1 Lecture 21: Ferber-McKinley-Samotij (supersaturation)(3/8/2019)

Counting F-free graphs is one of the important questions in extremal graph theory. A counting problem is called *degenerate* if $\text{ex}(n, F) = o(n^2)$. Otherwise, we shall call it *non-degenerate*. The non-degenerate problems, such as counting K_r-free graphs, are well studied. However, the degenerate problems are much more challenging. Examples of the degenerate problems includes counting C_4-free graphs, and also Sidon sets.

Among all the degenerate problems, perhaps the most interesting case is when F is an even cycle. It is well-known that $\text{ex}(n, C_{2k}) = O(n^{1+1/k})$. However, matching lower bounds are only known for $k \in \{2, 3, 5\}$. Many people believe the upper bound is sharp up a constant. Using the alternation method, one can show that $\text{ex}(n, C_{2k}) \geqslant \Omega(n^{1+\frac{1}{2k-1}})$. Kleitman and Wilson [3] proved that for $k \in \{3, 4\}$, the number of n-vertex C_{2k}-free graphs is at most $2^{cn^{1+1/k}}$ for some constant $c = c(k)$. Indeed, Erdős conjectured the following.

Conjecture 1.1. *For every $k \geqslant 2$, the number of n-vertex C_{2k}-free graphs is $2^{O(n^{1+1/k})}$.*

In fact, Erdős first proposed a stronger conjecture, that is, the number of C_{2k}-free graphs is $2^{(1+o(1))\text{ex}(n,C_{2k})}$. However, Morris and Saxton [4] showed this conjecture is false for C_6. The similar behavior also appears in Sidon sets, see Lecture 5 for more details.

To prove Conjecture 1.1, one may try to directly apply the hypergraph container method. However, this approach fails to work since the codegree conditions are not satisfied for the desired bound.

2 Ferber-McKinley-Samotij Theorem

Theorem 2.1 (Ferber-McKinley-Samotij [2])**.** *Let H be an r-uniform hypergraph and let α and A be positive constants. Let*

$$m_r(H) = \max\left\{\frac{e_F - 1}{v_F - 1}, F \subseteq H, v_F > 1\right\}.$$

Suppose that $r > \alpha > r - 1/m_r(H)$ and $\mathrm{ex}(n, H) \leqslant An^{\alpha}$ for all n. Then there exists a constant $C = C(\alpha, A, H)$ such that for all n, the number of H-free r-uniform hypergraphs on n vertices is at most $2^{Cn^{\alpha}}$.

Let us check a special case of this theorem. Suppose that $r = 2$ and $H = C_{2k}$. It is easy to see that $m_2(C_{2k}) = \frac{2k-1}{2k-2}$. Recall that $\mathrm{ex}(n, C_{2k}) = O(n^{1+1/k})$, and $1 + 1/k > 2 - \frac{2k-2}{2k-1} = 1 + \frac{1}{2k-1}$. Therefore, Theorem 2.1 indicates that the number of C_{2k}-free graphs is at most $2^{Cn^{1+1/k}}$ for some constant C, which confirms the Erdős conjecture. However, if $\mathrm{ex}(n, C_{2k}) = \Theta(n^{1+\frac{1}{2k-1}})$, then Theorem 2.1 would only give $2^{Cn^{1+\frac{1}{2k-1}+\varepsilon}}$ as an upper bound.

Let $\mathcal{H}$ be a hypergraph with the vertex set $V(\mathcal{H}) = E(K_n^{(r)})$ and the edge set $E(\mathcal{H}) = \{\text{copies of } H \text{ in } K_n^{(r)}\}$. The standard approach on such counting problems is to apply the hypergraph container method on $\mathcal{H}$. However, this approach fails to work because of the codegree condition. The key ingredient to overcome this difficulty is to find a family of subhypergraphs $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_t \subseteq \mathcal{H}$ satisfying the following two conditions:

- For every independent set I of $\mathcal{H}$, there exists a hypergraph $\mathcal{H}_i$ such that $I \subseteq \mathcal{H}_i$;
- Each $\mathcal{H}_i$ satisfies the codegree condition.

3 Supersaturation

The first supersaturation-type result is due to Varnavides, who studies the number of k-APs.

