Topics in Extremal Combinatorics - Notes

Asaf Shapira

Contents

1 Lecture 1

1.1 Ramsey numbers of bounded degree graphs: an improved bound

We recall that Ramsey's Theorem states that $2^{t/2} \leq r(K_t) \leq 2^{2t}$. In the first few lectures we will see several variants of this theorem, with the emphasize on using tools/techniques. We proved in the previous course that $r(G) \leq c(\Delta)n$ if G is an n-vertex graph of maximum degree Δ (this was conjectured by Burr and Erdős). The proof used the regularity-lemma so it only gave the very weak bound $c(\Delta) = \text{tower}(\Delta)$. We will now prove a much better bound, due to Graham-Rödl-Rucińnski '01, using an interesting technique that avoids using the regularity lemma. However, the "moral" of the regularity lemma will be important here. We also recall at this point statement of the the regularity lemma and Gowers's lower bound, which can be found in the previous lecture notes. We also recall the so called *triangle Counting Lemma*.

Lemma 1: If A, B, C are three vertex sets and the three bipartite graphs connecting them are ϵ-regular with densities $a, b, c > 2\epsilon$ then they contain at least $(1-2\epsilon)(a-\epsilon)(b-\epsilon)(c-\epsilon)|A||B||C|$ triangles.

Proof: Prove this lemma while observing that we only need one side of the density condition guaranteed by the notion of ϵ-regularity (this simple observation will be important soon). Furthermore, it is enough to assume that only 2 of the pairs are ϵ-regular. ■

In the previous course we actually proved the following more general result. Observe that the number of clusters as well as the parameters depend only on Δ and d and *not* on the size of H.

Lemma 2: For every d and Δ there is $\epsilon = \epsilon(d, \Delta)$ and $c = c(d, \Delta)$ so that if $V_1, \ldots, V_{\Delta+1}$ are all ϵ-regular with $d(V_i, V_j) \geq d$ and if $|V_i| \geq cn$ then they contain a copy of every n-vertex graph H of maximum degree Δ. Furthermore, we can take $\epsilon = (d/2)^{\Delta}/2\Delta$ and $c = 1/\epsilon$.

Observe that the assumption in the above lemma imply that for every $U_i \subseteq V_i$ and $U_j \subseteq V_j$ of sizes at least $\epsilon|V_i|$ and $\epsilon|V_j|$ we have $d(U_i, U_j) \geq d - \epsilon > d/2$. We now prove a "one-sided" version of the above lemma which makes only this weaker assumption.

Lemma 2: Given δ, Δ, set $\epsilon = \delta^{\Delta}/2\Delta$ and $c = 2/\delta^{\Delta}$. If $V_1, ..., V_{\Delta+1}$ have the property that every pair of subsets $U_i \subseteq V_i$ and $U_j \subseteq V_j$ of sizes at least $\epsilon|V_i|$ and $\epsilon|V_j|$ have density at least δ and $|V_i| \geq cn$ then they contain a copy of every n-vertex graph H with maximum degree Δ.

Proof: Suppose H's vertices are $[n]$. Since H is $(\Delta+1)$-colorable, we can fix a coloring into $\Delta+1$ sets and let $\sigma(i)$ denote the color of vertex i. We will greedily embed H into G by placing each vertex i into $V_{\sigma(i)}$. Start "running" the algorithm and mention what needs to happen.

Fix some $0 \leq j \leq n-1$ and suppose we have already found vertices $x_1, \ldots, x_j$ to play the role of $1, \ldots, j$. For every $i > j$ let Y_i^j denote the vertices of $V_{\sigma(i)}$ which can play the role of i given the choice of $x_1, \ldots, x_j$. When $j = 0$ we have $Y_i^0 = V_{\sigma(i)}$.

For every $0 \leq j < i \leq n$ let $b_{j,i}$ denote the number of neighbours of i among $[j]$. We claim that we can pick the vertices $x_1, \ldots, x_j$ so that $|Y_i^j| \geq \delta^{b_{j,i}} |V_{\sigma(i)}|$. This is true when $j = 0$ (since $b_{0,i} = 0$), so we proceed by induction on j. Suppose the claim holds for $j-1$. The definition of Y_j^{j-1} guarantees that any choice of vertex $x_j \in Y_j^{j-1}$ will be connected to the required vertices among $[j-1]$. We just need to pick x_j so that the sets Y_i^j will not shrink by much compared to Y_i^{j-1}. If $(i,j) \notin E(H)$ the choice of x_j does not affect Y_i^{j-1}. Suppose then that $(i,j) \in E(H)$ and note that there can be at most Δ such i. We need to make sure that $|Y_i^j| \geq \delta |Y_i^{j-1}|$. For each such i we know that $|Y_i^{j-1}| \geq \delta^{b(j-1,i)} |V_{\sigma(i)}| \geq \delta^{\Delta} |V_{\sigma(i)}| \geq \epsilon |V_{\sigma(i)}|$ and in particular $|Y_j^{j-1}| \geq \epsilon |V_{\sigma(j)}|$. The assumption on the sets $V_1, \ldots, V_{\Delta+1}$ thus means that Y_j^{j-1} contains at most $\epsilon |V_{\sigma(j)}|$ vertices with less than $\delta |Y_i^{j-1}|$ neighbors in Y_i^{j-1}. Since there are at most Δ relevant i, there are at most $\Delta\epsilon |V_{\sigma(j)}|$ vertices that violate this condition for some i. Hence, there are at least

$$|Y_j^{j-1}| - \Delta\epsilon |V_{\sigma(j)}| \geq \delta^{\Delta} |V_{\sigma(j)}| - \frac{1}{2}\delta^{\Delta} |V_{\sigma(j)}| = \frac{1}{2}\delta^{\Delta} |V_{\sigma(j)}| = n$$

choices of $x_j \in V_{\sigma(j)}$ which will make sure that $|Y_i^j| \geq \delta |Y_i^{j-1}|$ for all i. Furthermore, since we have n choices for x_j we will always be able to pick $x_j \in V_{\sigma(j)}$ that will be distinct from all other vertices chosen from $V_{\sigma(j)}$. ■

We say that G is (σ, δ)-dense if $d(X) \geq \delta$ for every set $X \subseteq V(G)$ of size at least $\delta|V|$. We say that G is bi-(σ, δ)-dense if $d(X, Y) \geq \delta$ for every pairs of disjoint sets $X, Y \subseteq V(G)$ of size at least $\delta|V|$. The following is an easy corollary of the above lemma.

Corollary 4: If G is a bi-$(\delta^{\Delta}/4\Delta^2, \delta)$-dense graph on at least $(4\Delta/\delta^{\Delta})n$ vertices then G contains a copy of every n vertex graph of maximum degree Δ.

There is a very simple *naive* "recursive algorithm" that finds a very sparse subset in a graph that has no "large" bi-dense subgraph. Hence we want to be able to deal with the case of very dense graphs. This is done in the following lemma.

Lemma 5: If G has $4n$ vertices and density at least $1 - 1/8\Delta$ then it contains a copy of every n-vertex graph of maximum degree Δ.

Proof: The density of the of the complement graph $\overline{G}$ is at most $1/8\Delta$. Hence, there are at most $|G|/2$ vertices with a degree (in $\overline{G}$) greater than $|G|/4\Delta$. We remove those vertices (and maybe a few more) and remain with exactly half the vertices. In the remaining graph, every vertex has a degree greater than $|G|/2 - |G|/4\Delta$.

Now we embed the vertices of H one by one. Suppose we embedded $1, \ldots, i$ and we attempt to embed $i+1$. vertex $i+1$ has at most Δ neighbors among $1, \ldots, i$ (it has at most Δ neighbors in general). For each such neighbor, there are at most $|G|/4\Delta$ vertices which are not connected to it. Hence there are at most $|G|/4$ "forbidden" vertices. Therefore we can pick one of $|G|/2 - |G|/4 = |G|/4 \geq n$ vertices, which means we can pick a vertex for $i+1$ which hasn't been picked before. ■

We now show how to actually find a large dense subset in a graph that has no large bi-dense subgraph. The precise statement is the following.

Lemma 6: The following holds for every $0 < \sigma, \delta < 1$ and $s \geq 2/\delta$: If G is $((\sigma/2)^s, \delta)$-dense graph on N vertices, then there is $U \subseteq V(G)$ of size at least $(\sigma/2)^s N$ so that $G[U]$ is bi-$(\sigma, \delta/4)$-dense.

