Subject: Sunflower Stuff/Smoother Spread Set Systems

Speaker: Sam Spiro

Subject: Sunflower Stuff/Smoother Spread Set Systems

Speaker: Sam Spiro

Synopsis: Several stellar scholars (see Shachar's site) studied spread set systems. Subsequently, said scholars surpassed some serious sunflower struggles. Sam shall summarize some significant strategies, sans some sporadic sums (so shall simplify some steps).

Subject: Sunflower Stuff/Smoother Spread Set Systems

Speaker: Sam Spiro

Synopsis: Several stellar scholars (see Shachar's site) studied spread set systems. Subsequently, said scholars surpassed some serious sunflower struggles. Sam shall summarize some significant strategies, sans some sporadic sums (so shall simplify some steps).

To appear at the Constant Consonant Conference Concerning Combinatorics

Introduction

Last week

Introduction

Last week, **Anthony**

Introduction

Last week, **Anthony** told us about sunflowers and applications to computer science.

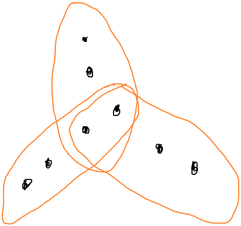
Introduction

Last week, **Anthony** told us about sunflowers and applications to computer science. Today we'll talk a bit more about the details of the recent breakthrough of Alweiss, Lovett, Wu, and Zhang related to sunflowers.

Introduction

Last week, **Anthony** told us about sunflowers and applications to computer science. Today we'll talk a bit more about the details of the recent breakthrough of Alweiss, Lovett, Wu, and Zhang related to sunflowers.

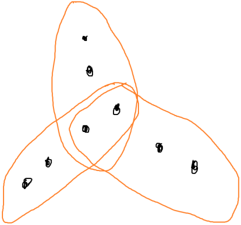
Recall that a *hypergraph* or *set system* $\mathcal{H}$ is a collection of sets called *edges*. The hypergraph is said to be *r-uniform* if every edge has size exactly r.



Introduction

Last week, **Anthony** told us about sunflowers and applications to computer science. Today we'll talk a bit more about the details of the recent breakthrough of Alweiss, Lovett, Wu, and Zhang related to sunflowers.

Recall that a *hypergraph* or *set system* $\mathcal{H}$ is a collection of sets called *edges*. The hypergraph is said to be *r-uniform* if every edge has size exactly r.



A hypergraph $\mathcal{S} = \{S_1, \ldots, S_p\}$ is called a *p-sunflower* if there exists a set K called the *kernel* such that $S_i \cap S_j = K$ for all $i \neq j$.

Introduction

Theorem (Erdős-Rado)

For all r, p, there exists a constant $f(r,p) = (pr)^r$ such that any r-uniform hypergraph $\mathcal{H}$ with more than $f(r,p)$ edges contains a p-sunflower.

Introduction

Theorem (Erdős-Rado)

For all r, p, there exists a constant $f(r,p) = (pr)^r$ such that any r-uniform hypergraph $\mathcal{H}$ with more than $f(r,p)$ edges contains a p-sunflower.

Conjecture (Sunflower Conjecture)

For all p there exists a constant $c = c(p)$ such that any r-uniform hypergraph $\mathcal{H}$ with more than c^r edges contains a p-sunflower.

Introduction

Theorem (Erdős-Rado)

For all r, p, there exists a constant $f(r,p) = (pr)^r$ such that any r-uniform hypergraph $\mathcal{H}$ with more than $f(r,p)$ edges contains a p-sunflower.

Conjecture (Sunflower Conjecture)

For all p there exists a constant $c = c(p)$ such that any r-uniform hypergraph $\mathcal{H}$ with more than c^r edges contains a p-sunflower.

Theorem (Alweiss-Lovett-Wu-Zhang; Rao; Bell-Chueluecha-Warnke)

There exists a constant $C > 0$ such that for all r, p, any r-uniform hypergraph $\mathcal{H}$ with more than $(Cp \log r)^r$ edges contains a p-sunflower.

Introduction

To prove this result, we will need to prove results about hypergraphs whose edges are "spread out."

Introduction

To prove this result, we will need to prove results about hypergraphs whose edges are "spread out." To this end, given a set of vertices A, define the *degree* $d(A)$ to be the number of edges of $\mathcal{H}$ which contain A.

Introduction

To prove this result, we will need to prove results about hypergraphs whose edges are "spread out." To this end, given a set of vertices A, define the *degree* $d(A)$ to be the number of edges of $\mathcal{H}$ which contain A. Given some $0 < q < 1$, we say that a hypergraph is *q-spread* if $d(A) \leq q^{|A|}|\mathcal{H}|$, i.e. any set of k vertices is contained in at most a q^k proportion of the edges of $\mathcal{H}$.

Introduction

To prove this result, we will need to prove results about hypergraphs whose edges are "spread out." To this end, given a set of vertices A, define the *degree* $d(A)$ to be the number of edges of $\mathcal{H}$ which contain A. Given some $0 < q < 1$, we say that a hypergraph is *q-spread* if $d(A) \leq q^{|A|}|\mathcal{H}|$, i.e. any set of k vertices is contained in at most a q^k proportion of the edges of $\mathcal{H}$.

