Rödl Nibble

David Brandfonbrener
Math 345

November 30, 2016

Definition 1. For $2 \leq l < k < n$, define the covering number $M(n,k,l)$ to be the minimal size of a family $\mathcal{K}$ of k-element subsets of $\{1,..,n\}$ such that every l-element subset of $\{1,..,n\}$ is contained in some $A \in \mathcal{K}$

Note 1. We always have $M(n,k,l) \geq \frac{\binom{n}{l}}{\binom{k}{l}}$. To see this, note that there are $\binom{n}{l}$ subsets of l elements, and we want to put each into a k-element subset, which can fit at most $\binom{k}{l}$ such l-element subsets.

Example 1.

- M(4,3,2). By Note 1, we have $M(4,3,2) \geq \frac{6}{3} = 2$. However, drawing it, we must have $|\mathcal{K}| = 3$.
- M(7,3,2). This is the smallest Steiner Triple. This can be illustrates with the Fano Plane, each line is an element of $\mathcal{K}$.

Theorem 1. For fixed $2 \leq l < k$, where $o(1) \to 0$ as $n \to \infty$:

$$M(n,k,l) \leq (1 + o(1))\frac{\binom{n}{l}}{\binom{k}{l}}$$

Definition 2. Let $H = (V,E)$ be an r-uniform hypergraph, with $x \in V$. Then define the degree of x in H, $d(x)$ to be the number of edges in E containing x. And for $x, y \in V$, define $d(x,y)$ to be the number of edges in E containing both x and y. And a covering of H is a set of edges in $C \subseteq E$ such that every vertex in V is in some edge in C.

Note 2. We will let $x = \pm y$ denote $-y \leq x \leq y$.

Note 3. The idea is to prove something more general about r-uniform hypergraphs with certain properties, then to create an r-uniform hypergraph with $r = \binom{k}{l}$, that will allow us to use the lemma to prove the theorem.

Lemma 1. For every integer $r \geq 2$ and reals $k \geq 1$, $a > 0$, there exist $\gamma = \gamma(r,k,a) > 0$ and $d_0 = d_0(r,k,a)$ such that for every $n \geq D \geq d_0$ the following holds.

Let $H = (V,E)$ be an r-uniform hypergraph with n vertices, all with positive degree and satisfying:

1. For all except at most γn vertices $x \in V$, $d(x) = (1 \pm \gamma)D$

2. For all $x \in V$, $d(x) < kD$

3. For any distinct $x, y \in V$, $d(x, y) < \gamma D$.

Then there exists a cover of H with at most $(1 + a)\frac{n}{r}$ edges.

Proof. (of Theorem 1). Let $r = \binom{k}{l}$. Let H be the r uniform hypergraph, with a vertex for each l-element subset of $\{1, ..., n\}$. Then each edge is a collection of $\binom{k}{l}$ l-element subsets that lie within one k-element subset. So we have that $|V| = \binom{n}{l}$. And each vertex has degree $D = \binom{n-l}{k-l}$, the number of possible ways to choose a k-subset that contains a specific l-subset. And for two distinct vertices (l-subsets) they take up at least $l+1$ points so we have at most $\binom{n-l-1}{k-l-1}$ common edges which is $o(D) \leq \gamma D$. Thus, we can apply lemma 1, to get that H has a cover of size $(1 + o(1))\frac{\binom{n}{l}}{\binom{k}{l}}$ and we are done. □

Note 4. The idea is to fix a small ϵ, then randomly choose a set E' of expected size $\epsilon\frac{n}{r}$ edges. Then with high probability only $O(\epsilon^2 n)$ vertices are covered more than once, so E' covers at least $\epsilon n - O(\epsilon^2 n)$ vertices. And, removing these vertices covered by E', the induced Hypergraph still satisfies the three desirable properties. So, we can repeatedly randomly take an ϵ fraction of the vertices removed until we have ϵn vertices left. Then we take one edge for each of the last ϵn vertices, which is inefficient, but good enough to prove Lemma 1. This is the motivation for lemma 2 (and also the idea of the "nibble").

Lemma 2. For every integer $r \geq 2$ and reals $K > 0$, $\epsilon > 0$, and every $\delta' > 0$, there exist $\delta = \delta(r, K, \epsilon, \delta')$ and $D_0 = D_0(r, K, \epsilon, \delta')$ such that for every $n \geq D \geq D_0$ the following holds.