Proof: If G is bi-$(\sigma, \delta/4)$-dense, we are done. Otherwise, there is a pair of vertex sets $U_{1,1}, U_{1,2}$ of size $\geq \sigma N$ such that $d(U_{1,1}, U_{1,2}) < \delta/4$. By averaging, $U_{1,1}$ contains a subset, W_1 of size $(\sigma/2)^s N$ such that $d(W_1, U_{1,2}) < \delta/4$. Again, by averaging, at most half of $U_{1,2}$'s vertices have density $> \delta/2$ in regard to W_1. Let V_2 denote those vertices of $U_{1,2}$ whose density (to W_1) is $\leq \delta/2$. Then, $|V_2| \geq \sigma/2N$ and if it is bi-$(\sigma, \delta/4)$-dense, we are done.

If V_2 is not bi-$(\sigma, \delta/4)$-dense, then there are two subsets of V_2, denote them by $U_{2,1}, U_{2,2}$, of size $(\sigma|V_2| \geq (\sigma/2)^s N$ such that $d(U_{2,1}, U_{2,2}) \leq \delta/4$. As before, take $W_2 \subseteq U_{2,1}$ of size $(\sigma/2)^s N$ such that $d(W_2, U_{2,2}) \leq \delta/4$, and as before define $V_3 \subseteq U_{2,2}$ with density $\leq \delta/2$ to W_2. Then, $|V_3| \geq (\sigma/2)^2 N$. We note that each vertex in V_3 has also density $< \delta/2$ to W_1.

We continue this process for at most $s = 2/\delta$ iterations. If somewhere along the way V_i is bi-$(\sigma, \delta/4)r$-dense, we are done. Otherwise, we've built vertex sets $W_1, \ldots, W_s$ each one of size $(\sigma/2)^s N$ such that every vertex in W_i has at most $(\delta/2)|W_j|$ neighbors in W_j for $j < i$. Therefore, if we take $W = \bigcup_{i=1}^{s} W_i$, there are few edges between the W_i's, the density between them is less than $\delta/2$. Hence, even if every W_i is a clique, the density of W is strictly less than δ, but as $|W| \geq (\sigma/2)^s N$, it is a contradiction to the assumption that G is $((\sigma/2)^s, \delta)$-dense. ■

We are now ready to state and prove our main result.

Theorem: $c(\Delta) \leq \Delta^{O(\Delta^2)}$

Proof: Take $\delta = 1/8\Delta$, $\sigma = \frac{1}{4\Delta^2}(\frac{\delta}{4})^\Delta$ and $s = 16\Delta \geq 2/\delta$. Consider a coloring of K_N with $N = 4\Delta(2/\sigma)^s(4/\delta)^\Delta n$. We need to show that that either the black edges or the red edges span a copy of H, where H is an arbitrary n vertex graph of max degree Δ.

If the red edges form a $((\frac{\sigma}{2})^s, \delta)$-dense then by Lemma 6 we can find a set of vertices of size at least $(\frac{\sigma}{2})^s N = (4\Delta/\delta^\Delta)n$ on which the red edges form a bi-$(\sigma, \delta/4)$-dense=bi-$(\frac{1}{4\Delta^2}(\frac{\delta}{4})^\Delta, \delta/4)$-dense graph. Corollary 4 then say that we can find a red H.

Otherwise, there is a subset of size at least $(\frac{\sigma}{2})^s N \geq 4n$ on which the black density is at least $1 - \delta = 1 - \frac{1}{8\Delta}$. Hence we can use Lemma 5 to find a black H. ■

2 Lecture 2

2.1 Lower bound for linear Ramsey numbers

We now want to prove a $2^{\Omega(\Delta)}n$ lower bound for Ramsey numbers of bounded degree graphs. Let's then try to reverse-engineer the above proof, that is, find a graph G and coloring of K_N, with

$N = 2^{\Omega(\Delta)}n$, so that the above proof will not find a copy of G in the coloring of K_N. Explain why a $o(n)$-blowup of a random graph of size $2^{\Omega(\Delta)}$ is a reasonable approach: The clusters are of size $o(n)$ so they do not contain H (that is, we don't care if they contain density $1 - \frac{1}{8\Delta}$ as in the second case of the proof). We want to make sure there will not be a $(\Delta+1)$-clique as otherwise the first case of the proof will allow to embed H (since all the pairs between the sets will be bi-dense). Therefore, a random graph of size $2^{\Omega(\Delta)}$ is a good choice since it has no Δ-clique. One can of course say that perhaps we do not need *all* pairs of the clusters to be dense (since perhaps there are no edges between some pairs) and perhaps we decide to break some color classes between various clusters. All of these are ruled out by the following fractional generalization of Erōds's lower bound for Ramsey's theorem.

Lemma 1: For every $k \geq 4$ there is a Red/Black coloring of $[K_k]$ so that for all functions $w : [k] \mapsto [0,1]$ with $\sum_{i=1}^{k} w(i) = x > (10^7 + 2)\log k$ we have

$$B = \sum_{(i,j)\in B} w(i)w(j) < 0.51\binom{x}{2} \quad \text{and} \quad R = \sum_{(i,j)\in R} w(i)w(j) < 0.51\binom{x}{2}$$

Proof: Prove that R/B is maximized when the vertices with weight in $(0,1)$ form a B/R clique. Argument is similar to the weight-shifting proof of Mantel's theorem. Take a coloring with no R/B copy of $K_{2\log k}$ and where every set of size $y \geq 10^7 \log k$ has $[0.499\binom{y}{2}, 0.501\binom{y}{2}]$ red/black edges[1].

Suppose w maximizes R. Then at most $2\log k$ vertices have weight in $(0,1)$. All the other $t \geq x - 2\log k \geq 10^7 \log k$ have weight 1 so they contribute at most $0.501\binom{t}{2} \leq 0.501\binom{x}{2}$. The contribution of edges with one vertex having weight in $(0,1)$ is at most $x \cdot 2\log k < 0.09\binom{x}{2}$ so all together $R < 0.51\binom{x}{2}$. Same argument also applies to B.

■

To use the above lemma, we would want our graph to have the property that no matter how one breaks it into pieces, there would be many large pairs with at least one edge between them. The next lemma does exactly that.

Lemma 1: There is an absolute constant $c > 1$ so that the following holds for every $\Delta > \Delta_0$, $t > 100$, $n > n_0(\Delta)$ and $k = c^{\Delta}$. There exists an n-vertex graph H of maximum degree Δ so that for every partition $V = V_1 \cup \ldots \cup V_k$ in which $|V_i| \leq n/t$ we have

$$\sum_{i<j \,:\, e(V_i,V_j)>0} |V_i||V_j| > 0.55\binom{n}{2} \tag{1}$$

Proof: Set $d = \Delta/300$. We pick a random graph $G(m, d/m)$ where $m = 1.01n$. Expected number of edges is $md/2$, so with very high probability we will not have more than dm edges. Observe that a graph with at most $dm = 1.01nd$ edges contains at most $n/100$ vertices of degree more than $300d$.

[1] Prove that such a graph exists using Chernoff saying $P[B(n,p) > (1+\epsilon)np] < 2^{-\epsilon^2 np/3}$.

Hence removing the $n/100$ vertices of highest degree gives a graph on n vertices with maximum degree at most $300d = \Delta$.

Take any partition of the m vertices into $k+1$ sets; one distinguished set B of size $n/100$ (this is the set we will remove) and the other k sets $V_1, \ldots, V_k$ of size at most n/t each. Since $t > 100$ the number of pairs of vertices within the sets $V_1, \ldots, V_k$ is bounded by $t(n/t)^2 < n^2/100$. Since $|B| = n/100$ the number of pairs of vertices with at least one vertex in B is at most $1.01n(n/100) \leq n^2/99$. Hence there are at least $\binom{1.01n}{2} - 2n^2/99$ pairs connecting two vertices from different sets V_i, V_j. Hence, if $V_1, \ldots, V_k$ violate (1), then

$$\sum_{i<j \,:\, e(V_i,V_j)=0} |V_i||V_j| \geq \binom{1.01n}{2} - 2n^2/99 - 0.55\binom{n}{2} \geq 0.1n^2 \,. \tag{2}$$

Therefore, the probability that a fixed partition $B, V_1, \ldots, V_k$ will violate (1) with respect to some fixed choice of pairs V_i, V_j as in (2) is at most $(1 - d/m)^{0.1n^2} = (1 - \Delta/303n)^{0.1n^2}$. Since there are at most $(k+1)^m$ such partitions, and for each such partition there are at most $2^{\binom{k}{2}}$ ways to pick the pairs V_i, V_j in (2), we see that the probability that in some partition $B, V_1, \ldots, V_k$, the sets $V_1, \ldots, V_k$ will violate (1) is at most

$$(k+1)^m 2^{k^2}(1 - \Delta/303n)^{0.1n^2} < c^{2\Delta n} 2^{k^2} e^{-\Delta n/3030} \,.$$

Hence for some $c > 1$ and[2] large enough n this is $o(1)$. We see that with positive probability, the graph has at most dm edges and in every partition $B, V_1, \ldots, B_k$, the sets $V_1, \ldots, V_k$ satisfy (1). Therefore, removing the $n/100$ vertices of highest degree, gives an n-vertex graph of maximum degree Δ that satisfies (1).

