# COS 598I Lecture 5: The hypergraph Moore bound and a shorter refutation for 3-SAT

Lecturer: Pravesh Kothari    Scribe: Ştefan Tudose

Spring 2025

## 1 The Moore bound

### 1.1 Graphs

The *girth* of a graph is defined as the length of its shortest cycle. Given a graph $G$ on $n$ vertices and of average degree $d$, what is the maximum possible girth? If $d = 2$ then we can take a cycle, so the girth can be $n$. However, as soon as $d > 2$, the bound becomes logarithmic.

Let us consider the case of a $d$-regular graph on $n$ vertices. If $g$ is the girth, then the ball of radius $\left\lfloor \frac{g-1}{2} \right\rfloor$ around any vertex (i.e. the vertices at distance at most $(g-1)/2$) will look like a tree (otherwise we would have a cycle of length less than $g$), so we have at least

$$1 + d \sum_{k=0}^{\lfloor (g-1)/2 \rfloor} (d-1)^k \sim d^{\lfloor (g+3)/2 \rfloor}$$

vertices. Since we have precisely $n$ vertices, this gives the so-called *Moore bound* $g \lesssim \log_{d-1} n$.

Let us now briefly discuss a spectral approach which will set the tone for the rest of the lecture. Let $A$ be its adjacency matrix and note that $\mathrm{tr} A^k$ counts the number of closed walks of length $k$ in the graph with $k$ smaller than the girth (and even). A closed walk starts off at a node (for which we have $n$ choices), and goes $k/2$ steps away from the original node and $k/2$ steps closer to the original node. At each step when it goes away from the original node, it has at most $d-1$ choices of an edge to use. At each step when it comes closer, because $k$ is less than the girth, it must come back through the edge that it came (for otherwise there will be a cycle). The number of ways to choose in which order to do the closer or further steps is $\left( \frac{k!}{(k/2)!(k/2)!} \right) \le 2^k$, so

$$\mathrm{tr} A^k \le n \cdot (d-1)^{k/2} \cdot 2^k$$

For a lower bound, note that the graph is $d$-regular so $\|A\| = d$ and since $k$ is even

$$\mathrm{tr} A^k \ge \|A^k\| = d^k$$

Together these give $k \le \log_{d/4} n$.

Combining this with the renormalization trick from the previous lecture[1] we can also get a similar bound in the case where the graph has $n$ vertices and average degree $d$, but it is no longer necessarily $d$-regular.

The bound in the irregular case was obtained by Alon-Hoory-Linial('02). They proved that

$$\text{girth} \leq 2 \log_{d-1} n$$

## 1.2 Hypergraphs

A hypergraph on $n$ vertices is a collection of subsets (called hyperedges) of the $n$ vertices. It is called $k$-uniform if each hyperedge has cardinality $k$. A graph corresponds to a 2-uniform hypergraph.

The notion of graph cycle does not have an immediate generalization to hypergraphs. We will choose to work with the following

**Definition:** A *cycle* or *even cover* in a hypergraph is a collection of hyperedges such that every vertex of the hypergraph appears in an even number of hyperedges in the collection, possibly zero.

We can now ask an analogous question to the graph Moore bound. Naor-Verstraëte('05) showed that any $k$-uniform hypergraph on $n$ vertices with $m \gtrsim n^{k/2}$ hyperedges has an even cover of size $O(\log n)$. Let us restrict our discussion to 4-uniform hypergraphs, so the Naor-Verstraëte bound is $m \gtrsim n^2$. Later Feige conjectured that if the number of hyperedges $m$ satisfies $m \gtrsim n^2/l$ then the the hypergraph has an even cover of size $O(l \log n)$.

Let us warm-up by proving the following

**Theorem 1.** Any 4-uniform hypergraph on $n$ vertices with $m \gtrsim n^2$ hyperedges has a cycle of size $O(\log n)$.

*Proof.* Consider the graph on $n^2$ vertices with vertices $\{i, j\}$ where we draw an edge between the vertices corresponding to $\{i, j\}$ and $\{k, l\}$ iff $\{i, j, k, l\}$ is an edge in the hypergraph. If

$$(x_1, y_1) \mapsto (x_2, y_2) \mapsto \ldots \mapsto (x_n, y_n) \mapsto (x_1, y_1)$$

