$$

Summing over i completes the proof. □

Observe that the space of closed monotone functions is closed under intersection: if f, g are closed monotone function then so is $f \wedge g$.

Exercise 5.1. *Prove this observation: if f, g are closed monotone functions, then $f \wedge g$ is also a closed monotone function.*

This allows us to define the closure of a monotone function.

Definition 5.5 (Closure of monotone function)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a monotone function. The* closure *of f, denoted f^*, is the minimal closed monotone function satisfying $f^* \geq f$; equivalently,*

$$f^* := \bigwedge \{g : g \text{ closed monotone function}, g \geq f\}.$$

Note that Lemma 5.4 implies that

$$\Pr[f^*(\mathcal{N}) = 1] \leq \Pr[f(\mathcal{N}) = 1] + \varepsilon|\mathcal{F}|.$$

The following lemma generalized this.

Lemma 5.6. *Let f be a monotone function. Then for any input x,*

$$\Pr[f^*(x \vee \mathcal{N}) = 1] \leq \Pr[f(x \vee \mathcal{N}) = 1] + \varepsilon|\mathcal{F}|.$$

Exercise 5.2. *Prove Lemma 5.6.*

5.4 Closed monotone functions and robust sunflowers

Next, we describe a connection between min-terms of closed monotone functions and robust sunflowers, which allows us to bound the number of min-terms.

Lemma 5.7. *Let f be a closed monotone function. Then $\mathcal{M}(f) \cap \mathcal{F}$ does not contain a (p, ε)-robust sunflower.*

Proof. Assume towards a contradiction there is a (p, ε)-robust sunflower $\mathcal{F}' \subset \mathcal{M}(f) \cap \mathcal{F}$ with kernel K. By the definition of robust sunflowers, this means that

$$\Pr[f(K \vee \mathcal{N}) = 1] \geq 1 - \varepsilon.$$

Next, note that $K \in \mathcal{F}$ since for any $M \in \mathcal{F}'$, we have $K < M$ and $\mathcal{F}$ is downward closed. Thus, as we assume f is closed, this implies that $f(K) = 1$. But then elements $M \in \mathcal{F}'$ cannot be min-terms since $K < M$ and $f(K) = 1$. □

Recall that $\mathcal{M}(f)$ denotes the set of min-terms of f; we denote by $\mathcal{M}_\ell(f) = \{S \in \mathcal{M}(f) : |S| = \ell\}$ the set of min-terms of f of size ℓ. Applying the bound for families without robust sunflowers (Theorem 2.12) gives the following corollary.

Corollary 5.8. *Let f be a closed monotone function. Then for all ℓ,*

$$|\mathcal{M}_\ell(f) \cap \mathcal{F}| \leq (cp^{-1} \log(\ell/\varepsilon))^\ell.$$

5.5 Approximating monotone circuits

We next move to approximate monotone circuits. Let $\mathcal{C}$ be a monotone circuit on n input bits. We define its *circuit closure* $\overline{\mathcal{C}}$ by iteratively replacing each node with its closure. That is:

- If v is an input node (computing some input x_i) then $\overline{v} = v^* = v$.
- If $v = v_1 \wedge v_2$ then $\overline{v} = (\overline{v_1} \wedge \overline{v_2})^* = \overline{v_1} \wedge \overline{v_2}$.
- If $v = v_1 \vee v_2$ then $\overline{v} = (\overline{v_1} \vee \overline{v_2})^*$.

Equivalently, we can obtain $\overline{\mathcal{C}}$ by replacing each OR operation $\vee$ in $\mathcal{C}$ with the operation $\overline{\vee}$ defined as

$$f \overline{\vee} g = (f \vee g)^*.$$

Define $\overline{\mathcal{C}}$ to be $\overline{v}$ for the output node v of $\mathcal{C}$. Note that for all nodes v,

$$v \leq v^* \leq \overline{v}.$$

We next show that if $\mathcal{C}$ is small (has few gates), then $\overline{\mathcal{C}}$ approximates $\mathcal{C}$ well. We denote by $|\mathcal{C}|$ the size (number of gates) in $\mathcal{C}$.