Let $H = (V, E)$ be an r-uniform hypergraph with n vertices, satisfying:

1. For all except at most δn vertices $x \in V$, $d(x) = (1 \pm \delta)D$

2. For all $x \in V$, $d(x) < KD$

3. For any distinct $x, y \in V$, $d(x, y) < \delta D$.

Then there exist a set of edges $E' \subseteq E$ that has the following properties:

4. $|E'| = \frac{\epsilon n}{r}(1 \pm \delta')$

5. Define $V' = V \setminus \cup_{e \in E'} e$. Then we have $|V'| = ne^{-\epsilon}(1 \pm \delta')$

6. For all except at most $\delta'|V'|$ vertices $x \in V'$ the degree $d'(x)$ in the induced hypergraph of H on V' satisfies $d'(x) = De^{-\epsilon(r-1)}(1 \pm \delta')$

Proof. (of Lemma 2). Notes: D, n can be assumed to be as large as needed. Each δ_i is a constant that goes to zero as $\delta \to 0, D \to \infty$ (ie can be smaller than δ').

Choose E' by selecting each edge in E independently and uniformly with probability $p = \frac{\epsilon}{D}$.

Proof of 4) By assumption 1 we have V contains at least $(1-\delta)n$ vertices of degree at least $(1-\delta)D$, so $|E| \geq \frac{(1-\delta)^2 nD}{r}$. By assumptions 1 and 2, we have that $|E| \leq \frac{(1+\delta)nD+\delta nKD}{r}$. So, for some δ_1 we have $|E| = \frac{(1\pm\delta_1)nD}{r}$. Now we can easily calculate the expectation and variance of $|E'|$. This gives us $\mathbb{E}|E'| = p|E| = \frac{(1\pm\delta_1)\epsilon n}{r}$. And we have $Var(|E'|) = |E|p(1-p) \leq |E|p = \frac{(1\pm\delta_1)\epsilon n}{r}$. Then, by Chebyshev's inequality there exists $\delta_2 > 0$ such that

$$Pr(||E'| - \frac{\epsilon n}{r}| \geq \delta_2 \frac{\epsilon n}{r}) \leq \frac{Var(|E'|)}{(\delta_2 \frac{\epsilon n}{r})^2} = \frac{1+\delta_1}{\delta_2^2 \frac{\epsilon n}{r}} \leq .01$$

So we have

$$Pr(|E'| = (1 \pm \delta_2)\frac{\epsilon n}{r}) \geq .99$$

This gives us our result for the desired concentration of the size of E' for property 4.

Proof of 5) For each vertex $x \in V$ define the bernoulli indicator random variable I_x of the event that $x \notin \cup_{e \in E'} e$. So we have that $|V'| = \sum_{x \in V} I_x$. For $n - \delta n$ of the vertices, we have $d(x) = (1 \pm \delta)D$, so for those vertices

$$\mathbb{E}(I_x) = Pr(I_x = 1) = (1-p)^{d(x)} = (1 - \frac{\epsilon}{D})^{(1\pm\delta)D} = e^{-\epsilon}(1 \pm \delta_3)$$

So, since for the other δn vertices we have $0 \leq E(I_x) \leq 1$, we get that $\mathbb{E}(|V'|) = (n - \delta n)e^{-\epsilon}(1 \pm \delta_3) \pm \delta n = ne^{-\epsilon}(1 \pm \delta_4)$. Now we compute the variance:

$$Var(|V'|) = \sum_{x \in V} Var(I_x) + \sum_{x \neq y \in V} Cov(I_x, I_y) \leq \mathbb{E}(|V'|) + \sum_{x \neq y \in V} Cov(I_x, I_y)$$

But, we also have that, using assumption 3:

$$\begin{aligned} Cov(I_x, I_y) &= \mathbb{E}(I_x I_y) - \mathbb{E}(I_x)\mathbb{E}(I_y) = (1-p)^{d(x)+d(y)-d(x,y)} - (1-p)^{d(x)+d(y)} \\ &= (1-p)^{d(x)+d(y)}((1-p)^{-d(x,y)} - 1) \leq (1-p)^{-d(x,y)} - 1 \\ &\leq (1 - \frac{\epsilon}{D})^{-\delta D} - 1 \leq e^{\epsilon\delta} - 1 = \delta_5 \end{aligned}$$

So, putting it together, we get that, since δ_6 just needs to go to 0 as $n \to \infty$ and can be larger than $1/n$,

$$Var(|V'|) \leq \mathbb{E}(|V'|) + n^2\delta_5 \leq \delta_6(\mathbb{E}(|V'|))^2$$

Then, by Chebyshev's inequality, we can choose some $\delta_7 > 0$ so that:

$$Pr(||V'| - \mathbb{E}(|V'|)| \geq \delta_7 \mathbb{E}(|V'|)) \leq \frac{Var(|V'|)}{(\delta_7 \mathbb{E}(|V'|))^2} = \frac{\delta_6}{\delta_7^2} \leq .01$$

So we have with probability .99 that

$$V' = \mathbb{E}(|V'|)(1 \pm \delta_7) = ne^{-\epsilon}(1 \pm \delta_8)$$

This gives us our result for the desired concentration of the size of V' for property 5.