is a cycle in the graph, this means we have hyperedges $\{x_1, y_1, x_2, y_2\}$, $\{x_2, y_2, x_3, y_3\}, \ldots, \{x_n, y_n, x_1, y_1\}$ in the hypergraph, so each element appears in an even number of hyperedges. The issue is that it might be the case that each hyperedge appears multiple times[2], in which case the cycle in the graph does not give any information about the original hypergraph[3]. The fix is simple: for each hyperedge choose just *one* way of splitting it into two. For example, if $\{i, j, k, l\}$ is a hyperedge with $i < j < k < l$, we draw an edge only between $\{i, j\}$ and $\{k, l\}$. This way we have one edge for each hyperedge instead of three and now a cycle in the graph truly corresponds to a cycle in the hypergraph. The graph has $n^2$ vertices with $m \geq cn^2$ edges, so the average degree is at least $c$. By the Moore bound for graphs there exists a cycle of length $O(\log n)$, so there exists an even cover in the hypergraph of size $O(\log n)$. $\square$

[1] This simply means that we work with $\Gamma^{-1/2} A \Gamma^{-1/2}$ instead of $A$, where $\Gamma$ is the diagonal matrix with entries $d_{\text{avg}} + \deg v_i$.

[2] Note that each hyperedge has 4 elements which corresponds to $\frac{1}{2}\binom{4}{2} = 3$ edges in the graph.

[3] We could have a cycle $\{a, b\} \mapsto \{c, d\} \mapsto \{e, f\} \mapsto \{a, c\} \mapsto \{b, d\} \mapsto \{g, h\} \mapsto \{a, b\}$ in the graph, which corresponds to hyperedges $\{a, b, c, d\}$, $\{c, d, e, f\}, \{e, f, a, c\}, \{b, d, g, h\}$ and $\{g, h, a, b\}$. Note that this is not an even cover since $a$ appears in three hyperedges.

We are now going to use Kikuchi graphs to prove Feige's conjecture (up to a log factor):

**Theorem 2.** Any 4-uniform hypergraph $H$ on $n$ vertices with $m \geq n^2 \log n / l$ hyperedges has a cycle of length $O(l \log n)$.

*Proof.* Consider the Kikuchi graph $K_l(H)$ having as vertices the subsets of $\{1, 2, \dots, n\}$ with $l$ elements (where $r$ will be specified later) and we draw an edge between $S$ and $T$ iff $S \Delta T = c \in H$, i.e. we draw an edge between $S$ and $T$ if their symmetric difference is a 4-element set appearing as a hyperedge in $H$.

Note that any closed walk $S_1 \mapsto S_2 \mapsto \dots \mapsto S_p \mapsto S_1$ corresponds to a collection $c_1, \dots, c_p$ of hyperedges such that each element in $\{1, 2, \dots, n\}$ appears an even number of times[4] in $c_1, \dots, c_p$. Indeed,

$$c_1 \Delta c_2 \Delta \dots \Delta c_p = (S_1 \Delta S_2) \Delta (S_2 \Delta S_3) \Delta \dots \Delta (S_p \Delta S_1) = \emptyset$$

because $A \Delta A = \emptyset$. Note that $c_1 \Delta c_2 \Delta \dots \Delta c_p$ is precisely the set of elements which appear in an odd number of $c_1, \dots, c_p$, hence the claim.

Let us call a closed walk *trivial* if the corresponding set of hyperedges $c_1, \dots, c_p$ does not contain an even cover.

We are going to count the number of closed walks of length $t \sim \log \binom{n}{l} \sim l \log n$ in $K_l(H)$, the number of trivial closed walks in $K_l(H)$ and show that the latter is less than the former. This will mean that we can find an even cover of the hypergraph of size $O(l \log n)$.

The number of edges in $K_r(H)$ is $m \binom{4}{2} \binom{n-4}{l-2}$ and the number of vertices is $\binom{n}{r}$ so the average degree is

$$d_{\text{avg}} = \frac{m \binom{4}{2} \binom{n-4}{l-2}}{\binom{n}{l}} \sim \frac{ml^2}{n^2}$$

If we denote by $A$ the adjacency matrix of $K_l(H)$, the number of closed walks of length $t \sim l \log n$ (take $t$ even) is given by $\text{tr} A^t$. We have the bound[5]

$$(\text{tr} A^t)^{1/t} \geq \|A\| \geq \frac{1^T A 1}{1^T 1} = d_{\text{avg}}$$

To get an upper bound on the number of trivial closed walks of length $t$, let $A_c$ be the adjacency matrix of the Kikuchi graph corresponding only to element $c \in H$, i.e. $A_c$ at position $(S, T)$ has entry $1_{S \Delta T = c}$. In particular note that $A = \sum_{c \in H} A_c$. Here's the trick for counting the trivial closed walks: the numebr of trivial closed walks of length $t$ is precisely given by

$$\mathbb{E} \text{tr} \left( \sum_{c \in H} b_c A_c \right)^t$$

[4] Just like before, it might be the case that this is just because each hyperedge appears an even number of times in $c_1, \dots, c_p$.

