1 Lecture 4: A Random Sperner Theorem

Recall that $2^{[n]}$ is the Boolean lattice. A family $\mathcal{A} \subseteq 2^{[n]}$ is an *antichain* if for every $B, C \in \mathcal{A}$, $B \not\subseteq C$ and $C \not\subseteq B$ (*i.e.*, B and C are not comparable in the containment poset on $2^{[n]}$).

A fundamental theorem about antichains is Sperner's famous result from 1928 [16]:

Theorem 1.1 (Sperner)**.** *If $\mathcal{A}$ is an antichain, then $|\mathcal{A}| \leqslant \binom{n}{\lfloor \frac{n}{2} \rfloor} =: m$. Furthermore, equality is obtained only by the "middle layer(s)" of $2^{[n]}$ (i.e., all subsets of size $\lfloor \frac{n}{2} \rfloor$, or also all subsets of size $\lceil \frac{n}{2} \rceil$ when n is odd).*

A natural question is: what is the number of antichains in $2^{[n]}$?

Theorem 1.2 (Kleitman [19])**.** *The number of antichains in $2^{[n]}$ is $2^{(1+o(1))m}$, where m is as above.*

Idea of proof: Define a graph G by $V(G) := 2^{[n]}$ and $A \sim B \iff A \subseteq B$ or $B \subseteq A$. Independent sets in G then correspond to antichains. We will run the graph container method on this setup. For a container $C \subseteq 2^{[n]}$, if $|C| > (1+\epsilon)m$, then C will contain many edges, which contradicts the graph container lemma. This gives an upper bound on C, and we can then use the counting lemma to obtain an upper bound on the number of antichains in $2^{[n]}$.

Recall a more formal statement of the graph container lemma.

Lemma 1.3 ((More) Formal Graph Container Lemma)**.** *Given G_N, a graph on N vertices, and a parameter t, there are $\{(S_i, C_i)\}$ such that*

- *For every i, $|S_i| < \frac{N}{t}$.*
- *For every i, $\Delta(G_N[C_i]) < t$.*
- *For every independent set I in G_N, there is an i such that $S_i \subseteq I$ and $I \subseteq S_i \cup C_i$.*

Throughout this lecture, we set $N := 2^n$, $m := \binom{n}{\lfloor \frac{n}{2} \rfloor}$.

1.1 Supersaturation Condition

Kleitman [17] defined the so-called "centrality order" on the lattice $2^{[n]}$. In this ordering, we first take the elements of $\binom{[n]}{\lfloor \frac{n}{2} \rfloor}$ and order these elements arbitrarily. Then we take the elements of $\binom{[n]}{\lfloor \frac{n}{2} \rfloor+1}$ and order them arbitrarily, then the elements of $\binom{[n]}{\lfloor \frac{n}{2} \rfloor-1}$, then the elements of $\binom{[n]}{\lfloor \frac{n}{2} \rfloor+2}$, and so on.

Define $J_r := \{A_1, \dots, A_r\}$ to be an initial segment of length r of a centrality order.

Theorem 1.4 (Kleitman (1966))**.** *Let $U \subseteq 2^{[n]}$ with $|U| = r$. Then, the number of comparable pairs (i.e., the number of pairs (A, B) with $A \subseteq B$) in U is at least the number of comparable pairs in J_r.*

That is, J_r has the minimal number of comparable pairs among all families of size r in $2^{[n]}$. It will therefore be sufficient for us to study J_r in the sequel.

Theorem 1.4 immediately implies the following claim:

Claim 1.5. *(i) If $r = (1 + \gamma)m$, with $\gamma < \frac{1}{3}$ a small constant, then the number of comparable pairs in J_r is at least $\frac{\gamma mn}{3}$.*

(ii) If $r = (2+\gamma)m$, γ as above, then the number of comparable pairs in J_r is at least $\frac{\gamma mn^2}{9}$.

Note that a more precise version of this statement is Corollary 5 in [18].