Theorem (Frankston-Kahn-Narayanan-Park)

Let $\mathcal{H}$ be an r-uniform q-spread hypergraph with vertex set V. There exists an absolute constant C_0 such that if W is a uniformly random set of size $Cq \log r \cdot |V|$ chosen from V with $C \geq C_0$, then

$$\Pr[W \textit{ contains an edge of } \mathcal{H}] \geq 1 - \frac{C_0}{C \log r}.$$

I.e. a random set of proportion $q \log r$ is likely to contain an edge.

Applications

Applications

Theorem

Let $G_{n,m}$ be a graph with n vertices and m edges chosen uniformly at random.

Applications

Theorem

Let $G_{n,m}$ be a graph with n vertices and m edges chosen uniformly at random. There exists a constant K such that if $m \geq Kn \log n$ and n is even, then $G_{n,m}$ contains a perfect matching with high probability.

Applications

Theorem

Let $G_{n,m}$ be a graph with n vertices and m edges chosen uniformly at random. There exists a constant K such that if $m \geq Kn \log n$ and n is even, then $G_{n,m}$ contains a perfect matching with high probability.

Roughly this says $G_{n,p}$ with $p \gg \log n/n$ contains a perfect matching with high probability.

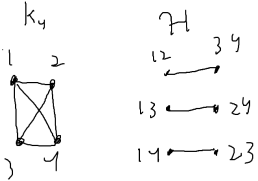
Applications

Theorem

Let $G_{n,m}$ be a graph with n vertices and m edges chosen uniformly at random. There exists a constant K such that if $m \geq Kn \log n$ and n is even, then $G_{n,m}$ contains a perfect matching with high probability.

Roughly this says $G_{n,p}$ with $p \gg \log n/n$ contains a perfect matching with high probability.

Let $\mathcal{H}$ be the hypergraph with vertex set $E(K_n)$ where each hyperedge S is a perfect matching of K_n.



Applications

Note that $|\mathcal{H}| = \frac{n!}{(n/2)!2^{n/2}}$

Applications

Note that $|\mathcal{H}| = \frac{n!}{(n/2)!2^{n/2}}$, and for $A \subseteq E(K_n)$ we have

$$d(A) = \frac{(n-2|A|)!}{(n/2-|A|)!2^{n/2-|A|}} \leq (en/2)^{-|A|}|\mathcal{H}|.$$

Applications

Note that $|\mathcal{H}| = \frac{n!}{(n/2)!2^{n/2}}$, and for $A \subseteq E(K_n)$ we have

$$d(A) = \frac{(n-2|A|)!}{(n/2-|A|)!2^{n/2-|A|}} \leq (en/2)^{-|A|}|\mathcal{H}|.$$

This implies that $\mathcal{H}$ is $(en/2)^{-1}$-spread. It is also $(n/2)$-uniform and has a ground set $V = E(K_n)$ of size $\binom{n}{2}$.

Applications

Note that $|\mathcal{H}| = \frac{n!}{(n/2)!2^{n/2}}$, and for $A \subseteq E(K_n)$ we have

$$d(A) = \frac{(n-2|A|)!}{(n/2-|A|)!2^{n/2-|A|}} \leq (en/2)^{-|A|}|\mathcal{H}|.$$

This implies that $\mathcal{H}$ is $(en/2)^{-1}$-spread. It is also $(n/2)$-uniform and has a ground set $V = E(K_n)$ of size $\binom{n}{2}$. Applying the theorem, a set of size roughly $n^{-1}\log(n/2)\binom{n}{2} \approx n\log n$ edges of K_n is very likely to contain a hyperedge, i.e. a perfect matching. □

Applications

Note that $|\mathcal{H}| = \frac{n!}{(n/2)!2^{n/2}}$, and for $A \subseteq E(K_n)$ we have

$$d(A) = \frac{(n-2|A|)!}{(n/2-|A|)!2^{n/2-|A|}} \leq (en/2)^{-|A|}|\mathcal{H}|.$$

This implies that $\mathcal{H}$ is $(en/2)^{-1}$-spread. It is also $(n/2)$-uniform and has a ground set $V = E(K_n)$ of size $\binom{n}{2}$. Applying the theorem, a set of size roughly $n^{-1}\log(n/2)\binom{n}{2} \approx n \log n$ edges of K_n is very likely to contain a hyperedge, i.e. a perfect matching. $\square$

This result easily extends to perfect matchings in random r-uniform hypergraphs (which was previously thought to be much harder!)

Applications

We just found the threshold for $G_{n,m}$ to contain a perfect matching.

Applications

We just found the threshold for $G_{n,m}$ to contain a perfect matching. A similar argument can be used to find the threshold for $G_{n,m}$ to contain any (fixed) subgraph F (e.g. $F = K_3$).