[5] Note that for a $N \times N$ symmetric real matrix, if $t \sim \log N$ is even, then

$$\|A\| \leq (\text{tr} A^t)^{1/t} = (\lambda_1^t + \dots + \lambda_N^t)^{1/t} \leq N^{1/t} \|A\| \leq c \|A\|$$

so lower bounding the trace by the norm is sharp up to a constant.

where $b_c$ are i.i.d. Rademacher random variables. Indeed, this is a simple consequence of the fact that $\mathbb{E}b_c^{2k+1} = 0$ and $\mathbb{E}b_c^{2k} = 1$.

We can use NCK[6] to obtain

$$\left(\mathbb{E}\mathrm{tr}\left(\sum_{c\in H} b_c A_c\right)^t\right)^{1/t} \lesssim \sqrt{l\log n}\left\|\sum_{c\in H} A_c^2\right\|^{1/2} = \sqrt{l\log n}\cdot\sqrt{d_{\max}}$$

Therefore, if

$$\sqrt{l\log n}\cdot\sqrt{d_{\max}} \lesssim d_{\text{avg}}$$

we have shown that the number of trivial closed walks in $K_l(H)$ of length $t$ is less than the number of closed walks in $K_l(H)$ of length $t$, so there must exist an even cover of size $O(t) = O(l\log n)$.

If the graph is regular, $d_{\max} \sim d_{\text{avg}} \sim \dfrac{ml^2}{n^2}$, then this condition is precisely equivalent to $m \gtrsim n^2 \log n/l$. However, this graph does not need to be regular (and in fact it isn't). To get rid of this regularity assumption, we can proceed as before and use the renormalization trick where we look at $\Gamma^{-1/2}A\Gamma^{-1/2}$ instead of $A$ and show that there are more *weighted* closed walks of length $t$ than *weighted* trivial closed walks of length $t$. We skip the details as they are similar to what we did in the previous lecture. □

## 2 A shorter refutation for random 3-SAT

Let $n$ be the number of variables and let $m$ be the number of clauses. Let $H \subset \binom{[n]}{3}$ be a collection of sets of three elements contained in $[n] = \{1, 2, \ldots, n\}$ chosen uniformly at random. For each $c \in H$, let $f_c^1, f_c^2, f_c^3$ be the negation patterns.[7] We choose these uniformly at random as well.

We are going to show that there exists a polynomial size refutation algorithm for the random 3-SAT if $m \gtrsim n^{1.4}$. This is a result of Feige, Kim and Offek ('06). Note that the bounds we obtained last time could be obtained only when $m \gtrsim n^{1.5}$. The difference in the two is that the algorithm with $m \gtrsim n^{1.5}$ is explicit, whereas in this section we will show that *there exists* an algorithm when $m \gtrsim n^{1.4}$, but this algorithm will not be "computable" in the sense that it will rely on the existence of a decomposition which is not easily computable.

Last time we showed that if $m \gtrsim n^{k/2}\log n$ then we can find a polynomial time refutation for random $k$-XOR. For $m \gtrsim n^{k/2}\log n/l^{k/2-1}$ we showed that there exists a refutation algorithm for the $k$-XOR which runs in $n^{O(l)}$ time.[8] We saw that there was a connection between $k$-XOR and 3-SAT. To the clause $x_1 \vee x_2 \vee x_3$ we associate the polynomial

$$P(x_1, x_2, x_3) = 1 - \frac{1}{8}(1-x_1)(1-x_2)(1-x_3) = \frac{7}{8} + \frac{1}{8}(x_1+x_2+x_3) - \frac{1}{8}(x_1x_2 + x_2x_3 + x_3x_1) - \frac{1}{8}x_1x_2x_3$$

which is 1 if $x_1 \vee x_2 \vee x_3$ is true and 0 otherwise (we identify true and false with 1 and $-1$). We then consider the polynomial

$$\Psi_{3\text{SAT}} = \frac{1}{m}\sum_{c\in H} P(f_c^1, f_c^2, f_c^3)$$

[6] See the second lecture.