All that is left is to combine the above two lemmas to prove our lower bound.

Main Result: There are $c', \Delta_0' > 1$ so that for every $\Delta \geq \Delta_0'$ and large enough $n \geq n_0(\Delta)$ there is an n vertex graph H of maximum degree Δ satisfying $r(H) \geq c'^{\Delta} n$.

Proof: Take $\Delta > \Delta_0$ and $t > 100$ to be chosen later, and suppose $n \geq n_0(\Delta)$. Let H be the graph from Lemma 2 (which satisfies (1)). Let $k = c^{\Delta}$ be the constant from Lemma 2 and let R be the R/B coloring of $[K_k]$ given by Lemma 1. Define a R/B coloring of K_N where

$$N = kn/t = c^{\Delta} n/t \,, \tag{3}$$

as follows. Partition $[N]$ into k sets $U_1, \ldots, U_k$ of size n/t. Inside U_i color black. Color (U_i, U_j) with color $R(i,j)$. Suppose the graph contains a black H (same argument works for red H). Let $V_i = V(H) \cap U_i$. Then $|V_i| \leq |U_i| = n/t$ so $V_1, \ldots, V_k$ satisfies the condition of Lemma 1, hence it must satisfy (1). Set $w(i) = |V_i|/|U_i|$. Then

$$x = \sum_{i=1}^{k} w(i) = \sum_{i=1}^{k} \frac{|V_i|}{|U_i|} = \frac{t}{n} \sum_{i=1}^{k} |V_i| = t > (10^7 + 2) \log k \tag{4}$$

[2] In the inequality we use the assumption that $\Delta > \Delta_0$ to get $k + 1 = c^{\Delta} + 1 \leq c^{2\Delta}$

where we choose $t = 10^8 \log k$ in order to satisfy the last inequality. Note that this means that $N = c^{\Delta} n/10^8 \log(c^{\Delta}) > c'^{\Delta} n$ for some[3] $c' > 1$ and all large enough $\Delta > \Delta'_0$. Lemma 1 then contradicts (1), since it gives

$$\sum_{i<j \,:\, e(V_i,V_j)>0} |V_i||V_j| = \sum_{(i,j)\in B} |V_i||V_j| = \frac{N^2}{k^2} \sum_{(i,j)\in B} w(i)w(j) < \frac{N^2}{k^2}(0.51)\binom{x}{2} < 0.51\binom{n}{2},$$

where the second equality uses (3) which gives $N/k = n/t$, and last inequality uses (3) and (4) which together give $N/k = n/x$.

3 Lecture 3

3.1 Krivelevich's proof that $r(3,k) \geq ck^2/\log^2 k$

Equivalently there is an n-vertex triangle-free graphs with $\alpha(G) = O(\sqrt{n}\log n)$. If we want $G(n,p)$ to be triangle-free, then we need $p = c/n$ but then $G(n,p)$ will contain cn edges and thus contain an independent set of linear size. We thus want to use the alternation method. We want the expected number of triangles (which is $p^3n^3/6$) to be less than $n/2$. We can then remove one vertex from each triangle and get half the vertices. This gives $p \approx n^{-2/3}$. The probability of having an independent set of size k is at most $\binom{n}{k}(1-1/n^{2/3})^{k^2/2}$. If $k \approx n^{2/3}\log n$ then this is $o(1)$.

Let's consider the following more subtle alternation method. Given a set of vertices S, let $M(S)$ denote the number of edges in S, and let $T(S)$ denote the size of the largest collection of edge disjoint triangles $\Delta_1, \Delta_2, \ldots$ with the property that each triangle Δ_i has at least one edge in S. Suppose G has the property that for every subset of vertices S of size k, we have $M(S) > 3 \cdot T(S)$. Then, if we turn G into a triangle free graph G' by iteratively removing triangles from it (thus constructing a collection of edge disjoint triangles), then the above property guarantees that $\alpha(G') < k$.

Enough to prove that there are $0 < \alpha < 1 < \beta$ so that $G(n, \alpha/\sqrt{n})$ satisfies the above property with $k = \beta\sqrt{n}\log n$. Since there are at most $n^k \leq 2^{\beta\sqrt{n}\log^2 n}$ sets S of size k, it is enough to show that for any given S, the probability S violates the above condition is at most $2^{-2\beta\sqrt{n}\log^2 n}$.

Fix S of size k, and let M denote the number of edges in S. Then $\mathbb{E}[M] = pk^2/2 = \frac{1}{2}\alpha\beta^2\sqrt{n}\log^2 n$. By Chernoff[4] we get $\mathbb{P}[M \leq \frac{1}{2}\mathbb{E}[M]] < 2^{-\frac{1}{8}\mathbb{E}[M]}$.

Let T^* be the largest collection of edge disjoint triangles with at least one edge in S. We want $T^* < \frac{1}{3} \cdot \frac{1}{2}\mathbb{E}[M]$. The probability that some fixed choice of $\frac{1}{6}\mathbb{E}[M]$ pairs of vertices in S will form a collection of edge disjoint triangles is at most $n^{\frac{1}{6}\mathbb{E}[M]}(p^3)^{\frac{1}{6}\mathbb{E}[M]}$. Since there are at most $\binom{|S|^2}{\frac{1}{6}\mathbb{E}[M]}$ ways to pick the collection of edges within S, we get that the probability that $T^* \geq \frac{1}{6}\mathbb{E}[M]$ is at most

$$\binom{|S|^2}{\frac{1}{6}\mathbb{E}[M]}(np^3)^{\frac{1}{6}\mathbb{E}[M]} < (6enp^3|S|^2/\mathbb{E}[M])^{\frac{1}{6}\mathbb{E}[M]}.$$

[3] If $c = 1 + \epsilon$ we can take $c' = 1 + \epsilon/2$

[4] $\mathbb{P}[B(n,p) < (1-\epsilon)pn] < e^{-\epsilon^2 pn/2}$

Since $|S| = k$, $p = \alpha/\sqrt{n}$ and $\mathbb{E}[M] = pk^2/2$, this is bounded from above by $(12e\alpha^2)^{\frac{1}{6}\mathbb{E}[M]}$ so if $\alpha = 0.1$ then this is at most $2^{-\frac{1}{6}\mathbb{E}[M]}$. Hence, the probability that S will not satisfy $T^* \leq \frac{1}{3}M$ is at most $2^{-\frac{1}{6}\mathbb{E}[M]} + 2^{-\frac{1}{8}\mathbb{E}[M]} < 2^{-\frac{1}{9}\mathbb{E}[M]}$. Recalling that $\mathbb{E}[M] = \frac{1}{2}\alpha\beta^2\sqrt{n}\log^2 n$ (and that we have already chosen α), we can choose $\beta > 1$ so that $\frac{1}{9}\mathbb{E}[M]$ will be larger than $2\beta\sqrt{n}\log^2 n$, as needed.

Observe that in $G(n, 1/\sqrt{n})$ with very high probability, all vertices v will satisfy $|N(v)| \geq \sqrt{n}$ so the probability of v not belonging to any triangle is at most the probability of $N(x)$ being independent which is thus at most $(1 - 1/\sqrt{n})^{\Theta(n)} < 2^{-\sqrt{n}}$ so we expect all vertices to belong to at least one triangle. This tells us that in this regime we have no chance of efficiently making G triangle free by removal of vertices and thus have to analyze the more subtle process that uses only edge removal.

3.2 Ajtai-Komlós-Szemerédi's $r(3,k) < ck^2/\log k$

Erdős-Szekeres bound of $r(s,t) \leq \binom{s+t-2}{s-2}$ implies that $r(3,k) \leq \binom{k+1}{2}$. Let's prove this bound in an alternative way, which will lead to the improved bound: if G has a vertex of degree k, then we are done, otherwise G has average degree at most k (actually, maximum degree but we ignore this), and it is known that a graph with average degree $r \geq 1$ has an independent set of size at least $n/2t$. One way to prove this fact, it to first observe that at least half the vertices have degree at most $2t$ and then use a greedy algorithm that adds vertices to the independent set and removes their neighbors.

Note that when we look for the independent set in the graph of average degree t (which in the proof for $r(3,t)$ will be $t = \sqrt{n\log n}$) we never use the fact that G is triangle-free. It is easy to see that if we could show that a triangle-free graph with average degree t has an independent set of size $\frac{n}{12t}\log t$ then we would get that any triangle free graph has an independent set of size $c\sqrt{n\log n}$ which would prove that $r(3,k) \leq ck^2/\log k$.

