Lecture 11: Restricted Intersections of Sets

Combinatorial Methods (Winter 2023)
University of Toronto
Swastik Kopparty
Scribes: Clement Yung

Fix some positive integers n, a, b. Consider familes of sets $\mathcal{Y} \subseteq \binom{[n]}{a}$ such that for all $u, v \in \mathcal{Y}$:

$$|u \cap v| \leq b$$

We are interested in finding out how big/small can the set $\mathcal{Y}$ be, and in this lecture we present various approaches to this problem.

We write $a = \alpha n$ and $b = \beta n$ for some $\alpha, \beta \in [0, 1]$.

1 Conditions on β for $|\mathcal{Y}|$ to be small

Theorem 1. *If $\beta < 2\alpha - 1$, then $|\mathcal{Y}| \leq 1$.*

Proof. Suppose otherwise. If $u, v \in \mathcal{Y}$ are such that $u \neq y$, then:

$$\beta n = b \geq |u \cap v| = |u| + |v| - |u \cup v| \geq 2a - n = (2\alpha - 1)n$$

a contradiction. □

Conversely, if $\beta > \alpha$, then trivially $\mathcal{Y}$ can be arbitrarily large (e.g. it is possible that $\mathcal{Y} = \binom{[n]}{a}$). The growth rate is exponential: By Stirling's approximation for $n!$, we have that:

$$\binom{n}{\alpha n} \sim 2^{c_\alpha \cdot n} \cdot \text{poly}(n)$$

where $\text{poly}(n)$ is some polynomial with variable n, and:

$$c_\alpha = H(\alpha) = \alpha \log \frac{1}{\alpha} + (1 - \alpha) \log \left(\frac{1}{1 - \alpha} \right)$$

2 Growth rates of $|\mathcal{Y}|$

We now consider the following questions:

1. For which α, β can $|\mathcal{Y}|$ be exponentially big?
2. For which α, β must $|\mathcal{Y}|$ be $\leq \text{poly}(n)$?

2.1 Exponential growth rate

Fix some m, and write:

$$\mathcal{Y} = \{u_1, u_2, \ldots, u_m\}$$

where each u_i is picked independently and uniformly from $\binom{[n]}{a}$. Since each number in $[n]$ has a uniform probability of $\frac{1}{n}$ to be in any of the v_i's, we "expect" the "typical intersection" size to be $\frac{a^2}{n} = \alpha^2 \cdot n$. To see this: Instead of taking $u \in \binom{[n]}{a}$, we take each element of $[n]$ into u with probability $\frac{a}{n}$. Thus, for a specific $i \in [n]$ and two fixed sets in $\binom{[n]}{a}$, there is $\left(\frac{a}{n}\right)^2$ probability of i belonging in both sets. Since there are n many such i's, the expected size of intersection is $n \cdot \left(\frac{a}{n}\right)^2$.

This also tells us the following:

Theorem 2. *If $\beta > \alpha^2$, then there exists exponentially large $\mathcal{Y}$.*

Intuitively, this is because if we take a large family of random sets, then by the above reasoning most of the families should have intersection at most $\alpha^2 n < b$, which satisfies the requirement $|u \cap v| \leq b$.

Proof. Pick b such that $a^2 < p < b$. For each $i \in [m]$, where m is to be determined, let u_i be a set where each element is taken independently with probability p. Then:

$$\mathbb{E}[|u_i|] = \mathbb{E}\left[\sum_{k=1}^{n} \mathbb{1}_{k \in u_i}\right] = \sum_{k=1}^{n} \mathbb{E}[\mathbb{1}_{k \in u_i}] = pn$$

By Chernoff's bound, we have that for all i and $\varepsilon > 0$:

$$\Pr[||u_i| - pn| > \varepsilon n] \leq e^{-\frac{\varepsilon^2 n}{4}}$$

Choose $\varepsilon := p - \alpha$. Then:

$$\Pr[|u_i| < \alpha n] \leq \Pr[||u_i| - pn| > \varepsilon n] \leq e^{-\frac{\varepsilon^2 n}{4}}$$