[7] So for example, if $c = \{4, 5, 8\}$ and $f_c^1 = 1,\ f_c^2 = -1,\ f_c^3 = 1$, this corresponds to the clause $x_4 \vee \neg x_5 \vee x_8$.

[8] We only showed it for $k = 4$, but the proof worked for any even $k$. The case when $k$ is odd is more difficult and we did not cover it, but the result is still true.

which gives the number of satisfied clauses.

Let us split $\Psi_{3\text{SAT}} = \frac{7}{8} + \Psi_{\text{linear}} + \Psi_{\text{quadratic}} + \Psi_{\text{cubic}}$ into linear, quadratic and cubic terms and analyze them one by one.

We have

$$\Psi_{\text{linear}} = \frac{1}{8m} \sum_{c \in H} (f_c^1 + f_c^2 + f_c^3)$$

**Lemma 3.** With probability at least 2/3

$$\Psi_{\text{linear}} \lesssim \sqrt{\frac{m}{n}}$$

*Proof.* Let $\Psi_{\text{linear},i}$ be the number of times $i$ appears with a plus in $\Psi_{\text{linear}}$ minus the number of times it appears with a minus[9], so that

$$\Psi_{\text{linear}} = \frac{1}{8m} \left(\Psi_{\text{linear},1} + \ldots + \Psi_{\text{linear},n}\right)$$

Note that

$$\begin{aligned}
\mathbb{E}\Psi_{\text{linear},i}^2 &= \sum_j \mathbb{E}[\Psi_{\text{linear},i}^2 | x_i \text{ appears } j \text{ times in} \Psi_{\text{linear}}] \cdot \mathbb{P}(x_i \text{ appears } j \text{ times in } \Psi_{\text{linear}}) \\
&= \sum_j \mathbb{E}[(f_1 + f_2 + \ldots + f_j)^2] \cdot \mathbb{P}(x_i \text{ appears } j \text{ times in } \Psi_{\text{linear}}) \quad \text{with } f_j \text{ i.i.d. Rademacher} \\
&= \sum_j j \cdot \mathbb{P}(x_i \text{ appears } j \text{ times in } \Psi_{\text{linear}}) \\
&= \mathbb{E}(\text{number of times } x_i \text{ appears in } \Psi_{\text{linear}}) \\
&= \frac{3m}{n}
\end{aligned}$$

because we have $m$ clauses wih $n$ variables and the clauses are chosen uniformly at random. Then

$$\mathbb{E}\Psi_{\text{linear},i} \leq \left(\mathbb{E}\Psi_{\text{linear},i}^2\right)^{1/2} = \sqrt{\frac{3m}{n}}$$

so $\mathbb{E}\Psi_{\text{linear}} \leq \frac{1}{8m} \cdot n \cdot \sqrt{\frac{3m}{n}}$. We can use Markov's (at the expense of a constant) to infer $\square$

Let us move on to the quadratic term

$$\Psi_{\text{quadratic}} = \frac{1}{8m} \sum_{c \in H} (f_c^1 f_c^2 + f_c^2 f_c^3 + f_c^3 f_c^1)$$

We move on to the quadratic terms

**Lemma 4.** With probability at least 2/3

$$\Psi_{\text{quadratic}} \lesssim \sqrt{\frac{m}{n}}$$

[9] In other words, we count how many times $x_i$ appears as it is in a clause minus the number of times it appears negated.

*Proof.* We write $\Psi_{\text{quadratic}} = \frac{1}{8m} 1^T A 1$ where $A$ is the $n \times n$ which has

$$A_{ij} = \frac{1}{2}(\text{number of clauses } \{i,j,k\} \text{ with } \text{sign}(i) = \text{sign}(j) - \text{number of clauses } \{i,j,k\} \text{ with } \text{sign}(i) \neq \text{sign}(j))$$

Indeed, to see the equality just note that if we have a clause $c$ on $\{i,j,k\}$, if $i,j$ appear with the same sign, i.e. $f_c^1 = f_c^2$, then $f_c^1 f_c^2 = 1$ and otherwise $f_c^1 f_c^2 = -1$. The $1/2$ in the definition of $A_{ij}$ is because both $(i,j)$ and $(j,i)$ contribute to the sum $1^T A 1$.