3.3 AKS Intuition

We want to prove that a triangle-free graph with average degree t satisfies $\alpha(G) \geq \frac{n}{12t}\log t$. If $\log t \leq 6$, then we can just use the basic result that $\alpha(G) \geq \frac{n}{2t} \geq \frac{n}{12t}\log t$. So let's assume from now on that $\log t > 6$.

Suppose we find a vertex of degree at least $2t$. It makes sense to remove this vertex and continue by induction since the average degree will significantly decrease. Indeed, the new average degree would be at most

$$\frac{tn - 4t}{n-1} = \frac{tn - t - 3t}{n-1} = t - \frac{3t}{n-1} \leq t(1 - \frac{3}{n}) .$$

While the number of vertices went down by a factor of $(1 - 1/n)$, the average degree went down by a factor of $(1 - 3/n)$, so their ratio increased by $1 + 2/n$. To get a feeling of the effect of one such step, observe that after $n/2$ such iterations we will get a graph with $n/2$ vertices and average degree less than $t/4$. Since $\frac{(n/2)}{12(t/4)}\log(t/4) > \frac{n}{12t}\log t$ (here we use the assumption that t is large) we see that we "win" in this case, since we can apply induction on the smaller subgraph.

So suppose now that all vertices are of degree at most $2t$. Remember that the naive algorithm would pick an arbitrary vertex, add it to the independent set, remove its (at most) $2t$ neighbors, and continue. This, of course, produces an independent set of size $O(n/t)$ as evidenced by the union of several K_{t+1}. So the key idea is to show that *if G is triangle free*, then we can pick a vertex v, so that after removing v it and its $d(v)$ neighbors[5], the average degree will go down by $1 - \frac{d(v)+1}{n}$, that is, by the *same* factor the number of vertices went down[6]. As in the first case, to get a feeling of the effect of this step, suppose we applied it until $n/2$ vertices are left. Since we always had $d \le 2t$ we increased the independent set by $(n/2)/(2t) = n/4t$. Now, and here comes the clincher, the graph spanned by the remaining $n/2$ vertices will have average degree $t/2$ (since t and n always decreased by the same ration). Hence, if we can apply the same step till there are $n/4$ vertices, we will add to the independent set at least $(n/4)/(2(t/2)) = n/4t$ vertices. We see that in each iteration we increase the independent set by the same $n/4t$. If we do this for $\log t$ iterations (remember that we assume that $\log t > 6$) we get the required independent set of size $\Omega((n/t)\log t)$.

How do we find a vertex v as above? We obviously want the d vertices to be adjacent to as many edges as possible. The key observation is that since G is triangle-free, an edge cannot be adjacent to two neighbors of v, hence the number of edges incident with the neighbors of v is exactly the sum of the degrees of v's neighbors. Since the average degree is t it seems like the best we can hope for, is for the sum of the degrees of v's neighbors to be td. Recapping, if we could find such a vertex v, then removing it along with its d neighbors would remove dt edges from G, thus giving us a new graph on $n-d-1$ vertices with average degree

$$\frac{tn-2td}{n-d-1} = \frac{tn - t(d+1) - t(d-1)}{n-d-1} = t\left(1 - \frac{d-1}{n-d-1}\right) < t\left(1 - \frac{d-1}{n}\right) .$$

This is not quite the $t(1 - \frac{d+1}{n})$ that we wanted, but this will turn out to suffice[7]. Hence we just need to prove the following:

Lemma: Every graph with average degree t contains a vertex v satisfying $\sum_{x \in N(v)} d(x) \ge t \cdot d(v)$.

[5] Of course, we add v to the independent set as before.

[6] Note that this is not doable in all graphs of average degree t, e.g. in the union of K_{t+1}

[7] Since $\frac{1-(d+1)/n}{1-(d-1)/n} \approx 1 - 2/n$ we see that n/t didn't remain the same (as in the intuitive explanation we had before) but instead went down by a factor $1 - 2/n$. However, assuming we always remove a vertex of degree t this means that when we are left with $n/2$ vertices, the ration n/t will go down by a factor

$$(1-2/n)(1-2/(n-t))(1-2/(n-2t)) \cdots (1-2/(n/2)) \approx 1 - 2/t ,$$

so after the $\log t$ iterations of the process we will lose only a constant factor. Of course this is false if we remove vertices of degrees $\ll t$, but then we can claim that by the time we are left with $n/2$ vertices we will have produced an independent set of size larger than $n/2t$.

Proof: Set $d_2(v) = \sum_{x \in N(v)} d(x)$. Then $d_2(v)$ is the number of 2-walks starting with v. Thus $\sum_v d_2(v)$ gives the number of ordered triples x, y, z so that $(x, y), (y, z)$ are edges. This last quantity is just $\sum_v d^2(v) \geq nt^2 = t(\sum_v d(v))$. We thus get that $\sum_v d_2(v) \geq t(\sum_v d(v))$ so there must be some v for which $d_2(v) \geq t \cdot d(v)$. ■

3.4 AKS formal proof

Let $\phi(n, t)$ be the least integer so that every triangle free graph of average degree t has an independent set of size $\phi(n, t)$. We need to prove that $\phi(n, t) \geq \frac{n}{12t} \log t$. We apply induction on n, recalling that the result holds for $\log t \leq 6$. Also, if $t \geq n/12$ then $\frac{n}{12t} \log t \leq \log t \leq \log n$, so we can just use the simple bound $r(3, k) \leq k^2$, which implies that every triangle-free graph has an independent set of size $\sqrt{n} \geq \log n$ (regardless of its average degree). We will thus also assume from now on that $t \leq n/12$.

(i) If we removed a vertex then we know that the average degree became at most $t(1 - 3/n)$, so we need to verify[8] that $\phi(n-1, t(1-3/n)) \geq \phi(n, t)$, i.e. that $\frac{n-1}{12t(1-3/n)} \log(t(1 - 3/n)) \geq \frac{n}{12t} \log t$. This is equivalent to showing that

$$\frac{1 - 1/n}{1 - 3/n}(\log t + \log(1 - 3/n)) \geq \log t \; .$$

Since $\frac{1-1/n}{1-3/n} \geq 1 + 2/n$, it is enough to show that $\frac{2}{n} \log t + (1 + 2/n) \log(1 - 3/n) \geq 0$. Since $n \geq 6$ we have $\log(1 - 3/n) \geq -6/n$ so $(1 + 2/n) \log(1 - 3/n) \geq -10/n$, hence the inequality holds when $\log t \geq 6$.

(ii) Suppose we are in the second case. We know that if we removed a vertex of degree d, then $d \leq 2t$ and the average degree went down by a factor $(1 - \frac{d-1}{n})$. On the other hand the number of vertices went down by a factor of $(1 - \frac{d+1}{n})$. We thus need to verify that $1 + \phi(n(1-(d+1)/n), t(1-(d-1)/n)) \geq \phi(n, t)$, that is, that

$$\frac{n(1 - (d+1)/n)}{12t(1 - (d-1)/n)} \log(t(1 - (d-1)/n)) + 1 \geq \frac{n}{12t} \log t \; ,$$

whenever $d \leq 2t$. As we assume that $t \leq n/12$ and $d \leq 2t$, we infer that $d \leq n/4$, guaranteeing that $(1 - \frac{d+1}{n})/(1 - \frac{d-1}{n}) > 1 - \frac{3}{n}$. Thus, dividing both sides by $n/12t$ we need to prove that

$$(1 - 3/n) \log(t(1 - (d-1)/n)) + 12t/n \geq \log t \; ,$$

whenever $d \leq 2t$. The above follows from

$$\frac{12t}{n} \geq \frac{3 \log t}{n} - \log(1 - (d-1)/n) \; .$$

As noted above $d \leq n/4$, implying that $\log(1 - \frac{d-1}{n}) \geq -2(d-1)/n$. Hence we need to verify that $\frac{12t}{n} \geq \frac{3 \log t}{n} + \frac{2(d-1)}{n}$ which holds since we assume that $d \leq 2t$.

[8] If the average degree is smaller than $t(1 - 3/n)$ we of course only gain more, so we can assume that the average degree is exactly $t(1 - 3/n)$. This can also be justified by the fact that $\frac{\log t}{t}$ is decreasing for $t \geq t_0$ which is the case we consider.

3.5 Improved bound for $r(4,n)$

Erdős-Szekeres gives $r(4,n) = O(n^3)$ while probabilistic proofs give $r(4,n) = \Omega((n/\log n)^{2.5})$. We can slightly improve the upper bound by using the improved bound for $r(3,n)$. We will need the following "robust" version of the AKS bound.