Applications

We just found the threshold for $G_{n,m}$ to contain a perfect matching. A similar argument can be used to find the threshold for $G_{n,m}$ to contain any (fixed) subgraph F (e.g. $F = K_3$).

Proposition

Let F be an graph and define

$$t(F) = \max\left\{\frac{|E(F')|}{|V(F')|} : F' \subseteq F\right\}.$$

There exists a constant $C(F)$ such that if $m \geq C(F)n^{2-1/t(F)}$, then $G_{n,m}$ contains a copy of F with high probability.

Applications

We just found the threshold for $G_{n,m}$ to contain a perfect matching. A similar argument can be used to find the threshold for $G_{n,m}$ to contain any (fixed) subgraph F (e.g. $F = K_3$).

Proposition

Let F be an graph and define

$$t(F) = \max\left\{\frac{|E(F')|}{|V(F')|} : F' \subseteq F\right\}.$$

There exists a constant $C(F)$ such that if $m \geq C(F)n^{2-1/t(F)}$, then $G_{n,m}$ contains a copy of F with high probability.

It's not hard to prove this with a slightly fiddily second moment argument, but with spread hypergraphs the proof is much cleaner.

Applications

Let $\mathcal{H}$ be the hypergraph on $E(K_n)$ whose hyperedges correspond to copies of F. Note that each set $A \subseteq E(K_n)$ of positive degree in $\mathcal{H}$ corresponds to some subgraph $F_A \subseteq F$ with $|E(F_A)| = |A|$

Applications

Let $\mathcal{H}$ be the hypergraph on $E(K_n)$ whose hyperedges correspond to copies of F. Note that each set $A \subseteq E(K_n)$ of positive degree in $\mathcal{H}$ corresponds to some subgraph $F_A \subseteq F$ with $|E(F_A)| = |A|$, and in this case

$$\left(\frac{d(A)}{|\mathcal{H}|}\right)^{1/|A|} \leq \left(\frac{n^{|V(F)|-|V(F_A)|}}{\binom{n}{|V(F)|}}\right)^{1/|A|} \approx n^{-|V(F_A)|/|E(F_A)|}.$$

Applications

Let $\mathcal{H}$ be the hypergraph on $E(K_n)$ whose hyperedges correspond to copies of F. Note that each set $A \subseteq E(K_n)$ of positive degree in $\mathcal{H}$ corresponds to some subgraph $F_A \subseteq F$ with $|E(F_A)| = |A|$, and in this case

$$\left(\frac{d(A)}{|\mathcal{H}|}\right)^{1/|A|} \leq \left(\frac{n^{|V(F)|-|V(F_A)|}}{\binom{n}{|V(F)|}}\right)^{1/|A|} \approx n^{-|V(F_A)|/|E(F_A)|}.$$

Thus $\mathcal{H}$ is q-spread with

$$q = \max\{n^{-|V(F')|/|E(F')|} : F' \subseteq F\} = n^{-1/t(F)}.$$

Plugging this into the theorem gives the result. □

Applications

Theorem

There exists a constant C such that if $\mathcal{H}$ is an r-graph with more than $(Cp \log r)^r$ edges, then $\mathcal{H}$ contains a p-sunflower.

Applications

Theorem

There exists a constant C such that if $\mathcal{H}$ is an r-graph with more than $(Cp\log r)^r$ edges, then $\mathcal{H}$ contains a p-sunflower.

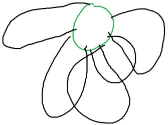
We prove the result by induction on r, the $r = 1$ case being trivial. Let $\mathcal{H}$ be an r-graph with at least this many edges.

Applications

Theorem

There exists a constant C such that if $\mathcal{H}$ is an r-graph with more than $(Cp\log r)^r$ edges, then $\mathcal{H}$ contains a p-sunflower.

We prove the result by induction on r, the $r = 1$ case being trivial. Let $\mathcal{H}$ be an r-graph with at least this many edges. If $\mathcal{H}$ is not q-spread with $q = (Cp\log r)^{-1}$, then there exists some $A \subseteq V(H)$ such that $d(A) \geq (Cp\log r)^{r-|A|}$

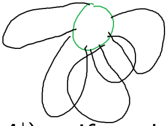


Applications

Theorem

There exists a constant C such that if $\mathcal{H}$ is an r-graph with more than $(Cp\log r)^r$ edges, then $\mathcal{H}$ contains a p-sunflower.

We prove the result by induction on r, the $r = 1$ case being trivial. Let $\mathcal{H}$ be an r-graph with at least this many edges. If $\mathcal{H}$ is not q-spread with $q = (Cp\log r)^{-1}$, then there exists some $A \subseteq V(H)$ such that $d(A) \geq (Cp\log r)^{r-|A|}$



So by induction, the $(r - |A|)$-uniform hypergraph $\mathcal{H}' = \{S \setminus A : A \subseteq S \in \mathcal{H}\}$ contains a sunflower $\{S_1, \ldots, S_p\}$, which means $\mathcal{H}$ contains the sunflower $\{S_1 \cup A, \ldots, S_p \cup A\}$.