Theorem 3.1 (Varnavides [5]). *For every k and ε, there exists an integer n_0 and a constant δ such that for all $n > n_0$, every set $A \subseteq [n]$ of size $|A| > \varepsilon n$ contains at least δn^2 k-APs.*

Proof. By Szemerédi's Theorem, there exists a large integer M such that every set $A \subseteq [M]$ of size $|A| \geqslant \frac{\varepsilon M}{2}$ contains a k-AP. The main ingredient of the proof is to show that there are $\delta_1 n^2$ M-APs in $[n]$ (referred to as P_M) such that $|A \cap P_M| \geqslant \frac{\varepsilon M}{2}$. By Szemerédi's Theorem, each $A \cap P_M$ contains a k-AP, and therefore we obtain $\delta_1 n^2$ k-APs. However, this estimation double counts some k-APs. So the next step is to show that the same k-AP cannot occur more than δ_2 times, where δ_2 is a constant depending only on M. From here, we can conclude that the number of k-APs in A is at least $\frac{\delta_1}{\delta_2} N^2 = \delta N^2$. □

Another famous supersaturation result is the Erdős-Simonovits theorem.

Theorem 3.2 (Erdős-Simonovits). *For every $r \geqslant 2$ and $\varepsilon > 0$, there exists an integer n_0 and a constant δ such that for all $n > n_0$, every n-vertex graph G with $e(G) \geqslant (1 - \frac{1}{r-1} + \varepsilon)\frac{n^2}{2}$ contains at least δn^r copies of K_r.*

Proof. We first claim that the number of subsets of $V(G)$ of size t, which contain more than $\left(1-\frac{1}{r-1}\right)\frac{t^2}{2}$ edges, is at least $\frac{\varepsilon}{2}\binom{n}{t}$. Otherwise, we have

$$\sum_{|T|=t} e(G[T]) \leqslant \binom{n}{t}\left(1-\frac{1}{r-1}\right)\frac{t^2}{2}+\frac{\varepsilon}{2}\binom{n}{t}\frac{t^2}{2}.$$

On the other hand, we have

$$\sum_{|T|=t} e(G[T]) \geqslant e(G)\binom{n-2}{t-2} \geqslant \left(1-\frac{1}{r-1}+\varepsilon\right)\frac{n^2}{2}\binom{n-2}{t-2},$$

which contradicts the above upper bound when $n \gg t \gg 1$. By Turán's Theorem, any graph of order t with more than $(1-\frac{1}{r-1})\frac{t^2}{2}$ edges yields a copy of K_r. Therefore, the number of copies of K_r in G is at least

$$\frac{\frac{\varepsilon}{2}\binom{n}{t}}{\binom{n-r}{t-r}} = \delta n^r.$$

□

The next supersaturation is for the degenerate case, where the forbidden graph is C_4.

Theorem 3.3. *Let $C \geqslant \frac{\sqrt{3}}{2}$ be a constant. For every n-vertex graph G with $e(G) \geqslant Cn^{3/2}$, G contains at least $\frac{4}{9}C^4n^2$ copies of C_4.*

Proof. Let $v_1, v_2, \ldots, v_n$ be the vertices in G and $b_i = d_G(v_i)$ for every $i \in [n]$. Let S be the set of paths of length 2 (or 3-paths) in G. We will count 3-paths in two ways.

First, for a vertex v_i, the number of 3-paths containing v_i as the middle point is exactly $\binom{b_i}{2}$. Therefore, we have

$$|S| = \sum_{i=1}^{n}\binom{b_i}{2} \geqslant n\binom{(\sum_{i=1}^{n} b_i)/n}{2} = n\binom{2e(G)/n}{2} \geqslant \frac{4}{3}C^2n^2.$$

On the other hand, let c_{ij} be the number of common neighbors of v_i and v_j, for $1 \leqslant i < j \leqslant n$. Then $|S| = \sum_{i<j} c_{ij}$. Hence, the number of copies of C_4 in G is

$$\frac{1}{2}\sum_{i<j}\binom{c_{ij}}{2} \geqslant \frac{1}{2}\binom{n}{2}\binom{(\sum_{i<j} c_{ij})/\binom{n}{2}}{2} = \frac{1}{2}\binom{n}{2}\binom{|S|/\binom{n}{2}}{2} \geqslant \frac{|S|^2}{4n^2} \geqslant \frac{4}{9}C^4n^2.$$