Lemma 1: If G has average degree $\leq t$ and at most $nt^{2-\delta}$ triangles, then $\alpha(G) > c\delta(n/t)\log t$.

We now use this lemma to prove that $r(4,n) = O(n^3/\log^2 n)$. Take a graph G on $r = Cn^3/\log^2 n$ vertices and assume it is K_4-free. If some vertex has degree at least $r(3,n)$ then we are done. In particular, this means that G's average degree is at most $t = r(3,n) \leq n^2/\log n$. For each edge (u,v) the vertices of $N(u) \cap N(v)$ must be independent, as an edge in this set will give a K_4 with u and v. If for some u, v we have $|N(u) \cap N(v)| > n$ then we are done, so suppose this is not the case. Then each edges belongs to at most n triangles, and since G has tr edges, G has at most $trn = Cn^6/\log^3$ triangles. This means that G contains at most $rt^{2-\delta} = (Cn^3/\log^2 n)(n^2/\log n)^{2-\delta}$ triangles for some $\delta > 0$ (any $\delta < \frac{1}{2}$ will work) and hence G has an independent set of size $c\delta\frac{r}{t}\log t = c\delta\frac{Cn}{\log n}\log n$, so choosing C large enough gives an independent set of size n.

Proof: Pick every vertex with probability p. Then whp (Chernoff) we will get at least $pn/2$ vertices, (Markov) at most $2p^2tn$ edges (or less if the average degree is actually less than t) and (Markov) at most $3p^3nt^{2-\delta}$ triangles. Let's pick p so that expected number of triangles is about half the expected number of vertices, i.e. $p = 1/t^{1-\delta/2}$. Then removing a vertex from each triangle will give us a triangle free graph G' with (about) pn vertices and average degree (about) $(2p^2tn)/(pn/2) = 4pt$. AKS theorem implies that G' (and thus also G) contains an independent set of size $c\frac{pn}{4pt}\log(4pt) = c\frac{n}{t}\log(t^{\delta/2}) = c\delta\frac{n}{t}\log t$. ■

4 Lecture 4

4.1 Ajtai-Komlós-Szemerédi a-la Shearer

We want to prove that every n-vertex triangle-free graph of average degree d has an independent set of size at least $f(d)n$. Let's try to do this by induction on n and see what kind of f we can pull. Let's "guess" that the $f : [0,\infty) \mapsto [0,1]$ we will choose will be continuous, differentiable, satisfy $f' < 0$ (since increasing average degree decreases $\alpha(G)$) and convex [9]. We use $d_2(x)$ for sum of the degrees of x's neighbors, G_x for the graph we get after removing x and $N(x)$, and d_x for the average degree of G_x. Since G is triangle free, we have

$$e(G_x) = e(G) - \sum_{y \in N(x)} d(y) = \frac{1}{2}nd - d_2(x) \; . \tag{5}$$

[9] The trivial bound is $f(d) = 1/d$ which is convex, so we expect the same here.

Since we assume that f is convex, we have

$$f(d_x) \geq f(d) + (d_x - d)f'(d) \; . \tag{6}$$

For every vertex x, we have the trivial bound $\alpha(G) \geq 1 + \alpha(G_x)$. Averaging this inequality over all x, then using induction and then (6) we deduce that

$$\alpha(G) \geq 1 + \frac{1}{n}\sum_x \alpha(G_x) \geq 1 + \frac{1}{n}\sum_x f(d_x)(n - d(x) - 1) \geq 1 + \frac{1}{n}\sum_x (f(d) + (d_x - d)f'(d))(n - d(x) - 1)$$

which can be rearranged (using $\sum_x d(x) = nd$) into

$$1 + f(d)n - f(d)(d+1) + (d^2 + d - nd)f'(d) + \frac{f'(d)}{n}\sum_x d_x(n - d(x) - 1) \; .$$

Since $d_x(n - d(x) - 1) = 2e(G_x)$ we can use (5) to simply the above into

$$1 + f(d)n - f(d)(d+1) + (d^2 + d - nd)f'(d) + \frac{f'(d)}{n}\sum_x (nd - 2d_2(x)) \; .$$

Cancelling the $\pm ndf'(d)$ terms we get

$$1 + f(d)n - f(d)(d+1) + (d^2 + d)f'(d) + \frac{f'(d)}{n}\sum_x (-2d_2(x)) \; .$$

As in AKS, $\sum_x d_2(x)$ is just the number of 2-walks in G implying that $\sum_x d_2(x) \geq nd^2$. Putting this in the above expression and using the fact that $f'(d) < 0$ we finally get

$$\alpha(G) \geq f(d)n + \left(1 - f(d)(d+1) + f'(d)(d - d^2)\right) \; ,$$

so if we want the induction to work, we need to find an f so that the second term above would vanish, and that f will be continuous, differentiable, convex and satisfy $f'(d) < 0$. It can be checked that $f(d) = \frac{d\ln(d) - d + 1}{(d-1)^2}$ with $f(0) = 1, f(1) = 1/2$ satisfies all these conditions. Furthermore, since for every $d \geq 3$ we have $f(d) \geq \ln(d)/d$, we get the improved upper bound $r(3,n) \leq n^2/\ln n$.

4.2 Erdős-Hajnal Conjecture/Theorem

The lower bound for Ramsey numbers is obtained by taking a random graph. It is natural to ask if this is necessary. One obvious property of random graphs is that they are universal, that is, contain an induced copy of every fixed graph H. Let $hom(G)$ denote the size of the largest homogenous subset of vertices in G. In other words, $hom(G) = \max\{\alpha(G), \omega(G)\}$. The famous Erdős-Hajnal conjecture is that if G is induced H-free then $hom(G) \geq n^c$ for some $c = c(H) > 0$. While it is easy to see that this conjecture holds for many H (convince yourself it holds for $H = P_2$), it is open already for $H = C_5$. The following is still the best known general bound.

Theorem: If G is induced H-free then $hom(G) \geq 2^{c\sqrt{\log n}}$ for some $c = c(H) > 0$.

We first prove the following intermediate lemma.

Lemma 1: Suppose G is induced H free. Then the following holds for every $\delta > 0$: there are two disjoint sets A, B satisfying the following:

- $|A| \geq \delta^n(\frac{n}{h^2})$ and $|B| \geq \delta^n(\frac{n}{2h})$
- Either every $v \in A$ has has at most $\delta|B|$ neighbors in B or every $v \in A$ has at least $(1-\delta)|B|$ neighbors in B.

Proof: Partition $V(G)$ into h sets $V_1, \ldots, V_h$ of equal size n/h and let's try to embed H by placing $i \in V(H)$ into V_i. We are bound to fail at some point. Let us do the embedding in the following "controlled" way. As in previous embedding lemmas, once we pick vertex i, we are restricted to pick vertex $j > i$ only from a subset, denoted[10] by A_j^{i-1}, of the previous vertices from which we could choose j (so initially $A_j^0 = V_j$). So the plan is to pick vertex i so that for any $j > i$ the set of available vertices for vertex j, will not shrink by more than δ, i.e. we want to make sure[11] that $|A_j^i| \geq \delta|A_j^{i-1}|$. Since G is H free, we know that at some point the process will fail. Suppose we fail at step i for the first time. This means that each vertex in A_i^{i-1} has a problematic $j > i$, that is, for each vertex $x \in A$ there is some $j > i$ such that either $(i,j) \in E(H)$ and yet x has at most $\delta|A_j^{i-1}|$ neighbors in A_j^{i-1} or $(i,j) \notin E(H)$ and yet x has at least $(1-\delta)|A_j^{i-1}|$ neighbors in A_j^{i-1}. Hence, there must be $A \subseteq A_i^{i-1}$ of size at least $|A_i^{i-1}|/h \geq \delta^h(n/h^2)$ for which the *same* j is problematic. Setting $B = A_j^{i-1}$ we have thus found the required sets A, B. ■

Let us now define a family of graphs called *Coraphs*: K_1 is a cograph, and the family is closed under disjoint-union and complementation. Observe that we thus also get closure under taking union by adding a complete bipartite graph.

Lemma 2: If G is a cograph[12] then $\alpha(G)\omega(G) \geq n$. In particular, $hom(G) \geq \sqrt{n}$.