Lemma 5.9. *Let $\mathcal{C}$ be a monotone circuit. Then for all inputs x,*

$$\Pr[\overline{\mathcal{C}}(x \vee \mathcal{N}) = 1] \leq \Pr[\mathcal{C}(x \vee \mathcal{N}) = 1] + \varepsilon|\mathcal{F}||\mathcal{C}|.$$

Proof. Let $s = |\mathcal{C}|$ denote the size of $\mathcal{C}$. Let $v_1, \ldots, v_s$ denote the nodes of $\mathcal{C}$ in topological order, from the inputs $v_1, \ldots, v_n$ to the output v_s. The proof proceeds by a hybrid argument.

For $i = 0, \ldots, s$ define a "hybrid" circuit $\mathcal{C}^i$, obtained from $\mathcal{C}$ by replacing $\vee$ with $\overline{\vee}$ for the first i gates. Note that $\mathcal{C}^0 = \mathcal{C}$ and $\mathcal{C}^s = \overline{\mathcal{C}}$. Let v_j^i denote the function computed by the j-th node in $\mathcal{C}^i$. Note that by construction, $v_j^i = v_j^{i-1}$ for $j < i$ and $v_j^i \geq v_j^{i-1}$ for $j \geq i$. Moreover, $v_i^i = (v_i^{i-1})^*$.

Observe that for any input y, if $\mathcal{C}^i(y) \neq \mathcal{C}^{i-1}(y)$ then it must be that $v_i^i(y) \neq v_i^{i-1}(y)$. Applying Lemma 5.6 gives

$$\Pr[\mathcal{C}^i(x \vee \mathcal{N}) = 1] - \Pr[\mathcal{C}^{i-1}(x \vee \mathcal{N}) = 1] \leq \Pr[(v_i^{i-1})^*(x \vee \mathcal{N}) = 1] - \Pr[v_i^{i-1}(x \vee \mathcal{N}) = 1] \leq \varepsilon|\mathcal{F}|.$$

Summing over all i gives the claimed bound. $\square$

5.6 Completing the proof

We describe here in very high level how to complete the proof of Theorem 5.2. The problem session will guide you through some of the missing steps; for the full details, we refer to the original paper [Ros14]. We will restrict our attention to the setting of fixed k and large n (recall $n = \binom{m}{2}$); with some more work, the argument extends to larger values of k. We use the following parameters:

- The family $\mathcal{F}$ is all graphs with at most k non-isolated nodes. Note that $|\mathcal{F}| = \Theta(m^k)$.
- The error parameter is $\varepsilon = m^{-10k}$ so that $\varepsilon|\mathcal{F}||\mathcal{C}| \ll 1$.

The argument will crucially use Janson's inequality, which we record below for convenience.

Lemma 5.10 (Janson's inequality)**.** *Let $\mathcal{F}$ be a family of subsets of $[n]$, and let $W \sim Bin([n], p)$. Define μ and Δ by*

$$\mu = \sum_{S \in \mathcal{F}} \Pr[S \subset W] \quad \textit{and} \quad \Delta = \sum_{\substack{S_1, S_2 \in \mathcal{F} \\ S_1 \neq S_2, S_1 \cap S_2 \neq \emptyset}} \Pr[S_1 \cup S_2 \subset W].$$

Then,

$$\Pr\left[\bigwedge_{S \in \mathcal{F}} S \not\subset W\right] \leq \exp\left(-\min\left\{\frac{\mu}{2}, \frac{\mu^2}{2\Delta}\right\}\right).$$

Let $\mathcal{C}$ be a monotone circuit as given in Theorem 5.2, and let $\overline{\mathcal{C}}$ denote its circuit closure. The first step is to establish the following lemma.

Lemma 5.11. *Let $H \sim \mathcal{Y}$ be a random k-clique. Then*

$$\Pr[H \in \mathcal{M}(\overline{\mathcal{C}})] \geq 0.9 - o(1).$$

Proof sketch. We have by assumption that $\Pr[\mathcal{C}(H) = 1] \geq 0.9$, and hence also $\Pr[\overline{\mathcal{C}}(H) = 1] \geq 0.9$. In order to prove that most of them are min-terms, we need to show that $\overline{\mathcal{C}}$ does not accept most of their sub-graphs. We consider

$$\mathcal{Q} = \{Q : Q \text{ is a } k\text{-clique minus an edge}\}.$$

We thus need to show that

$$\Pr_{Q \in \mathcal{Q}}[\overline{\mathcal{C}}(Q) = 1] = o(1).$$

Assume towards a contradiction that this is not the case, and $\overline{\mathcal{C}}$ accepts a constant fraction of $Q \in \mathcal{Q}$.