□

4 Proof of Theorem 2.1

Lemma 4.1 (Setting up iterations for containers)**.** *Let H be an r-uniform hypergraph and let $\mathcal{H}$ be the e_H-uniform hypergraph with the vertex set $V(\mathcal{H}) = E(K_n^{(r)})$ and the edge set $E(\mathcal{H}) =$copies of H in $K_n^{(r)}$. Let K be a constant, and $\gamma = \frac{1}{1-\delta} > 1$, where $\delta = \delta(e_H, K)$ is defined in the container lemma [1, Proposition 3.1]. Suppose that for a given integer n, there exist M and t_0 such that the following holds: for all integers $t \geqslant t_0$ and all $G \subseteq V(\mathcal{H})$ satisfying*

$$\gamma^t M < |G| \leqslant \gamma^{t+1} M,$$

there exists a subhypergraph $\mathcal{H}_G \subseteq \mathcal{H}[G]$ for which

$$\Delta_\ell(\mathcal{H}_G) \leqslant K \cdot \left(\frac{b_t}{|G|}\right)^{\ell-1} \cdot \frac{e(\mathcal{H}_G)}{|G|}$$

where $b_t = \frac{M}{(t+1)^3}$, for all $\ell \in \{1, \ldots, e_H\}$. Then there is a constant $C = C(K, t_0, e_H)$ such that the number of H-free hypergraphs of order n is at most $2^{C \cdot M}$.

Sketch of the proof. We first apply the hypergraph container theorem on $\mathcal{H}$ and obtain a collection of containers $\{G_i^1\}$. By the assumption, for each G_i^1, there exists a subhypergraph $\mathcal{H}_{G_i^1} \subseteq \mathcal{H}[G_i^1]$ satisfying the condition. Next we apply the hypergraph container theorem on each $\mathcal{H}_{G_i^1}$, and obtain a new collection of containers $\{G_i^2\}$. After doing this process $O(\log n)$ steps, each container is sufficiently small. One can show that the number of containers is also small enough so that we obtain the desired bound. □

Next theorem shows that such subhypergraph $\mathcal{H}_G$ always exists. This together with Lemma 4.1 would imply Theorem 2.1

Theorem 4.2. *Let H be an r-uniform hypergraph, and let $\mathcal{H}$ be the e_H-uniform hypergraph with the vertex set $V(\mathcal{H}) = E(K_n^{(r)})$ and the edge set $E(\mathcal{H}) =$copies of H in $K_n^{(r)}$. Let $\gamma > 1$, and $\alpha > r - 1/m_r(H)$. Suppose that M is such that for every $s \in \{1, \ldots, n\}$,*

$$\mathrm{ex}(s, H) \leqslant M \cdot \left(\frac{s}{n}\right)^\alpha.$$

Then there exists a constant $t_0 = t_0(\alpha, \gamma, H)$ such that the following holds. If $G \subseteq V(\mathcal{H})$ satisfies

$$\gamma^t M < |G| \leqslant \gamma^{t+1} M,$$

for some $t \geqslant t_0$, then there is a subhypergraph $\mathcal{H}_G \subseteq \mathcal{H}[G]$ for which

$$\Delta_\ell(\mathcal{H}_G) \leqslant 2^{2e_H+3} \cdot \left(\frac{b_t}{|G|}\right)^{\ell-1} \cdot \frac{e(\mathcal{H}_G)}{|G|} \tag{1}$$

where $b_t = \frac{M}{(t+1)^3}$*, for all* $\ell \in \{1, \dots, e_H\}$.

Proof. Let $\mathcal{H}_0$ be the initial empty hypergraph with $V(\mathcal{H}_0) = V(\mathcal{H}[G])$ and $E(\mathcal{H}_0) = \emptyset$. Suppose that we have already defined $\mathcal{H}_i$. Our goal is to add one hyperedge to H_i to form H_{i+1} without violating the codegree condition (1) for every $\ell \in [e_H - 1]$. $\mathcal{H}_G$ shall be constructed by such a way after N steps, and it will have exactly N hyperedges.

Let $m = |G|$. First, we claim that $N = \left(\frac{\gamma^{t+1}M}{b_t}\right)^{e_H-1} \cdot m$. Observe that $\Delta_{e_H}(\mathcal{H}_G) = 1$. To fulfill the codegree condition (1) for $\ell = e_H$, we need

$$N = e(\mathcal{H}_G) \geqslant \left(\frac{m}{b_t}\right)^{e_H-1} \cdot 2^{-2e_H-3} \cdot m,$$

which is satisfied by the choice of N.