Applications

Thus we can assume $\mathcal{H}$ is q-spread, and further that the size of the vertex set V of $\mathcal{H}$ is a multiple of $2p$.

Applications

Thus we can assume $\mathcal{H}$ is q-spread, and further that the size of the vertex set V of $\mathcal{H}$ is a multiple of $2p$. Let $V_1, \ldots, V_{2p}$ be a random partition of V such that each $V_i \subseteq V$ has size $(2p)^{-1}|V|$.

Applications

Thus we can assume $\mathcal{H}$ is q-spread, and further that the size of the vertex set V of $\mathcal{H}$ is a multiple of $2p$. Let $V_1, \ldots, V_{2p}$ be a random partition of V such that each $V_i \subseteq V$ has size $(2p)^{-1}|V|$. This means that each V_i is a uniformly chosen set of V of size $(2p)^{-1}|V| = \frac{1}{2}Cq\log(r)|V|$.

Applications

Thus we can assume $\mathcal{H}$ is q-spread, and further that the size of the vertex set V of $\mathcal{H}$ is a multiple of $2p$. Let $V_1, \ldots, V_{2p}$ be a random partition of V such that each $V_i \subseteq V$ has size $(2p)^{-1}|V|$. This means that each V_i is a uniformly chosen set of V of size $(2p)^{-1}|V| = \frac{1}{2}Cq\log(r)|V|$.

Let 1_i be the indicator variable for V_i containing an edge of $\mathcal{H}$. By the theorem, we have $\Pr[1_i = 1] \geq \frac{1}{2}$ provided C is sufficiently large.

Applications

Thus we can assume $\mathcal{H}$ is q-spread, and further that the size of the vertex set V of $\mathcal{H}$ is a multiple of $2p$. Let $V_1, \ldots, V_{2p}$ be a random partition of V such that each $V_i \subseteq V$ has size $(2p)^{-1}|V|$. This means that each V_i is a uniformly chosen set of V of size $(2p)^{-1}|V| = \frac{1}{2}Cq\log(r)|V|$.

Let 1_i be the indicator variable for V_i containing an edge of $\mathcal{H}$. By the theorem, we have $\Pr[1_i = 1] \geq \frac{1}{2}$ provided C is sufficiently large. In this case, $\mathbb{E}[\sum 1_i] \geq p$, and hence there exists some partition $V_1, \ldots, V_{2p}$ such that $\sum 1_i \geq p$

Applications

Thus we can assume $\mathcal{H}$ is q-spread, and further that the size of the vertex set V of $\mathcal{H}$ is a multiple of $2p$. Let $V_1, \ldots, V_{2p}$ be a random partition of V such that each $V_i \subseteq V$ has size $(2p)^{-1}|V|$. This means that each V_i is a uniformly chosen set of V of size $(2p)^{-1}|V| = \frac{1}{2}Cq\log(r)|V|$.

Let 1_i be the indicator variable for V_i containing an edge of $\mathcal{H}$. By the theorem, we have $\Pr[1_i = 1] \geq \frac{1}{2}$ provided C is sufficiently large. In this case, $\mathbb{E}[\sum 1_i] \geq p$, and hence there exists some partition $V_1, \ldots, V_{2p}$ such that $\sum 1_i \geq p$, which in particular means there exist p disjoint edges of $\mathcal{H}$. Thus $\mathcal{H}$ contains a sunflower, a contradiction. $\square$

Proof of Main Theorem

Proof of Main Theorem

The main strategy is that we iteratively generate $\log r$ random sets W_i of size $q|V|$

Proof of Main Theorem

The main strategy is that we iteratively generate $\log r$ random sets W_i of size $q|V|$, we then win if the following holds:

Lemma

If $\mathcal{H}$ is a q-spread r-uniform hypergraph and you randomly choose a set $W \subseteq V(\mathcal{H})$ of size $q|V|$, then it's very likely that almost every $S \in \mathcal{H}$ has at least half its vertices covered, i.e. $|S \setminus W| \leq r/2$.

Proof of Main Theorem

The main strategy is that we iteratively generate $\log r$ random sets W_i of size $q|V|$, we then win if the following holds:

Lemma (False)

If $\mathcal{H}$ is a q-spread r-uniform hypergraph and you randomly choose a set $W \subseteq V(\mathcal{H})$ of size $q|V|$, then it's very likely that almost every $S \in \mathcal{H}$ has at least half its vertices covered, i.e. $|S \setminus W| \leq r/2$.

Proof of Main Theorem

The main strategy is that we iteratively generate $\log r$ random sets W_i of size $q|V|$, we then win if the following holds:

Lemma (False)

If $\mathcal{H}$ is a q-spread r-uniform hypergraph and you randomly choose a set $W \subseteq V(\mathcal{H})$ of size $q|V|$, then it's very likely that almost every $S \in \mathcal{H}$ has at least half its vertices covered, i.e. $|S \setminus W| \leq r/2$.

Let's just pretend this is true for a second.

Proof of Main Theorem

- Start with $\mathcal{H}_1 = \mathcal{H}$ and W_1 a random set of size $q|V|$.