Proof: The claim is true for K_1, so we use induction on the length of the sequence of operations for building G. If last operation was complementation then this is clear. So suppose last operation was a disjoint union of G_1 and G_2. Then induction gives $\alpha(G_1)\omega(G_1) \geq n_1$ and $\alpha(G_2)\omega(G_2) \geq n_2$. Since $\omega(G) = \max\{\omega(G_1), \omega(G_2)\}$ and $\alpha(G) = \alpha(G_1) + \alpha(G_2)$ we get

$$\omega(G)\alpha(G) = \omega(G)\alpha(G_1) + \omega(G)\alpha(G_2) \geq \omega(G_1)\alpha(G_1) + \alpha(G_2)\omega(G_2) \geq n_1 + n_2 = n \ ,$$

as needed. ■

[10] A_j^{i-1} means the vertices available for j after $i-1$ iterations

[11] In particular we always maintain $|A_j^i| \geq 1$, so that if the process never stops we will obtain a copy of H.

[12] It can be shown that G is a cograph iff G is induced P_3-free. So this lemma proves the Erdős-Hajnal Conjecture for P_3.

We are now ready to prove Theorem 1.

Proof: We want to prove that every induced H-free graph G on n vertices has a homogenous set of size $f(n)$ where $f(n) = 2^{c\sqrt{\log n}}$ with $c = c(H) > 0$. By Lemma 2 it is enough to find a co-graph of size $f(n)$. The proof proceeds by induction on n.

Assuming G is induced H-free, let A, B be the sets from Lemma 1 where δ will be chosen shortly. Apply induction on A to find a cograph on a set X of size at least $f(|A|) \geq f(\delta^h(n/h^2))$. If $|X| \geq f(n)$ then we are done, so assume $|X| < f(n)$, and remove from B all vertices adjacent to a vertex of X. We thus remove at most $|X| \cdot \delta|B| < f(n)\delta|B|$ vertices from H. Thus, if we take $\delta = 1/2f(n)$ we are guaranteed to be left with a set B' of size at least $|B|/2 \geq \delta^h(n/4h) \geq \delta^h(n/h^2)$, so that there are no edges between B' and X. We now use induction to find a cograph of size at least $f(\delta^h(n/h^2))$ in B'. Since we made sure there are no edges between X and B' we have thus found a cograph of size at least

$$2f(\delta^h(n/h^2)) = 2f\left(\frac{n}{2^h(f(n)^h)h^2}\right) , \tag{7}$$

where we used our choice of δ in the last equality. Hence we just need to verify that there is a $c = c(H) > 0$ so that the function $f(n) = 2^{c\sqrt{\ln n}}$ satisfies

$$2f\left(\frac{n}{2^h(f(n)^h)h^2}\right) \geq f(n) \tag{8}$$

as (7) would then allow us to complete the proof by induction on n. If we plug $f(n) = 2^{c\sqrt{\ln n}}$ into (8) we see that it "almost" works, in the sense that no matter how large c is, we seem to lack an additive error in the exponent. Instead we prove that the function[13]

$$f(n) = e^{\frac{1}{h}\sqrt{\ln n} - M} \text{ where } M = \frac{h\ln(2) + 2\ln(h)}{h} .$$

satisfies (8), which is clearly enough for our needs. Indeed

$$\begin{aligned}
2f\left(\frac{n}{2^h(f(n)^h)h^2}\right) &= 2e^{\frac{1}{h}\sqrt{\ln\left(\frac{n}{2^h(f(n)^h)h^2}\right)} - M} \\
&= \exp(\ln(2) + \frac{1}{h}\sqrt{\ln(n) - h\ln(2) - h\ln(f(n)) - 2\ln(h)} - M) \\
&= \exp(\ln(2) + \frac{1}{h}\sqrt{\ln(n) - h\ln(2) - h\left(\frac{1}{h}\sqrt{\ln n} - M\right) - 2\ln(h)} - M) \\
&= \exp(\ln(2) + \frac{1}{h}\sqrt{\ln(n) - \sqrt{\ln(n)} + Mh - h\ln(2) - 2\ln(h)} - M) \\
&= \exp(\ln(2) + \frac{1}{h}\sqrt{\ln(n) - \sqrt{\ln(n)}} - M)
\end{aligned}$$

[13] Note that $f(2) \leq 2$ so the claim holds for the base of the induction.

so we just need to verify that $\ln(2) + \frac{1}{h}\sqrt{\ln(n) - \sqrt{\ln(n)}} \geq \frac{1}{h}\sqrt{\ln(n)}$, that is, that

$$\frac{1}{h}(\sqrt{\ln(n)} - \sqrt{\ln(n) - \sqrt{\ln(n)}}) \leq \ln(2)\ .$$

But since $h \geq 3$ and $\ln(2) \geq 1/3$ it is enough to prove that $\sqrt{x} - \sqrt{x - \sqrt{x}} \leq 1$ which follows by multiplying both sides by $\frac{\sqrt{x}+\sqrt{x-\sqrt{x}}}{\sqrt{x}}$, that is, from the fact that

$$\frac{(\sqrt{x} - \sqrt{x - \sqrt{x}})(\sqrt{x} + \sqrt{x - \sqrt{x}})}{\sqrt{x}} = 1 \leq \frac{\sqrt{x} + \sqrt{x - \sqrt{x}}}{\sqrt{x}}\ .$$

■

5 Lecture 5

5.1 Erdős-Szemerédi Theorem (Ramsey's Theorem for sparse graphs)

Theorem: If G has ϵn^2 edges then $hom(G) = \Omega(\frac{\log n}{\epsilon \log(1/\epsilon)})$.

Proof: We first recall that $r(s,t) \leq \binom{s+t-2}{s-1}$. Suppose G has ϵn^2 edges. We can clearly assume that G has maximum degree ϵn (check that you see why). We wish to find a homogenous set of largest possible size s, where the exact value of s will be chosen later. Let K be largest IS[14] in G. If $k = |K| \geq s$, then we are done, so suppose $k < s$. Since G has maximum degree ϵn, there are at most $\epsilon k n$ edges connecting K to $V \setminus K$, hence at least[15] $\frac{1}{2}(n-k) \geq n/4$ of the vertices of $V \setminus K$ have at most $2\epsilon k$ neighbors in K. Call these vertices L. The choice of L implies that there are $\sum_{i=i}^{2\epsilon k}\binom{k}{i} < 2\binom{k}{2\epsilon k}$ ways to pick the neighborhood of a vertex from L in K. Thus, there must be a subset, call it L', of at least $n/8\binom{k}{2\epsilon k} \geq n/8\binom{s}{2\epsilon s}$ vertices in L with the same neighborhood, call it K', in K. By the choice of L, we have $|K'| \leq 2\epsilon k$. Hence if L' contains an IS larger than $2\epsilon s > 2\epsilon k$, then adding this IS to $K \setminus K'$ would give an IS larger than K, a contradiction. By Erdős-Szekeres, if L' also has no clique of size s then $|L'| \leq r(s, 2\epsilon s) \leq \binom{s+2\epsilon s}{2\epsilon s}$. But we know that $|L'| > n/8\binom{s}{2\epsilon s}$, implying that $8\binom{s+2\epsilon s}{2\epsilon s}\binom{s}{2\epsilon s} > n/8$. This inequality does not hold when $s = \frac{\log n}{8\epsilon \log(1/\epsilon)}$, thus completing the proof. ■

We now observe that the above estimate is tight. $G(n,\delta)$ only gives $hom(G) = O(\frac{\log n}{\delta})$. Actually, $\omega(G(n,\delta)) = O(\frac{\log n}{\log(1/\delta)})$ and $\alpha(G(n,\delta)) = O(\frac{\log n}{\delta})$. Taking $\delta = \epsilon \log(1/\epsilon)$ we get a graph G satisfying $\omega(G) = O(\frac{\log n}{\log(1/\epsilon)})$ and $\alpha(G) = O(\frac{\log n}{\epsilon \log(1/\epsilon)})$. Only problem is that we have too many edges. So taking $1/\epsilon$ disjoint copies of $\overline{G}$ we get a graph R of density at most ϵ with $\alpha(R) = \frac{1}{\epsilon} \cdot \omega(G) = O(\frac{\log n}{\epsilon \log(1/\epsilon)})$ and $\omega(R) = \alpha(G) = O(\frac{\log n}{\epsilon \log(1/\epsilon)})$.

[14] It is (perhaps) tempting to think that since G is sparse we can in fact always find an IS of size $\Omega(\frac{\log n}{\epsilon \log(1/\epsilon)})$. It is easy to see that this is not the case.

[15] We assume that $k = o(n)$ otherwise we have found an IS of size cn.

5.2 Beck's Theorem (Size-Ramsey number of a path)

Let us start by observing that $s(K_t) = \binom{r(K_t)}{2}$. The $\leq$ part is trivial. Suppose G has the property that every 2-coloring has a monochromatic K_t. We claim that $\chi(G) \geq r(K_t)$. Indeed, if this is false then let R be a 2-coloring of $K_{\chi(G)}$ without a monochromatic K_t, and color all the edges of G that connect color class i and color class j using the color $R(i,j)$. It is easy to check that this coloring of $E(G)$ does not contain a monochromatic K_t. So we infer that $\chi(G) \geq r(K_t)$ which implies that G has at least $\binom{r(K_t)}{2}$ edges.