Given $F \subset V(\mathcal{H}_i)$, let $\deg_{\mathcal{H}_i}(F)$ be the number of hyperedges in $\mathcal{H}_i$ containing F. A set $F \subset V(\mathcal{H}_i)$ (or an r-uniform hypergraph F) is 'bad' if $\deg_{\mathcal{H}_i}(F)$ reach the upper bound in (1) for $\ell = e(F)$. Let $\mathcal{B}_F(\mathcal{H}_i)$ be the collections of 'bad' copies of F in G. We will give an upper bound on $|\mathcal{B}_F(\mathcal{H}_i)|$. By the choice of $\mathcal{B}_F(\mathcal{H}_i)$, we have

$$\sum_{F \in \mathcal{B}_F(\mathcal{H}_i)} \deg_{\mathcal{H}_i}(F) \geqslant |\mathcal{B}_F(\mathcal{H}_i)| 2^{2e_H+2} \cdot \left(\frac{b_t}{m}\right)^{e_F-1} \cdot \frac{N}{m}.$$

On the other hand,

$$\sum_{F \in \mathcal{B}_F(\mathcal{H}_i)} \deg_{\mathcal{H}_i}(F) \leqslant e(\mathcal{H}_i) \binom{e_H}{e_F} \leqslant 2^{e_H} \cdot N.$$

Combining the above two inequalities, we obtain that

$$|\mathcal{B}_F(\mathcal{H}_i)| \leqslant 2^{-e_H-2} \left(\frac{m}{b_t}\right)^{e_F-1} \cdot m. \tag{2}$$

A hyperedge $E \in \mathcal{H}[G]$ is *good* if it doesn not contain any 'bad' sets. If there exists a good hyperedge $E \in \mathcal{H}[G] \setminus \mathcal{H}_i$, then we let $\mathcal{H}_{i+1} = \mathcal{H}_i \cup \{E\}$. One can easily check that this new hypergraph $\mathcal{H}_{i+1}$ satisfies the codegree condition. The key ingredient of this proof is to show the existence of such good E.

Recall that G is s subset of $V(\mathcal{H})$. By the definition of $V(\mathcal{H})$, G can also be viewed as an r-uniform n-vertex hypergraph. Fix some $p \in (0, 1]$ (p will be chosen later) and let R be an uniformly chosen random subset of $V(G)$ of order pn. Denote by G' the subgraph of G induced by R. Let G'' be an r-graph obtained from G' by removing one r-edge from each 'bad' copies of F in G', for every nonempty $F \subset H$. By the definition, any copy of H in G'' is good.

Let Z be the total number of good copies of H in G, and let X be the number of copies of H in G''. Then we have

$$\mathbb{E}[X] \leqslant \text{Expected number of good copies of } H \text{ in } G' = Z \cdot \binom{n-v_H}{pn-v_H} / \binom{n}{pn} \leqslant Z \cdot p^{v_H}.$$

On the other hand, since $v(G'') = pn$, then we have

$$X \geqslant e(G'') - \mathrm{ex}(pn, H) \geqslant e(G'') - M \cdot p^{\alpha}.$$

Clearly,

$$\mathbb{E}[e(G'')] \geqslant \mathbb{E}[e(G')] - \sum_{F \subset H} |\mathcal{B}_F(\mathcal{H}_i)| \cdot \mathrm{P}(F \subset G') \geqslant \mathbb{E}[e(G')] - \sum_{F \subset H} |\mathcal{B}_F(\mathcal{H}_i)| \cdot p^{v_F}.$$

Moreover, if $pn \geqslant 2r^2$, then

$$\mathbb{E}[e(G')] = m \cdot \binom{n-r}{pn-r} / \binom{n}{pn} \geqslant \frac{m \cdot p^r}{2}.$$

Therefore, we obtain that

$$Z \cdot p^{v_H} \geqslant \mathbb{E}[X] \geqslant \frac{m \cdot p^r}{2} - \sum_{F \subset H} |\mathcal{B}_F(\mathcal{H}_i)| \cdot p^{v_F} - M \cdot p^{\alpha}. \tag{3}$$

Using (2) and some technical computation, one can show that there is a $p \in [2r^2/n, 1]$ such that the right-hand side of (3) is at least $N \cdot p^{v_H}$, and thus $Z \geqslant N$. This inequality would imply that there is always a good hyperedge $E \in \mathcal{H}[G] \setminus \mathcal{H}_i$ whenever $i < N$, which completes the proof. □

References

[1] J. Balogh, R. Morris and W. Samotij, *Independent sets in Hypergraphs*, J. Amer. Math. Soc. 28 (2015), no. 3, 669–709.

[2] A. Ferber, G. McKinley and W. Samotij, *Supersaturated sparse graphs and hypergraphs*, to appear in Int. Math. Res. Not. IMRN.

[3] D. Kleitman and D. Wilson, *On the number of graphs which lack small cycles*, (1996), manuscript.

[4] R. Morris and D. Saxton. *The number of $C_{2\ell}$-free graphs*, Advances in Mathematics 298 (2016): 534–580.

[5] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. 34 (1959), 358–360.