Proof of Main Theorem

- Start with $\mathcal{H}_1 = \mathcal{H}$ and W_1 a random set of size $q|V|$.
- Let $\mathcal{H}_2 = \{S_1 \setminus W_1 : S_1 \in \mathcal{H}_1, |S_1 \setminus W_1| \leq r/2\}$.

Proof of Main Theorem

- Start with $\mathcal{H}_1 = \mathcal{H}$ and W_1 a random set of size $q|V|$.
- Let $\mathcal{H}_2 = \{S_1 \setminus W_1 : S_1 \in \mathcal{H}_1, |S_1 \setminus W_1| \leq r/2\}$. Similarly define W_2 to be a random set of size $q|V|$ and $\mathcal{H}_3 = \{S_2 \setminus W_2 : S_2 \in \mathcal{H}_2, |S_2 \setminus W_2| \leq r/4\}$ and so on.

Proof of Main Theorem

- Start with $\mathcal{H}_1 = \mathcal{H}$ and W_1 a random set of size $q|V|$.
- Let $\mathcal{H}_2 = \{S_1 \setminus W_1 : S_1 \in \mathcal{H}_1, |S_1 \setminus W_1| \leq r/2\}$. Similarly define W_2 to be a random set of size $q|V|$ and $\mathcal{H}_3 = \{S_2 \setminus W_2 : S_2 \in \mathcal{H}_2, |S_2 \setminus W_2| \leq r/4\}$ and so on.
- By the "lemma", $\mathcal{H}_2$ will basically contain as many edges as $\mathcal{H}_1$

Proof of Main Theorem

- Start with $\mathcal{H}_1 = \mathcal{H}$ and W_1 a random set of size $q|V|$.
- Let $\mathcal{H}_2 = \{S_1 \setminus W_1 : S_1 \in \mathcal{H}_1, |S_1 \setminus W_1| \leq r/2\}$. Similarly define W_2 to be a random set of size $q|V|$ and $\mathcal{H}_3 = \{S_2 \setminus W_2 : S_2 \in \mathcal{H}_2, |S_2 \setminus W_2| \leq r/4\}$ and so on.
- By the "lemma", $\mathcal{H}_2$ will basically contain as many edges as $\mathcal{H}_1$, so $d(A) \leq q^{-|A|}|\mathcal{H}| \approx q^{-|A|}|\mathcal{H}_2|$, i.e. $\mathcal{H}_2$ is basically q-spread and $(r/2)$-uniform.

Proof of Main Theorem

- Start with $\mathcal{H}_1 = \mathcal{H}$ and W_1 a random set of size $q|V|$.
- Let $\mathcal{H}_2 = \{S_1 \setminus W_1 : S_1 \in \mathcal{H}_1, |S_1 \setminus W_1| \leq r/2\}$. Similarly define W_2 to be a random set of size $q|V|$ and $\mathcal{H}_3 = \{S_2 \setminus W_2 : S_2 \in \mathcal{H}_2, |S_2 \setminus W_2| \leq r/4\}$ and so on.
- By the "lemma", $\mathcal{H}_2$ will basically contain as many edges as $\mathcal{H}_1$, so $d(A) \leq q^{-|A|}|\mathcal{H}| \approx q^{-|A|}|\mathcal{H}_2|$, i.e. $\mathcal{H}_2$ is basically q-spread and $(r/2)$-uniform. By the "lemma" again, $\mathcal{H}_3$ has basically as many edges as $\mathcal{H}_2$ and is q-spread and $(r/4)$-uniform.

Proof of Main Theorem

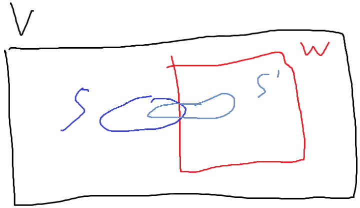
- Start with $\mathcal{H}_1 = \mathcal{H}$ and W_1 a random set of size $q|V|$.
- Let $\mathcal{H}_2 = \{S_1 \setminus W_1 : S_1 \in \mathcal{H}_1, |S_1 \setminus W_1| \leq r/2\}$. Similarly define W_2 to be a random set of size $q|V|$ and $\mathcal{H}_3 = \{S_2 \setminus W_2 : S_2 \in \mathcal{H}_2, |S_2 \setminus W_2| \leq r/4\}$ and so on.
- By the "lemma", $\mathcal{H}_2$ will basically contain as many edges as $\mathcal{H}_1$, so $d(A) \leq q^{-|A|}|\mathcal{H}| \approx q^{-|A|}|\mathcal{H}_2|$, i.e. $\mathcal{H}_2$ is basically q-spread and $(r/2)$-uniform. By the "lemma" again, $\mathcal{H}_3$ has basically as many edges as $\mathcal{H}_2$ and is q-spread and $(r/4)$-uniform.
- After about $\log r$ steps, $\mathcal{H}_i$ is going to have some empty edges, i.e. there exists $S \in \mathcal{H}$ such that $S \subseteq W_1 \cup W_2 \cdots \cup W_i$.