We come back to the question of counting the number of antichains, a question considered by Kleitman [19] and independently Sapozhenko [20]. Recall that $N = 2^n$ and $m = \binom{n}{\lfloor \frac{n}{2} \rfloor} \sim \frac{1}{\sqrt{n}} 2^n$. We prove that the number of antichains in $2^{[n]}$ is $2^{m(1+o(1))}$.

By the counting lemma from the graph container theorem, the number of antichains is at most $\binom{N}{N/t} 2^{\max |C_i|}$ (with the $\binom{N}{N/t}$ term coming from the choices for the certificate S_i). We want to consider the case $|C_i| < (1 + o(1))m$. If $|C_i| > (1 + \gamma)m$, then by Claim 1.5(i), we have that $\Delta(G[C_i]) \geqslant \frac{\gamma n}{4} =: t$. Therefore, using standard estimates (such as $\binom{n}{k} \leqslant \left(\frac{ne}{k}\right)^k$), we obtain (taking $\gamma \approx \frac{\log n}{\sqrt{n}}$)

$$\binom{N}{\frac{N}{t}} \leqslant \binom{2^n}{4\frac{2^n}{\gamma n}} \leqslant 2^{c \log n \frac{2^n}{n}} \leqslant 2^{m \frac{\log n}{\sqrt{n}}} = 2^{o(m)}.$$

1.2 A Random Version of the Sperner Theorem

This work is based on a paper of Balogh-Mycroft-Treglown [18].

Recall that G_N is the comparability graph of $2^{[n]}$, (*i.e.*, $V(G_N) = 2^{[n]}$, and $A \sim B \iff A \subseteq B$ or $B \subseteq A$). Antichains correspond to independent sets in G_N.

Let $0 \leqslant p \leqslant 1$ denote a probability. Define (G_N, p) by keeping each vertex with probability p independently of other choices. (Note that this is *not* the Erdős-Rényi random graph!).

Observe that $\mathbb{E}(|V(G_N, p)|) = pN$ and $\mathbb{E}(\text{max size antichain in } (G_N, p)) = pm$.

Question: For what p is it true that $\alpha(G_N, p) = (1 + o(1))pm$ with high probability?

- For $p = 1$, this is true by Sperner's theorem.

- For $p = \frac{c}{n}$, $c \in \mathbb{R}^+$ arbitrary, and $n \to \infty$, this is false. To see this, take all the sets randomly chosen from the middle layer, and then add elements under the middle layer *if possible* (meaning that a set under the middle layer is added if it is in (G_N, p) and if none of its neighbors in the middle layer are in $G(N, p)$). If X is a random variable that counts the number of sets below the middle layer added during this process, then we have that

$$\mathbb{E}(X) = (1 + o(1))mp(1 - p)^{\frac{n}{2}}.$$

  For $p = \frac{c}{n}$, we have that $(1 - p)^{\frac{n}{2}} \approx e^{-\frac{c}{2}} = c'$, so the expected number of additions is $c'mp$, giving an independent set of size at least $(1+c')mp > (1+o(1))mp$ in expectation.

The above observations beget the natural question: what if $p \gg \frac{1}{n}$? Osthus [21] proved that the statement in the question is true if $p \gg \frac{\log n}{n}$. To remove the logarithmic factor, it is not enough to use the usual graph containers method — we must instead use a two-phase approach. However, it may be helpful to first see how the regular graph container method fails, and for this purpose, we make a short first moment method calculation.

1.3 First Moment Method Failure

Set $L = (1+c)pm$, and let X be a random variable corresponding to the number of antichains in (G_N, p) of size L. There are $\binom{N}{N/n}$ choices for the certificate sets S_i and $\binom{(1+\gamma)m}{(1+c)pm}$ choices for the L-sets in a container C_i, so we may give an upper bound for the number of choices of L as

$$\binom{N}{\frac{N}{n}}\binom{(1+\gamma)m}{(1+c)pm} \leqslant 2^{\log n \frac{N}{n}} 2^{(1+c)pm \log n}.$$

However, this is not sufficient to conclude the theorem. Observe that

$$\mathbb{E}(X) \leqslant p^{(1+c)pm},$$

and we could conclude the theorem if $\mathbb{E}(X) = o(1)$. But $p^{(1+c)pm} \sim 2^{-pm \log n}$, and $2^{\log n \frac{N}{n}} \gg 2^{pm \log n}$ since $N = 2^n$, $p = \frac{c}{n}$, and $m = \binom{n}{\lfloor \frac{n}{2} \rfloor} \sim \frac{2^n}{\sqrt{n}}$.