Recall that $\mathcal{N} \sim G(m, p)$ where we choose $p = m^{-\frac{2}{k-1}-\delta}$ for a small δ. We will choose δ so that

1. With high probability, $G(m, p)$ does not contain a k-clique.
2. With high probability, $G(m, p)$ contains many copies of k-clique minus an edge.

We will show that, under our assumption towards a contradiction,

$$\Pr[\overline{\mathcal{C}}(\mathcal{N}) = 1] \geq 1 - o(1).$$

Assuming this, Lemma 5.9 and our choice of ε then gives the same for $\mathcal{C}$:

$$\Pr[\mathcal{C}(\mathcal{N}) = 1] \geq 1 - o(1).$$

However, this is impossible since we assumed $\Pr[\mathcal{C}(\mathcal{N}) = 1] \leq 0.1$.

So to conclude we need to show $\Pr[\overline{\mathcal{C}}(\mathcal{N}) = 1] \geq 1 - o(1)$. We sketch the argument here, which is based on Janson's inequality. Define

$$\mathcal{Q}' = \{Q \in \mathcal{Q} : \overline{\mathcal{C}}(Q) = 1\}.$$

By our assumption towards a contradiction, $|\mathcal{Q}'| = \Omega(|\mathcal{Q}|)$. For each $Q \in \mathcal{Q}'$ define the random variable

$$X_Q = \mathbf{1}[Q \subset \mathcal{N}],$$

and let $X = \sum_{Q \in \mathcal{Q}'} X_Q$. We can thus recast our goal as

$$\Pr[\overline{\mathcal{C}}(\mathcal{N}) = 1] = \Pr[X \geq 1].$$

In order to apply Janson's inequality, we need to estimate two quantities:

- $\mu = \sum_{Q \in \mathcal{Q}'} \mathbb{E}[X_Q]$

- $\Delta = \sum_{Q_1,Q_2 \in \mathcal{Q}', Q_1 \sim Q_2} \mathbb{E}[X_{Q_1} X_{Q_2}]$, where the notation $Q_1 \sim Q_2$ means that Q_1, Q_2 share at least one common edge.

Janson's inequality gives a meaningful bound whenever μ is large and $\Delta = o(\mu^2)$, so we need to show both.

By our assumptions, since $\mathcal{N}$ is super-critical to the emergence of k-clique minus an edge,

$$\begin{aligned}\mu &= \sum_{Q \in \mathcal{Q}'} \mathbb{E}[X_Q] = |\mathcal{Q}'| p^{\binom{k}{2}-1} \\ &= \frac{|\mathcal{Q}'|}{|\mathcal{Q}|} \cdot (\text{Expected number of } k\text{-cliques minus an edge in } G(m,p)) \gg 1.\end{aligned}$$

Next, we upper bound Δ by summing over all $Q \in \mathcal{Q}$, in which case it is well-known that $\Delta = o(\mu^2)$ for graph counts for *strictly balanced* graphs, where k-clique minus an edge is an example of such a graph (see for example [AS16, section 4.4]). Overall, Janson's inequality then gives that

$$\Pr[\overline{\mathcal{C}}(\mathcal{N}) = 1] = \Pr[X \geq 1] \geq 1 - \exp(-\mu^2/(2\Delta + \mu)) = 1 - o(1). \qquad \square$$

We established that $\overline{\mathcal{C}}$ has $\Omega(m^k)$ min-terms formed by k-cliques; these have size $\ell = \binom{k}{2}$, and so we have

$$\mathcal{M}_\ell(\overline{\mathcal{C}}) \geq \Omega(m^k).$$

On the other hand, Lemma 5.7 shows that $\mathcal{M}_\ell(\overline{\mathcal{C}})$ does not contain a (p,ε)-robust sunflower, and the robust sunflower lemma (Theorem 2.12) gives

$$\mathcal{M}_\ell(\overline{\mathcal{C}}) \leq (cp^{-1}\log(\ell/\varepsilon))^\ell \approx m^{\left(\frac{2}{k-1}+\delta\right)\ell} = m^{k+\delta'},$$

so it seems like we got nowhere. However, there is a way forward; we will be super-sketchy here as this part of the analysis is not related to threshold phenomena, but you will be guided through a full proof in today's problem session.