Therefore:

$$\Pr[\exists i \in [m] \text{ such that } |u_i| < \alpha n] = \sum_{i=1}^{n} \Pr[|u_i| < \alpha n] \leq m e^{-\frac{\varepsilon^2 n}{4}}$$

Now fix $i, j \in [m]$ with $i \neq j$. Then:

$$|v_i \cap v_j| = \mathbb{E}\left[\sum_{k=1}^{n} \mathbb{1}_{k \in u_i} \cdot \mathbb{1}_{k \in u_j}\right] = \sum_{k=1}^{n} \Pr[k \in u_i \wedge k \in u_j] = p^2 n$$

Let $X_k := \mathbb{1}_{k \in u_i} \cdot \mathbb{1}_{k \in u_j}$ (so $\Pr[X_k = 1] = p^2$). Then, by Chernoff's bound again:

$$\Pr\left[\left|\sum X_k - p^2 n\right| > \epsilon^2 n\right] \leq e^{-\frac{\epsilon^2 n}{4}} \tag{1}$$

Choose $\epsilon = \beta - p^2$ here, and we have that:

$$\Pr[|u_i \cap u_j| > \beta n] \leq e^{-\frac{\epsilon^2 n}{4}}$$

Therefore:

$$\Pr[\exists i, j \in [m],\ i \neq j \text{ such that } |u_i \cap u_j| > \beta n] \leq \binom{m}{2} e^{-\frac{\epsilon^2 n}{4}} \tag{2}$$

Thus, if m is chosen such that $me^{-\frac{\varepsilon^2 n}{4}} < 1$ (inequality (1)) and $\binom{m}{2}e^{-\frac{\epsilon^2 n}{4}} < 1$ (inequality (2)), then with positive probability we obtain a family with size of intersections at most $\beta n = b$. □

2.2 Polynomial growth rate

Theorem 3. *If $\beta < \alpha^2$, then $|\mathcal{Y}| \leq n + 1$.*

Proof. We consider each u_i as a vector $u_i \in \{0, 1\}^n \subseteq \mathbb{R}^n$, such that for all $i \neq j$:

$$\langle u_i, u_i \rangle = \alpha n,\ \langle u_i, u_j \rangle \leq \beta n$$

Write $\mathcal{Y} = \{u_1, \ldots, u_m\}$.

Lemma 4. *If $\tilde{u}_i, \ldots, \tilde{u}_m \in \mathbb{R}^n \setminus \{0\}$ are such that $\langle \tilde{u}_i, \tilde{u}_j \rangle < 0$ for all $i \neq j$, then $m \leq n + 1$.*

Proof. We induct on n. Assume WLOG that $\tilde{u}_i$ are all unit vectors, and by rotating the space if necessary we also assume that $\tilde{u}_1 = e_1$. Then write out the vectors as follow:

$$\begin{aligned}
\tilde{u}_1 &= \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \\
\tilde{u}_2 &= \begin{bmatrix} c_2 & \text{—}\ \vec{r}_2\ \text{—} \end{bmatrix} \\
\tilde{u}_3 &= \begin{bmatrix} c_3 & \text{—}\ \vec{r}_3\ \text{—} \end{bmatrix} \\
&\vdots \\
\tilde{u}_m &= \begin{bmatrix} c_m & \text{—}\ \vec{r}_m\ \text{—} \end{bmatrix}
\end{aligned}$$