1.4 Two Phase Graph Container algorithm

We will save the graph containers method for this problem by instead running **two** phases of the graph containers algorithm. The goal is to gain a smaller certificate by dividing it into two specially chosen certificates S_1 and S_2. The basic ideas for Phases I and II are as follows:

- In Phase I, we will add a vertex u to the certificate S_1 if $\deg(u) \geqslant t_1 \sim n^{1.99}$.
- In Phase II, we will add u to a different certificate S_2 if $\deg(u) \geqslant t_2 \sim \gamma n$.

We now give full details for implementation of the algorithm.

Input: Fix an order of $V(G_N)$ to break ties. Let I be an independent set in G_N.

Phase I: Build the certificate S_1.

Start: Initialize $S_1 := \emptyset$, $C := 2^{[n]}$.

Step: Pick $v_i \in C$ such that v_i has maximum degree in $G[C]$ (breaking ties by the arbitrary order at the beginning). Is $v_i \in I$?

NO: Set $S_1 := S_1$, $C := C - v_i$. Continue with Phase I.

YES: Set $S_1 := S_1 \cup \{v_i\}$, $C := C - v_i - N(v_i)$. Is $|N(v_i)| > t_1$?

NO: Continue with Phase II.

YES: Continue Phase I.

Phase II: Set $S := S - S_1$. We work with the parameter t_2.

Start: Initialize $S_2 := \emptyset$.

Step: Pick $v_i \in C$ as in Phase I. Is $v_i \in I$?

NO: Set $S_2 := S_2$, $C := C - v_i$. Continue with Phase II.

YES: Set $S_2 := S_2 \cup \{v_i\}$, $C := C - v_i - N(v_i)$. Is $|N(v_i)| > t_2$?

NO: Stop.

YES: Continue with Phase II.

We have the standard estimate $|S_1| \leqslant \frac{N}{t_1} \sim \frac{N}{n^{1.99}} \leqslant \frac{m}{n^{1.48}}$, but the real improvement is in the size of S_2. Indeed, we have that $|S_2| \leqslant \frac{|C_1|}{t_2} \sim \frac{|C_1|}{\gamma n}$, where C_1 is the container C at the end of Phase I. Observe that $\Delta(G[C_1]) \leqslant t_1 \sim n^{1.99}$. Now, by Claim 1.5 and the fact that $(2+\gamma)m \leqslant 3m$, we have that $|C_1| \leqslant 3m$. Hence, the bound for S_2 can now be written as $|S_2| \leqslant \frac{3m}{\gamma n}$.

1.5 Balogh-Mycroft-Treglown Theorem

We now state the formal theorem proved by Balogh, Mycroft, and Treglown [18]. With the setup that we have devised, the proof is essentially computation.

Theorem 1.6. *For any $\gamma > 0$, there exists a constant c such that if $p > \frac{c}{n}$, then with high probability the largest antichain in (G_N, p) has size at most $(1+\gamma)pm$.*

In their paper [18], Balogh, Mycroft and Treglown actually prove a slightly more general statement than Theorem 1.6.

Proof of Theorem 1.6. Let X be a random variable corresponding to the number of independent sets of size $(1+\gamma)mp$. It is our aim to show that $\mathbb{E}(X) = o(1)$. By the 2-phase graph container algorithm, for any independent set I, we can find a certificate $S_1 \subseteq I$. Associated to S_1 is the container C_1 such that $I \subseteq S_1 \cup C_1$. Additionally, from Phase II, there is a certificate $S_2 \subseteq C_1$, associated to which is a container C_2, with the property that $I \subseteq S_1 \cup S_2 \cup C_2$. We also have the following bounds on the sizes of S_1, S_2, C_1 and C_2.