By analyzing the min-terms structure in a monotone circuit, one can show that because $\overline{\mathcal{C}}$ contains a constant fraction of k-cliques as min-terms, there must be:

1. A node $\overline{v}$ in $\overline{\mathcal{C}}$,
2. A graph H on $h \leq k$ nodes with $c\binom{k}{2}$ edges, with $c \in [1/3, 2/3]$,

such that the min-terms of $\overline{v}$ contain $\Omega(m^h)/|\mathcal{C}|$ copies of H. Applying the same logic as above to $\overline{v}$ instead of to $\overline{\mathcal{C}}$, and assuming $|\mathcal{C}|$ is small enough, completes the proof.

References

[ALWZ21] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. *Annals of Mathematics*, 194(3):795–815, 2021.

[AS16] Noga Alon and Joel H Spencer. *The Probabilistic Method*. John Wiley & Sons, 2016.

[ASU12] Noga Alon, Amir Shpilka, and Christopher Umans. On sunflowers and matrix multiplication. In *2012 IEEE 27th Conference on Computational Complexity*, pages 214–223. IEEE, 2012.

[BCW21] Tolson Bell, Suchakree Chueluecha, and Lutz Warnke. Note on sunflowers. *Discrete Mathematics*, 344(7):112367, 2021.

[CKR22] Bruno Pasqualotto Cavalar, Mrinal Kumar, and Benjamin Rossman. Monotone circuit lower bounds from robust sunflowers. *Algorithmica*, 84(12):3655–3685, 2022.

[EG17] Jordan S Ellenberg and Dion Gijswijt. On large subsets of with no three-term arithmetic progression. *Annals of Mathematics*, pages 339–343, 2017.

[ER60] Paul Erdős and Richard Rado. Intersection theorems for systems of sets. *Journal of the London Mathematical Society*, 1(1):85–90, 1960.

[ES78] Paul Erdős and Endre Szemerédi. Combinatorial properties of systems of sets. *Journal of Combinatorial Theory, Series A*, 24(3):308–313, 1978.

[FKNP21] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park. Thresholds versus fractional expectation-thresholds. *Annals of Mathematics*, 194(2):475–495, 2021.

[Fuk18] Junichiro Fukuyama. Improved bound on sets including no sunflower with three petals. *arXiv preprint arXiv:1809.10318*, 2018.

[KK07] Jeff Kahn and Gil Kalai. Thresholds and expectation thresholds. *Combinatorics, Probability and Computing*, 16(3):495–502, 2007.

[Kos96] Alexandr V. Kostochka. An intersection theorem for systems of sets. *Random Structures & Algorithms*, 9(1-2):213–221, 1996.

[LLZ18] Xin Li, Shachar Lovett, and Jiapeng Zhang. Sunflowers and quasi-sunflowers from randomness extractors. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2018.

[LMM+22] Shachar Lovett, Raghu Meka, Ian Mertz, Toniann Pitassi, and Jiapeng Zhang. Lifting with sunflowers. *Leibniz international proceedings in informatics*, 215, 2022.

[PP24] Jinyoung Park and Huy Pham. A proof of the Kahn–Kalai conjecture. *Journal of the American Mathematical Society*, 37(1):235–243, 2024.

[Rao20] Anup Rao. Coding for sunflowers. *Discrete Analysis*, 2020.

[Raz85] Alexander Razborov. Lower bounds on the monotone complexity of some Boolean function. In *Soviet Math. Dokl.*, volume 31, pages 354–357, 1985.

[Ros14] Benjamin Rossman. The monotone complexity of k-clique on random graphs. *SIAM Journal on Computing*, 43(1):256–279, 2014.

[Tao20] Terence Tao. The sunflower lemma via Shannon entropy. `https://terrytao.wordpress.com/2020/07/20/the-sunflower-lemma-via-shannon-entropy/`, 2020.