We know that if G has bounded degree then $r(G) = O(n)$ implying that $s(G) = O(n^2)$. We will prove a theorem of Beck stating that $s(P_n) = O(n)$. This can be extended to trees, but not to graphs of bounded degree 3. It is true however that if G has bounded degree d then $s(G) = O(n^{2-1/d})$.

We will need a consequence of Pósa's famous lemma, whose proof is differed to the end. We use $N(U)$ to denote the neighbors of U *outside* U, and use P_k to denote a path on k vertices.

Lemma 1: If every $U \subset V(G)$, $|U| \leq u$ satisfies $|N(U)| \geq 2|U|$, then G contains P_{3u} .

We know that every graph has a subgraph with min-degree $\frac{1}{2}d(G)$. Let $e_G(X)$ by number of edges in G incident with X. We will need the following generalization.

Lemma 2: Every G has a subgraph H in which $e_H(X) > \frac{1}{2}d(G)|X|$ for every $X \subseteq V(H)$.

Proof: If G itself does not satisfy the condition, then remove from G a set X satisfying $e_G(X) \leq \frac{1}{2}d(G)|X|$. Then $d(G \setminus X) \geq \frac{d(G)n - d(G)|X|}{n-|X|} = d(G)$ implying that continuing with this process we must eventually end up with a non-empty subgraph satisfying the condition. ■

Corollary 1: Suppose every $X \subset Y \subset V(G)$ with $|Y| < 3|X| \leq 3u$ satisfies[16]

$$e(X,Y) \leq \frac{1}{4}d(G)|X| \tag{9}$$

Then every 2-coloring of G contains P_{3u}.

Proof: Let G' be the graph spanned by the popular color. Then $d(G') \geq \frac{1}{2}d(G)$. By Lemma 2, we can find a subgraph H in which every $X \subseteq V(H)$ satisfies

$$e_H(X) > \frac{1}{2}d(G')|X| \geq \frac{1}{4}d(G)|X| \ . \tag{10}$$

We now claim that H satisfies the condition of Lemma 1. Take any $U \subseteq V(H)$ of size at most u and assume $|N(U)| < 2|U|$. Then (9) (with $X = U$ and $Y = U \cup N(U)$) implies that

$$e_H(U) = e_H(U, U \cup N(U)) \leq e_G(U, U \cup N(U)) \leq \frac{1}{4}d(G)|U|$$

which contradicts (10). ■

[16]For $X \subset Y$, we use $e(X,Y)$ to denote $e(X) + e(X, Y \setminus X)$ where $e(X)$ is the number of edges inside X.

Theorem: There is an n-vertex sparse graph satisfying (9) with $u = \alpha n$. Hence $s(P_n) = O(n)$.

Proof: Consider $G(n,p)$ for some $p = c/n$ where $4d < c < c' = 5d$ will be chosen later. Clearly, whp G has at most $c'n$ edges. Also, whp $d(G) \geq 4d$. Hence, if some X, Y violate (9) then they must satisfy $e(X,Y) \geq d|X|$. We now prove that whp, $G(n,p)$ satisfies (9) with $u = \alpha n$. Fix some $1 \leq s \leq \alpha n$, and take $X \subset Y \subset V(G)$ with[17] $|X| = s, |Y| = 3s$. Then

$$\mathbb{P}[e(X,Y) \geq d|X|] \leq p^{ds} \binom{\frac{5}{2}s^2}{ds} \leq \left(\frac{5esc}{2dn}\right)^{ds} ,$$

implying that the probability of having such a pair (of some size $1 \leq s \leq u = \alpha n$) is at most

$$\sum_{s=1}^{\alpha n} \binom{n}{s}\binom{n}{2s}\left(\frac{5esc}{2dn}\right)^{ds} \leq \sum_{s=1}^{\alpha n}\left[\left(\frac{en}{s}\right)\left(\frac{e^2n^2}{s^2}\right)\left(\frac{5esc}{2dn}\right)^{d}\right]^{s} \leq \sum_{s=1}^{\alpha n}\left[\left(\frac{e^3n^3}{s^3}\right)\left(\frac{40s}{n}\right)^{d}\right]^{s} ,$$

where we use the fact that $c \leq 5d$. If $d = 4$ then we see that for small enough $\alpha > 0$, the base is smaller than $1/2$ so the sum is $o(1)$, implying that whp G satisfies all the required properties. ∎

5.3 Pósa's rotation-extension lemma

We now prove Lemma 1. We start with Pósa's Lemma. Suppose P is a longest path in some graph G that starts from a given vertex x_0. Let $A = A(x_0)$ be[18] the vertices of P which are endpoints of a (longest) path P' which can be obtained from P by a sequence of rotations[19]. Let $B = B(x_0)$ be the neighbors *on* P of the vertices in A, that is, the vertices that appear before/after some vertex of A on the path P. Clearly $|B| \leq 2|A|$. Actually, since $x_k \in A$ we have $|B| \leq 2|A| - 1$.

Lemma 2 (Pósa's Lemma): In the notation above, we have $N(A) \subseteq B$.

Proof: Since P is a longest path, a vertex of $a \in A$ cannot have a neighbor not in P, hence $N(A) \subseteq P$, so we just need to prove that if (a,b) is an edge then $b \in A \cup B$. An important observation is that if P' is a rotation of P then all the vertices of P, save for x_i, x_{i+1} and x_k, have the same pair of neighbors along P and P' (but the order of these pairs might switch). Moreover, note that $x_{i+1}, x_k \in A$ and $x_i \in B$.

Consider now the sequence of rotations that ends up with a path P^* whose last vertex is a. If in one of the rotations along the way, b was either x_i, x_{i+1} or x_k then $b \in A \cup B$ and we are done. If not, then by the previous paragraph, the vertex appearing after b on P^*, call it b^*, was also a neighbour of b in P. It is now easy to see that since (a,b) is an edge, we can perform another rotation that will place b^* at the end of the path. But this means that $b^* \in A$ and so $b \in B$. ∎

[17] Note that if a graph satisfies (9) whenever $|Y| = 3|X|$ then it also satisfies (9) whenever $|Y| \leq 3|X|$.

[18] Strictly speaking we should have written $A = A(x_0, P)$ since A depends on the specific path P that we chose to work with.

[19] If $P = x_0, x_1, \ldots, x_k$ then a rotation of P is a path of the form $x_0, x_1, \ldots, x_i, x_k, x_{k-1}, \ldots, x_{i+1}$.

Proof (of Lemma 1): Let P be a longest path, and suppose it starts at x_0. Setting $A = A(x_0)$, $B = B(x_0)$, we know from Lemma 2 that $N(A) \subseteq B$. But since $|B| \leq 2|A| - 1$ the assumption on G imply that $|A| > u$ (which already means that $|P| > u$). Pick any subset $A' \subseteq A$ of size u. Assumption on G, and the fact that $A' \cup N(A') \subseteq P$, now imply that $|P| \geq |A' \cup N(A')| \geq 3u$. ■

5.4 Hamiltonicity of sparse random graphs

Lemma 1: If $G \sim G(n,p)$ and $p \geq \frac{25\ln(n)}{n}$ then with probability $1 - o(1/n)$ every vertex x_0 satisfies $A(x_0) > n/4$.

Proof: The probability that for some $1 \leq k \leq n/4$ a graph $G \in G(n,p)$ contains a set of k vertices with no edge connecting it to another set of size at least $n - 3k$ is at most[20]

$$\sum_{k=1}^{n/4} \binom{n}{k}\binom{n}{3k}(1-p)^{k(n-3k)} \leq \sum_{k=1}^{n/4} n^{4k} e^{-pkn/4} = \sum_{k=1}^{n/4} \left[n^4 e^{-pn/4}\right]^k$$

which is $o(1/n)$ when $p = 25\ln(n)/n$. We claim that if the above holds, then G satisfies the assertion of the lemma. Indeed, suppose $A(x_0) = k \leq n/4$. By Pósa's Lemma we have $N(A(x_0)) \subseteq B(x_0)$. But by the definition of B we have $|B(x_0)| \leq 2|A(x_0)| = 2k$, implying that there is no edge between $A(x_0)$ (whose size is k) and $V \setminus (A(x_0) \cup B(x_0))$ (whose size is at least $n - 3k$). ■

Lemma 2: If $G \sim G(n,p)$ and $p \geq \frac{25\ln(n)}{n}$ then whp, the following holds for every vertex x_0; The longest path in G is longer than the longest path in $G \setminus x_0$. In particular, G has a Hamilton path.