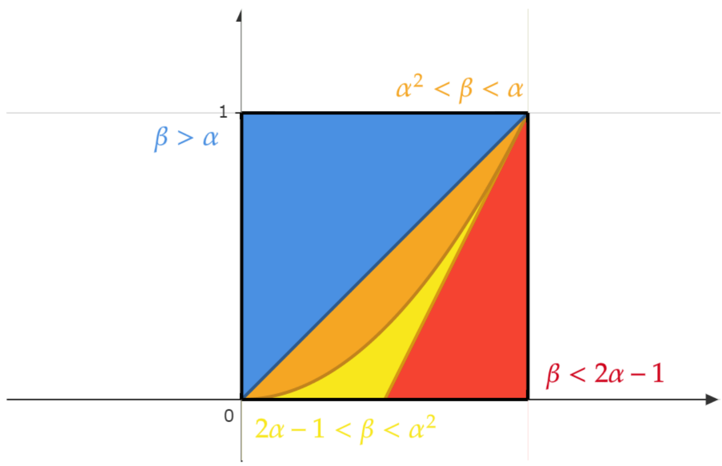
We have that $c_2, \ldots, c_m < 0$, and $\vec{r}_2, \ldots, \vec{r}_m \in \mathbb{R}^{n-1}$. This implies that $\langle \vec{r}_i, \vec{r}_j \rangle < 0$ for all $i \neq j$. By induction hypothesis, we have that $m - 1 \leq n$, so $m \leq n + 1$. ■

We now prove the theorem. Take $w = \gamma \cdot \vec{1} \in \mathbb{R}^n$ (where $\vec{1} = (1, \ldots, 1)$), where γ is to be determined. Let $\tilde{u}_i := u_i - w$. Then for $i \neq j$:

$$\begin{aligned}
\langle \tilde{u}_i, \tilde{u}_j \rangle &= \langle u_i - w, u_j - w \rangle \\
&= \langle u_i, u_j \rangle - \langle u_i, w \rangle - \langle u_j, w \rangle + \langle w, w \rangle \\
&\leq \beta n - \gamma \alpha n - \gamma \alpha n + n\gamma^2 \\
&= \gamma^2 - 2\alpha n \gamma + \beta n \\
&= n \left[(\gamma - \alpha)^2 + (\beta - \alpha^2) \right]
\end{aligned}$$

Choose $\gamma = \alpha$, and we have that $\langle \tilde{u}_i, \tilde{u}_j \rangle \leq n(\beta - \alpha^2) < 0$. By the Lemma above, we have that $m \leq n + 1$, as desired. □

In summary, we have the following:



1. If $\beta < \alpha$, then the family can be arbitrarily large.
2. If $\alpha^2 < \beta < \alpha$, then by Theorem 2 the family can be exponentially large.
3. If $2\alpha - 1 < \beta < \alpha$, then by Theorem 3 the family can be polynomially large.
4. If $\beta < \alpha$, then by Theorem 1 the family has at most one set.

3 Bounding $|\mathcal{Y}|$ with using bipartite graphs

Again, let $\mathcal{Y} = \{u_1, \ldots, u_m\} \subseteq \binom{[n]}{a}$. Consider a bipartite graph $G = (\mathcal{Y} \sqcup [n], E)$, such that for each i and k, $(u_i, k) \in E$ iff $k \in u_i$. The assertion that $|u_i \cap u_j| \leq \beta n$ for all $i \neq j$ is equivalent to saying that the graph G does not contain $K_{2,\beta n+1}$ as an induced subgraph. How can we use this property to bound m?

We start by counting the number of subgraphs of the shape ">". That is, an induced subgraph with vertices $\{u_i, k, u_j\}$ for some $u_i, u_j \in \mathcal{Y}$ and $k \in [n]$. We first note that:

$$\#(\text{“>”}) \leq \binom{m}{2} \cdot \beta n = \frac{m(m-1)\beta n}{2}$$

as there are $\binom{m}{2}$ many pairs in $\mathcal{Y}$, and each pair has at most βn many elements in their intersection. On the other hand:

$$\#(\text{“>”}) = \sum_{k \in [n]} \binom{\deg(k)}{2} \geq \sum_{k \in [n]} \binom{\overline{\deg}}{2} = n \cdot \binom{\alpha m}{2} = \frac{n(\alpha m)(\alpha m - 1)}{2}$$

where $\overline{\text{deg}}$ is the average degrees of nodes on the right. Combining both inequalities, we have that:

$$\frac{n(\alpha m)(\alpha m - 1)}{2} \leq \frac{m(m-1)\beta n}{2} \implies m(\alpha^2 - \beta) \leq \alpha - \beta$$
$$\implies m \leq \frac{\alpha - \beta}{\alpha^2 - \beta}$$