- $|S_1| \leqslant mn^{-1.48}$
- $|S_2| \leqslant \frac{3m}{\gamma n}$.
- $|C_1| \leqslant 3m$.
- $|C_2| \leqslant (1 + \frac{1}{10}\gamma)m$.

Observe that the number of ways to choose S_1 is $\binom{2^n}{\leqslant mn^{-1.48}}$, and for each possible choice of S_1 there is a $p^{|S_1|}$ probability that it will appear in (G_N, p). Similarly, the number of ways to choose S_2 is $\binom{3m}{\frac{3m}{\gamma n}}$, and for each possible choice of S_2 there is a $p^{|S_2|}$ probability that that choice will appear in (G_N, p). Finally, the number of choices for $I - S_1 - S_2 \subseteq C_2$ is $\binom{(1+\frac{1}{10}\gamma)m}{(1+\gamma)mp}$, each possibility occurring in (G_N, p) with probability $p^{(1+\gamma)mp}$. These probabilities should be summed up over all possible choices for $|S_1|$ $(0 \leqslant |S_1| \leqslant mn^{-1.48})$ and all possible choices for $|S_2|$ $(0 \leqslant |S_2| \leqslant \frac{3m}{\gamma n})$.

First, we fix a choice of S_1 and S_2. We have the following bounds:

$$p^{|S_1|}\binom{2^n}{mn^{-1.48}} \leqslant \left(\frac{2^n p}{mn^{-1.48}}\right)^{mn^{-1.48}} = \left(\frac{2^n \frac{c}{n}}{\frac{2^n}{\sqrt{n}} n^{-1.48}}\right)^{mn^{-1.48}} \ll 2^{mn^{-1.48}\log n} \tag{1}$$

and

$$p^{|S_2|}\binom{3m}{|S_2|} \leqslant \left(\frac{emp}{|S_2|}\right)^{|S_2|} \leqslant \left(\frac{c}{\gamma}\right)^{\frac{3}{\gamma}\frac{m}{n}}. \tag{2}$$

From the graph containers two phase method approach, we have that C_2 is a container of size $\sim (1+\frac{1}{10}\gamma)m$, so the expected number of random sets from (G_N, p) intersecting C_2 is $\sim (1+\frac{1}{10}\gamma)mp$. But the independent set I is contained in C_2, so I is a subset of the intersection of a random set in (G_N, p) and C_2. Suppose that $|I| > (1+\gamma)mp$. Then, by a standard Chernoff bound, the probability that I is contained in (G_N, p) is bounded above by

$$e^{-\frac{\gamma^2 m^2 p^2}{8mp}} \approx e^{-\frac{\gamma^2}{8}mp}. \tag{3}$$

We now sum over all possible choices of S_1 and S_2 and use the union bound and (1), (2) and (3). This gives that the probability that (G_N, p) contains an independent set is at most:

$$\begin{aligned}\Pi &:= \sum_{0\leqslant a\leqslant mn^{-1.48}} \sum_{0\leqslant b\leqslant \frac{3m}{\gamma n}} \binom{2^n}{a} p^a \binom{3m}{b} p^b e^{-\frac{\gamma^2}{8}mp} \\ &\leqslant \left(mn^{-1.48}+1\right)\left(\frac{3m}{\gamma n}+1\right) 2^{mn^{-1.48}\log n} \left(\frac{c}{\gamma}\right)^{\frac{3}{\gamma}\frac{m}{n}} e^{-\frac{\gamma^2}{8}mp}.\end{aligned}$$

For large n, we have that

$$\left(mn^{-1.48}+1\right)\left(\frac{3m}{\gamma n}+1\right) \leqslant e^{\frac{\gamma^2}{32}pm}$$

and

$$2^{mn^{-1.48}\log n} \leqslant e^{\frac{\gamma^2}{32}pm}.$$

Recall that $m \sim \frac{2^n}{\sqrt{n}}$ and $p \geqslant \frac{c}{n}$. The first one follows because the LHS is $\sim \frac{2^n}{n^{1.98}}\frac{2^n}{n^{1.5}} \ll x^2$, where $x = \frac{2^n}{n^{1.5}}$, while the RHS is $\sim e^{\frac{2^n}{n^{1.5}}} = e^x$. The second one follows since the LHS is $\sim 2^{\frac{2^n}{n^{1.98}}\log n}$, while the RHS is $\sim e^{\frac{2^n}{n^{1.5}}}$, and comparing exponents yields $\log n \ll n^{.48}$.