Proof of Main Theorem

- Start with $\mathcal{H}_1 = \mathcal{H}$ and W_1 a random set of size $q|V|$.
- Let $\mathcal{H}_2 = \{S_1 \setminus W_1 : S_1 \in \mathcal{H}_1, |S_1 \setminus W_1| \leq r/2\}$. Similarly define W_2 to be a random set of size $q|V|$ and $\mathcal{H}_3 = \{S_2 \setminus W_2 : S_2 \in \mathcal{H}_2, |S_2 \setminus W_2| \leq r/4\}$ and so on.
- By the "lemma", $\mathcal{H}_2$ will basically contain as many edges as $\mathcal{H}_1$, so $d(A) \leq q^{-|A|}|\mathcal{H}| \approx q^{-|A|}|\mathcal{H}_2|$, i.e. $\mathcal{H}_2$ is basically q-spread and $(r/2)$-uniform. By the "lemma" again, $\mathcal{H}_3$ has basically as many edges as $\mathcal{H}_2$ and is q-spread and $(r/4)$-uniform.
- After about $\log r$ steps, $\mathcal{H}_i$ is going to have some empty edges, i.e. there exists $S \in \mathcal{H}$ such that $S \subseteq W_1 \cup W_2 \cdots \cup W_i$.
- The set $W = W_1 \cup W_2 \cdots \cup W_i$ is basically a random set of size $q \log r|V|$, so we conclude that a set of this size is likely to contain an edge of $\mathcal{H}$. $\square$

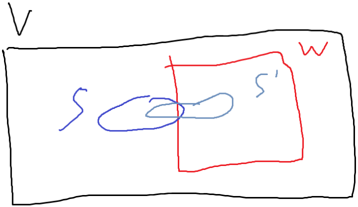
Proof of Main Theorem

Given an r-uniform hypergraph $\mathcal{H}$, say that a pair of sets (S, W) with $S \in \mathcal{H}$ is *good* if there exists some edge $S' \subseteq S \cup W$ with $|S' \setminus W| \leq r/2$ and that it's *bad* otherwise



Proof of Main Theorem

Given an r-uniform hypergraph $\mathcal{H}$, say that a pair of sets (S, W) with $S \in \mathcal{H}$ is *good* if there exists some edge $S' \subseteq S \cup W$ with $|S' \setminus W| \leq r/2$ and that it's *bad* otherwise



It turns out that the same approach as before works as long as almost all pairs (S, W) are good (e.g. we let $\mathcal{H}_2$ have edge set $S' \setminus W_1$ as opposed to $S \setminus W_1$).

Proof of Main Theorem

Lemma (Not False)

Let $\mathcal{H}$ be an r-uniform n-vertex hypergraph on V which is q-spread. If $p = Cq$, then

$$\left|\left\{(S, W) : S \in \mathcal{H},\ W \in \binom{V}{pn},\ (S, W) \textit{ is bad}\right\}\right| \leq 3(C/2)^{-r/4}|\mathcal{H}|\binom{n}{pn}$$

I.e. for large C almost every pair (S, W) is such that there exists $S' \subseteq S \cup W$ with $|S' \setminus W| \leq r/2$.

Proof of Main Theorem

Lemma (Not False)

Let $\mathcal{H}$ be an r-uniform n-vertex hypergraph on V which is q-spread. If $p = Cq$, then

$$\left|\left\{(S,W) : S \in \mathcal{H},\ W \in \binom{V}{pn},\ (S,W) \textit{ is bad}\right\}\right| \leq 3(C/2)^{-r/4}|\mathcal{H}|\binom{n}{pn}$$

I.e. for large C almost every pair (S, W) is such that there exists $S' \subseteq S \cup W$ with $|S' \setminus W| \leq r/2$. For $t \leq r$, define

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

Observe that the quantity we wish to bound is $\sum_t |\mathcal{B}_t|$, so it suffices to bound each term of this sum.

Proof of Main Theorem

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.

Proof of Main Theorem

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 $(S, W) \in \mathcal{B}_t$.

Proof of Main Theorem

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 $|S \cup W|$ and then bound how many $S, W \subseteq Z$ have $(S, W) \in \mathcal{B}_t$.

Proof of Main Theorem

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 $|S \cup W|$ 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 $\mathcal{B}_t$.

Proof of Main Theorem

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 $|S \cup W|$ 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 (somewhat imprecisely) say that a bad pair (S, W) is pathological if the number of bad pairs in $S \cup W$ is larger than some quantity N to be determined later.

Proof of Main Theorem

Let $w := pn - t$.

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

$$\binom{n}{r+w} N \binom{r}{t}.$$

Proof of Main Theorem

Let $w := pn - t$.
Claim: The number of $(S, W) \in \mathcal{B}_t$ which are non-pathological is at most

$$\binom{n}{r+w} N \binom{r}{t}.$$

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

Proof of Main Theorem

Let $w := pn - t$.
Claim: The number of $(S, W) \in \mathcal{B}_t$ which are non-pathological is at most

$$\binom{n}{r+w} N \binom{r}{t}.$$

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 $r + w$.