Proof of 6) First we claim that for all vertices $x \in V$ except $\delta_9 n$ of them that:
(A)$d(x) = (1 \pm \delta)D$
(B) All edges $e \in E$ except at most $\delta_{10}D$ such that $x \in E$ satisfy

$$|\{f \in E : x \notin f, f \cap e \neq \emptyset\}| = (1 \pm \delta_{11})(r-1)D$$

(A) holds obviously by assumption 1. For (B), note that the number of edges containing vertices with degrees outside of the δ band around D are at most δnKD. So the number of vertices contained in more than $\delta_{10}D$ such edges must be $\delta nKD/\delta_{10}D \leq \delta_9 n/2$ which is negligible, for appropriate choices. Then, if e contains all good vertices, the number of f that satisfy the condition is at most $(r-1)(1 \pm \delta)D$ and at least $(r-1)(1 \pm \delta)D - \binom{r-1}{2}\delta D$, so we have it.

Now we just need to show that most vertices (up to a delta window) that satisfy A and B also satisfy 6. Let x be such a vertex, and call an edge e with $x \in e$ good if e satisfies (B). Conditioning on $x \in V'$, the chance a good edge remains in the hypergraph is $(1-p)^{(1 \pm \delta_{11})(r-1)D}$. So, the expectation of $d'(x)$ becomes:

$$\mathbb{E}(d'(x)) = (1 \pm \delta \pm \delta_{10})D(1-p)^{(1 \pm \delta_{11})(r-1)D} \pm \delta_{10}D = e^{-\epsilon(r-1)}D(1 \pm \delta_{12})$$

Now for each e containing x, let I_e be the indicator random variable that indicates $e \subseteq V'$. Then $d'(x)$ is the sum of the I_e conditioned on $x \in V'$, so:

$$Var(d'(x)) \leq \mathbb{E}(d'(x)) + \sum_{x \in e, x \in f} Cov(I_e, I_f) \leq \mathbb{E}(d'(x)) + 2\delta_{10}D^2(1 \pm \delta) + \sum_{x \in e, x \in f, f has (B)} Cov(I_e, I_f)$$

$\square$

Proof. (of Lemma 1). Choose $\epsilon > 0$ such that

$$\frac{\epsilon}{1-e^{-\epsilon}} + r\epsilon < 1 + a$$

And we choose $1/10 > \delta > 0$ such that

$$(1+4\delta)\frac{\epsilon}{1-e^{-\epsilon}} + r\epsilon < 1 + a$$

Now we fix t large enough such that $e^{-\epsilon t} < \epsilon$. Then, we apply Lemma 2 t times to finish the proof. Let $\delta = \delta_t$ and then we can define $\delta_t > \delta_{t-1} > ... > \delta_0$ such that $\delta_i \leq \delta_{i+1}e^{-\epsilon(r-1)}$ AND $\prod_{i=0}^{t}(1+\delta_i) < 1+2\delta$ AND for each i from 0 to t and for $n \geq D \geq R_i$ AND we can apply the lemma with $r, K = ke^{i\epsilon(r-1)}, \epsilon, \delta' = \delta_{i+1}, \delta = \delta_i$ (ie choose a very small δ_i). This eventually inductively gives us the theorem for $\gamma = \delta_0, d_0 = \max R_i$. We get decreasing vertex sets $V_0 \supset V_1 \supset ... \supset V_t$ that are induced by each subsequent E' from the lemma on V_{i-1}, which gives us the edge sets to remove at each step called $E_1, ..., E_t$. So we have that

$$\begin{aligned}
|V_i| = |V_{i-1}|e^{-\epsilon}(1 \pm \delta_i) \leq |V_0|e^{-\epsilon i}(1+2\delta) = ne^{-\epsilon i}(1+2\delta) \\
|E_i|\frac{\epsilon|V_{i-1}|}{r}(1 \pm \delta_i) \leq (1+4\delta)\frac{\epsilon n}{r}e^{-(i-1)\epsilon} \\
D_i = D_{i-1}e^{-\epsilon(r-1)} = De^{-\epsilon i(r-1)}
\end{aligned}$$

Now we can cover the remaining vertices in V_t with one edge each. Thus we have that there exists a cover of size

$$\begin{aligned}
\sum_{i=1}^{t} E_i + |V_t| &\leq (1+4\delta)\frac{\epsilon n}{r}\sum_{i=0}^{t} e^{-i\epsilon} + ne^{-\epsilon t}(1+2\delta) \leq (1+4\delta)\frac{\epsilon n}{r}\frac{1}{1-e^{-\epsilon}} + ne^{-\epsilon t}(1+2\delta) \\
&\leq \frac{n}{r}((1+4\delta)(\frac{\epsilon}{1-e^{-\epsilon}} + r\epsilon)) < (1+a)\frac{n}{r}
\end{aligned}$$

This is the desired result, a cover of the proper size. $\square$