By taking c sufficiently large (say $c = 10^{10}\gamma^{-5}$), for sufficiently large n it follows that

$$\left(\frac{c}{\gamma}\right)^{\frac{3}{\gamma}\frac{m}{n}} \leqslant e^{\frac{\gamma^2}{32}pm}.$$

This gives that the upper bound on the probability Π is $o(1)$, completing the proof. $\square$

References

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

[2] J. Balogh, R. Morris, and W. Samotij, *The Method of Hypergraph Containers*, 2018, survey for ICM 2018.

[3] D. Saxton and A.G. Thomason, *Hypergraph Containers*. Inventiones Mathematicae, 201 (2015), 925–992.

[4] P. Turán, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok 48 (1941), 436–452.

[5] P. Erdős and M. Simonovits, *Supersaturated Graphs and Hypergraphs*. Combinatorica 3 (1983), 183.

[6] P. Erdős, *Some recent results on extremal problems in graph theory.* Results, Theory of Graphs (Internat. Sympos., Rome, 1966), Gordon and Breach, New York; Dunod, Paris, 1967, pp. 117–123 (English); pp. 124–130 (French).

[7] M. Simonovits, *A method for solving extremal problems in graph theory*, stability problems, Theory of Graphs (Proc. Colloq., Tihany, 1966), Academic Press, New York, 1968, pp. 279–319.

[8] J. Balogh, H. Liu and M. Sharifzadeh, 2016. *The number of subsets of integers with no k-term arithmetic progression.* International Mathematics Research Notices, 2017(20), 6168–6186.

[9] J. Balogh, N. Bushaw, M. Collares, H. Liu, R. Morris and M. Sharifzadeh (2017). *The typical structure of graphs with no large cliques.* Combinatorica, 37(4), 617–632.

[10] P. J. Cameron and P. Erdős. *On the number of sets of integers with various properties.* Number theory (Banff, AB, 1988), 61–79. (1990)

[11] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression.* Acta Arith, 27 (1975), 299–345.

[12] D. J. Kleitman and K. J. Winston, *The asymptotic number of lattices*, Ann. Discrete Math. 6 (1980), 243–249, Combinatorial mathematics, optimal designs and their applications (Proc. Sympos. Combin. Math. and Optimal Design, Colorado State Univ., Fort Collins, Colo., 1978).

[13] D. J. Kleitman and K. J. Winston, *On the number of graphs without 4-cycles*, Discrete Math. 41 (1982), 167–172.

[14] J. Kahn, *An entropy approach to the hard-core model on bipartite graphs*, Combinatorics, Probability and Computing (2001) 10, 219–237.

[15] Y. Zhao, *The number of independent sets in a regular graph*, Combinatorics, Probability and Computing (2010) 19, 315–320.

[16] E. Sperner, *Ein Satz über Untermengen einer endlichen Menge*, Mathematische Zeitschrift (1928) 27, 544–548.

[17] D. Kleitman, *A conjecture of Erdős-Katona on commmensurable pairs among subsets of an n-set*, Theory of Graphs, Proc. Colloq., Tihany (1966), 215–218.

[18] J. Balogh, R. Mycroft, and A. Treglown, *A random version of Sperner's theorem*, Journal of Combinatorial Theory, Series A (2014) 128, 104–110.

[19] D. Kleitman, *On Dedekind's problem: the number of monotone boolean functions*, Proc. Amer. Math. Soc. (1969) 21, 677–682.

[20] A. A. Sapozhenko, *The number of antichains in ranked partially ordered sets*, Diskret. Math. (1989) 1, 74–93.

[21] D. Osthus, *Maximum antichains in random sets of a finite set*, Journal of Combinatorial Theory, Series A (2000) 90, 36–46.