Proof of Main Theorem

Let $w := pn - t$.
Claim: The number of $(S, W) \in \mathcal{B}_t$ which are non-pathological is at most

$$\binom{n}{r+w} N \binom{r}{t}.$$

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 $r + w$. Because Z is non-pathological, there are at most N choices for S such that $(S, Z \setminus S)$ is bad.

Proof of Main Theorem

Let $w := pn - t$.
Claim: The number of $(S, W) \in \mathcal{B}_t$ which are non-pathological is at most

$$\binom{n}{r+w} N \binom{r}{t}.$$

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 $r + w$. Because Z is non-pathological, there are at most N choices for S such that $(S, Z \setminus S)$ is bad. Given S, there are at most $\binom{r}{t}$ choices for $S \cap W$.

Proof of Main Theorem

Let $w := pn - t$.
Claim: The number of $(S, W) \in \mathcal{B}_t$ which are non-pathological is at most

$$\binom{n}{r+w} N \binom{r}{t}.$$

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 $r + w$. Because Z is non-pathological, there are at most N choices for S such that $(S, Z \setminus S)$ is 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.

Proof of Main Theorem

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

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

Proof of Main Theorem

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

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

The general strategy is to identify these pairs by first specifying $S \in \mathcal{H}$, then $S \cap W$, then $W \setminus S$.

Proof of Main Theorem

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

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

The general strategy is to identify these pairs by first specifying $S \in \mathcal{H}$, then $S \cap W$, then $W \setminus S$. The first two steps can happen in $|\mathcal{H}| \cdot \binom{r}{t}$ ways.

Proof of Main Theorem

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

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

The general strategy is to identify these pairs by first specifying $S \in \mathcal{H}$, then $S \cap W$, then $W \setminus S$. The first two steps can happen in $|\mathcal{H}| \cdot \binom{r}{t}$ ways. To count the remaining choices for $W \setminus S$, we upper bound the probability that a random set of this form is pathological and bad.

Proof of Main Theorem

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

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

The general strategy is to identify these pairs by first specifying $S \in \mathcal{H}$, then $S \cap W$, then $W \setminus S$. The first two steps can happen in $|\mathcal{H}| \cdot \binom{r}{t}$ ways. To count the remaining choices for $W \setminus S$, we upper bound the probability that a random set of this form is pathological and bad. Observe that the probability that this happens is at most

$$\Pr[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\} \geq N] \leq \frac{\mathbb{E}[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\}]}{N}$$

(if (S, W) is bad then every $S' \subseteq S \cup W$ satisfies $|S' \cap S| \geq r/2$).

Proof of Main Theorem

$$\Pr[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\} \geq N] \leq \frac{\mathbb{E}[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\}]}{N}.$$

Proof of Main Theorem

$$\Pr[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\} \geq N] \leq \frac{\mathbb{E}[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\}]}{N}.$$

For any $j \geq r/2$, the number of S' with $|S' \cap S| = j \geq r/2$ is at most

$$\sum_{B \subseteq S : |B| = j} d(B)$$

Proof of Main Theorem

$$\Pr[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\} \geq N] \leq \frac{\mathbb{E}[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\}]}{N}.$$

For any $j \geq r/2$, the number of S' with $|S' \cap S| = j \geq r/2$ is at most

$$\sum_{B \subseteq S : |B| = j} d(B) \leq \binom{r}{j} q^j |\mathcal{H}| \approx q^j |\mathcal{H}|$$

Proof of Main Theorem

$$\Pr[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\} \geq N] \leq \frac{\mathbb{E}[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\}]}{N}.$$

For any $j \geq r/2$, the number of S' with $|S' \cap S| = j \geq r/2$ is at most

$$\sum_{B \subseteq S : |B| = j} d(B) \leq \binom{r}{j} q^j |\mathcal{H}| \approx q^j |\mathcal{H}|,$$

and one can work out that the probability that a given S' is in $S \cup W$ is about $(Cq)^{-j}$, so putting things together gives the claim. □

Proof of Main Theorem

$$\Pr[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\} \geq N] \leq \frac{\mathbb{E}[\#\{S' \subseteq S \cup W : |S' \cap S| \geq r/2\}]}{N}.$$

For any $j \geq r/2$, the number of S' with $|S' \cap S| = j \geq r/2$ is at most

$$\sum_{B \subseteq S : |B| = j} d(B) \leq \binom{r}{j} q^j |\mathcal{H}| \approx q^j |\mathcal{H}|,$$

and one can work out that the probability that a given S' is in $S \cup W$ is about $(Cq)^{-j}$, so putting things together gives the claim. □

We can now pick N so that the estimates of these two claims are about the same, and in total this shows there are few bad pairs, proving the lemma (and hence the theorem). □

Further Results

Theorem (Frankston-Kahn-Naryanan-Park)

If $\mathcal{H}$ is q-spread and r-uniform, then a random set of size $\gg q \log r \cdot |V|$ contains an edge of $\mathcal{H}$ with high probability.