Therefore

$$\Psi_{\text{quadratic}} = \frac{1}{8m} 1^T A 1 \leq \frac{n}{8m} \lambda_1(A)$$

so we need to estimate the spectral norm of $A$. For each triple $i < j < k$ we distinguish three pairs: pair 1 is $(i,j)$, pair 2 is $(i,k)$ and pair 3 is $(j,k)$. We define the matrices $A_1, A_2, A_3$ by

$$\begin{aligned}(A_l)_{ij} = \frac{1}{2}(&\text{number of clauses } \{i,j,k\} \text{ with } (i,j) \text{ as pair } l \text{ and } \text{sign}(i) = \text{sign}(j) \\ &- \text{number of clauses } \{i,j,k\} \text{ with } (i,j) \text{ as pair } l \text{ and } \text{sign}(i) \neq \text{sign}(j))\end{aligned}$$

Then $A_1, A_2, A_3$ are symmetric matrices with independent entries[10] of mean zero. The variance of each entry can be computed in a similar way to the previous lemma to be equal to

$$\mathbb{E}(\text{number of of times } x_i, x_j \text{ both appear in a clause}) = \frac{3m}{\binom{n}{2}} \sim \frac{m}{n^2}$$

By Füredi-Komlós('80), the largest eigenvalue of such a matrix is less than

$$\sqrt{\text{size of the matrix}} \cdot \sqrt{\text{variance of entries}}$$

with high probability, so

$$\lambda_1(A) \lesssim \sqrt{n} \sqrt{\frac{m}{n^2}} = \sqrt{\frac{m}{n}}$$

with high probability.

□

For the cubic terms we could try to use the bound for random 3-XOR to obtain that

$$\Psi_{\text{3SAT}} \leq \frac{7}{8} + \tilde{O}\left(\sqrt{\frac{n}{m}}\right) + \tilde{O}\left(\sqrt{\frac{n^{3/2}}{m}}\right)$$

but this clearly shows that any such refutation algorithm needs $m \gtrsim n^{1.5}$. We need to do better for the cubic terms.

We are going to use the following result from last time[11]

**Theorem.** Given a 3-uniform hypergraph $H$ on $n$ vertices with $|H| \geq cn\sqrt{n/l} \log n$, then $H$ contains an even cover of size $l \log n$.

[10] The entries with $i \geq j$ are independent.

[11] We only proved the result for 4-uniform hypergraphs (or more generally $k$-uniform hypergraphs with $k$ even), but it is true for any $k$-uniform hypergraph, although the numerology needs to be adjusted accordingly.

Assume that we start with a 3-uniform hypergraph $H$ with $|H| \geq 100cn\sqrt{n/l}\log n$. Find an even cover of size $O(l \log n)$, call it $H_1$. In $H - H_1$, find another even cover of size $O(l \log n)$ and so on until we can no longer do it. We obtain

$$H = H_1 \cup H_2 \cup \ldots \cup H_s \cup H_*$$

We have that $H_1, \ldots, H_s$ are all even covers and $H_*$ has at most $cn\sqrt{n/l}\log n$ elements (otherwise we would apply the theorem above again). Each $H_i$ is unstastifiable with probability at least 1/2. If $H_i$ is unsatisfiable, it contains at least one unsatisfiable equation. Therefore the number of unsatisfiable equations is at least

$$\frac{1}{2}0.99\frac{m}{l \log n}$$

so the fraction of satisfiable equations is $1 - O\left(\frac{1}{l \log n}\right)$. We thus get the bound

$$\frac{7}{8} + \tilde{O}\left(\sqrt{\frac{n}{m}}\right) + \frac{1}{8}\left(1 - O\left(\frac{1}{l \log n}\right)\right) = 1 + \tilde{O}\left(\sqrt{\frac{n}{m}}\right) - \frac{1}{8}O\left(\frac{1}{l \log n}\right)$$

so in order to get a refutation (i.e. the above less than 1), we need $l^2 \lesssim \frac{m}{n}$. We chose $m \gtrsim n\sqrt{n/l}$, so we can take $l \sim n^{1/5}$ to have $l^2 \simeq m/n \simeq \sqrt{n/l}$. For this choice of $l$, we get

$$m \gtrsim n^{1.4}$$