Further Results

Theorem (Frankston-Kahn-Naryanan-Park)

If $\mathcal{H}$ is q-spread and r-uniform, then a random set of size $\gg q \log r \cdot |V|$ contains an edge of $\mathcal{H}$ with high probability.

This result can be sharp (e.g. for perfect matchings in random graphs).

Further Results

Theorem (Frankston-Kahn-Naryanan-Park)

If $\mathcal{H}$ is q-spread and r-uniform, then a random set of size $\gg q \log r \cdot |V|$ contains an edge of $\mathcal{H}$ with high probability.

This result can be sharp (e.g. for perfect matchings in random graphs).

Theorem (Kahn-Naryanan-Park)

If $\mathcal{H}$ is the hypergraph encoding squares of Hamiltonian cycles of K_n, then one can remove the $\log r$ term.

Further Results

Theorem (Frankston-Kahn-Naryanan-Park)

If $\mathcal{H}$ is q-spread and r-uniform, then a random set of size $\gg q \log r \cdot |V|$ contains an edge of $\mathcal{H}$ with high probability.

This result can be sharp (e.g. for perfect matchings in random graphs).

Theorem (Kahn-Naryanan-Park)

If $\mathcal{H}$ is the hypergraph encoding squares of Hamiltonian cycles of K_n, then one can remove the $\log r$ term.

The proof is remarkably similar to the proof we just outlined

Further Results

Theorem (Frankston-Kahn-Naryanan-Park)

If $\mathcal{H}$ is q-spread and r-uniform, then a random set of size $\gg q \log r \cdot |V|$ contains an edge of $\mathcal{H}$ with high probability.

This result can be sharp (e.g. for perfect matchings in random graphs).

Theorem (Kahn-Naryanan-Park)

If $\mathcal{H}$ is the hypergraph encoding squares of Hamiltonian cycles of K_n, then one can remove the $\log r$ term.

The proof is remarkably similar to the proof we just outlined, so it's perhaps natural to ask if we can (1) generalize when we can drop the $\log r$ term

Further Results

Theorem (Frankston-Kahn-Naryanan-Park)

If $\mathcal{H}$ is q-spread and r-uniform, then a random set of size $\gg q \log r \cdot |V|$ contains an edge of $\mathcal{H}$ with high probability.

This result can be sharp (e.g. for perfect matchings in random graphs).

Theorem (Kahn-Naryanan-Park)

If $\mathcal{H}$ is the hypergraph encoding squares of Hamiltonian cycles of K_n, then one can remove the $\log r$ term.

The proof is remarkably similar to the proof we just outlined, so it's perhaps natural to ask if we can (1) generalize when we can drop the $\log r$ term, and (2) try and find some interpolation between these two proof methods.

Further Results

We say that a hypergraph $\mathcal{H}$ is $(q; r_1, r_2, \ldots, r_\ell)$-spread with $r_1 > r_2 > \cdots > r_\ell$ if it's r_1-uniform and for every $A \subseteq V$ with $|A| = r_i$ and $j \geq r_{i+1}$

$$\sum_{B \subseteq A : |B| = j} d(B) \leq q^j |\mathcal{H}|.$$

Basically, not too many edges intersect sets of size r_i in at least r_{i+1} vertices.

Further Results

We say that a hypergraph $\mathcal{H}$ is $(q; r_1, r_2, \ldots, r_\ell)$-spread with $r_1 > r_2 > \cdots > r_\ell$ if it's r_1-uniform and for every $A \subseteq V$ with $|A| = r_i$ and $j \geq r_{i+1}$

$$\sum_{B \subseteq A : |B| = j} d(B) \leq q^j |\mathcal{H}|.$$

Basically, not too many edges intersect sets of size r_i in at least r_{i+1} vertices.

Proposition

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-uniform, then it is $(4q; r, r/2, \ldots, 1)$-spread.*

Further Results

Theorem (S.)

If $\mathcal{H}$ is $(q; r_1, r_2, \ldots, r_\ell, 1)$-spread, then a random set of size $q\ell|V|$ is very likely to contain an edge of $\mathcal{H}$.

Further Results

Theorem (S.)

If $\mathcal{H}$ is $(q; r_1, r_2, \ldots, r_\ell, 1)$-spread, then a random set of size $q\ell|V|$ is very likely to contain an edge of $\mathcal{H}$.

The proof is basically the same as before, except now instead of going from edges of size $r, r/2, r/4, \ldots$ we do $r_1, r_2, r_3, \ldots$ (and the definition is designed precisely so that the proof still works).

Further Results

Theorem (S.)

If $\mathcal{H}$ is $(q; r_1, r_2, \ldots, r_\ell, 1)$-spread, then a random set of size $q\ell|V|$ is very likely to contain an edge of $\mathcal{H}$.

The proof is basically the same as before, except now instead of going from edges of size $r, r/2, r/4, \ldots$ we do $r_1, r_2, r_3, \ldots$ (and the definition is designed precisely so that the proof still works). This theorem succeeds in recovering/interpolating between basically all previously known results.

The End

Thank You!