Proof: Fix x_0. We first expose the edges in $G \setminus x_0$ which is just an instance of $G(n-1,p)$. By the previous lemmas with probability $1 - o(1/n)$, we have $A(y_0) > (n-1)/4$ for every vertex $y_0 \in G \setminus x_0$. Let P be a longest path in $G \setminus x_0$ and assume its first vertex is y_0. Suppose that $A(y_0) > (n-1)/4$. If we now expose the edges between x_0 and the rest of G we see that the probability of not having an edge from x_0 to $A(y_0)$ is at most $(1 - 25\ln(n)/n)^{(n-1)/4} = o(1/n)$. Assuming one such edge (x_0, v) exist, we can obtain a path (in G) longer than P by taking a path ending with v and extending it to x_0. Since in both steps of "exposing" G we got the properties we wanted with probability $1 - o(1/n)$ the claims follows.

Since the above holds with respect to x_0 with probability $1 - o(1/n)$ we get from a union bound that it holds for all x_0 whp. ■

Theorem: If $G \sim G(n, 26\ln(n)/n)$ then whp G is Hamiltonian.

Proof: Think of $G \sim G(n, 26\ln(n)/n)$ as the union of independent $G_1 \sim G(n, 25\ln(n)/n)$ and $G_2 \sim G(n, \ln/n)$ (check that you see why we made do so). By Lemma 2, whp G_1 is hamiltonian,

[20] It is clearly enough to consider only sets of size exactly $n - 3k$.

and from Lemma 1, whp we have $|A(x_0)| \geq n/4$ for all x_0. Suppose both properties hold, let P be a Hamilton path and suppose it starts with x_0. Then the probability that G_2 contains no edge between x_0 and $A(x_0)$ is $(1 - \ln/n)^{n/4} = o(1)$. Assuming such an edge (x_0, u) is present we get a Hamilton cycle by taking a Hamilton path ending with u and closing it with (x_0, u). ■

Bound is "tight" in the sense that $G(n, \frac{1}{2}\ln(n)/n)$ is not even connected whp. Actually, whp $G(n, \frac{1}{2}\ln(n)/n)$ has isolated vertices. Exact threshold is around $p = \frac{\log n + \log\log n \pm \omega(1)}{n}$. Mention random process and min-degree 2.

6 Lecture 6

6.1 Dependent random choice basic lemma

Basic Lemma: Let a, d, m, n, r be positive integers. Let $G = (V, E)$ be a graph with $|V| = n$ vertices and average degree $d = 2|E(G)|/n$. If there is a positive integer t such that

$$\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq a \ ,$$

then G contains a subset U of at least a vertices such that every r vertices in U have at least m common neighbors.

Proof: Pick t vertices at random (with repetitions) and let A be the vertices adjacent to all the t chosen vertices. Then by linearity of expectation and convexity, we have

$$\mathbb{E}[|A|] = \sum_v (d(v)/n)^t = n^{-t} \sum_v d(v)^{t-1} \geq d^t/n^{t-1} \ .$$

On the other, if S is a set of r vertices that have less than than m common neighbors, then the probability that $S \subseteq A$ is at most $(m/n)^t$. Hence, the expected number of such sets in A, denoted by Y, is at most $\binom{n}{r}(m/n)^t$. By linearity of expectation and the lemma's assumption we have $\mathbb{E}[|A| - Y] \geq a$, so if we remove a vertex from each set of r vertices with less than m common neighbors we get the required set. ■

For what follows we will need the following embedding lemma (given as home assignment).

Lemma 1: Suppose X, Y are vertex sets so that: every collection of p vertices in X have at least b common neighbors in Y. Then G contains a copy of every bipartite graph with vertex sets A, B satisfying $|A| \leq |X|$, $|B| \leq b$ and where every vertex in B has at most p vertices in A.

We now derive the following corollaries of the above two lemmas.

Theorem: If $H = (A, B, E)$ is a bipartite graph where every vertex in B has degree at most p then $ex(n, H) \leq cn^{2-1/p}$.

Proof: By Lemma 1 we just need to prove that for c large enough, there is a vertex set X of size at least h so that each subset of p vertices of its vertices have at least h common neighbors. If G has $cn^{2-1/p}$ edges, then in the basic lemma we have $d = 2cn^{1-1/p}$, $r = p$, $m = h$ and $a = h$, so it is enough[21] to find a t so that

$$(2c)^t n^{1-t/p} - n^{p-t} h^t \geq h \; .$$

Considering this difference, it is now clear that we should take $t = p$, in which case the inequality becomes $(2c)^p - h^p \geq h$, so taking $c = h$ makes sure the required inequality holds. ■

This bound is believed to be tight. Note that there is a caveat in the above proof since some vertices of the common neighborhood of a p-tuple of vertices from X might be in X itself (resulting in choosing the same vertex twice). Think how this can be fixed (either by first picking a bipartite subgraph or by replacing h with $2h$).

The above analysis is tight in that if we have $o(n^{2-1/p})$ edges then the first term would be $o(1)$ so we would have no chance of finding a set of size h satisfying the condition. So in this case it was easy to "guess" the best bound that this proof can give. We now give a more involved example.

Theorem: Let Q_b denote the b-dimensional Boolean cube. Then $r(Q_b) \leq 2^{3b} = |Q_b|^3$.

Proof: As usual, we look at the popular color, that is, we'll prove that for large enough n, every graph of average degree $n/2$ contains Q_b. By Lemma 1, it is enough to find a subset of size 2^{b-1} so that every set of b vertices have at least 2^{b-1} common neighbors. Hence, in the basic lemma we have $d = n/2$, $r = b$, $m = 2^{b-1}$ and $a = 2^{b-1}$ so

$$(1/2)^t n - n^b (2^{b-1}/n)^t \geq 2^{b-1} \; . \tag{11}$$

Since we know that $n \geq 2^b$ it makes sense to look for $n = 2^{\alpha b}$ with the smallest α. We clearly need to require the second term be smaller than the first, that is $2^{\alpha b - t} \geq 2^{\alpha b^2 - \alpha b t + bt}$ in which case the bound will be basically the first term that is $2^{\alpha b - t}$. So we require $\alpha b - t \geq b$ (so that the first term will be larger than 2^b) and $\alpha(t + 1 - b) \geq b + t/b$. This last requirement implies that t should be of order b so we set $t = \beta b$. Then we need $\alpha \geq 1 + \beta$ and $\alpha \geq \frac{1+\beta/b}{\beta - 1 + 1/b} = \frac{1+o(1)}{\beta - 1 + o(1)}$. We look for the case when both bounds meet that is when $1 + \beta = \frac{1}{\beta - 1}$ i.e. $\beta = \sqrt{2}$. For the sake of simplicity set $\beta = \frac{3}{2}$ (so $t = \frac{3}{2}b$) and $\alpha = 3$ (so $n = 2^{3n}$). It is now easy to check that this choice is valid in (11). ■

We will soon improve this to $r(Q_b) \leq b2^{2b}$. It is conjectured that $r(Q_b) = O(2^b)$.

Theorem: If G has ϵn^2 edges then G has a 1-subdivision of K_b with $b = \epsilon^{3/2}\sqrt{n}$.

[21] We consider here an expression which is *smaller* than the right hand side expression in the lemma. We will also do so in later proofs.

Proof: It is easy to see that if $U = b$ and every 2 vertices in U have at least $\binom{b}{2}$ common neighbors, then we can construct a 1-subdivision of K_b. Taking $d = (2\epsilon n)$, $r = 2$, $a = b$, $m = b^2$ in the basic lemma the situation is

$$(2\epsilon)^t n - \binom{n}{2}\left(\frac{b^2}{n}\right)^t \geq b \ .$$

If we want the second term to be smaller than the first then $b \leq \sqrt{n}$. It is clear that b should also depend on ϵ so let us parameterize $b = (2\epsilon)^\alpha \sqrt{n}$. Again, we clearly need the second term to be smaller than the first so we require

$$(2\epsilon)^t n \geq n^2 (2\epsilon)^{t2\alpha} \ ,$$

and the goal is to find the smallest α for which there is an t so the the above holds. In this case, we will get at least 1/2 of the term. We know that $(2\epsilon)^t \approx 1/\sqrt{n}$ (since, again, $b \leq \sqrt{n}$) implying that $\alpha \geq 3/2$. It is now easy to check that taking $\alpha = 3/2$ (i.e. $b = (2\epsilon)^{3/2} \log n$) and choosing t so that $(2\epsilon)^t = 1/\sqrt{n}$ (i.e. $t = \frac{\log n}{2\log(1/\epsilon)}$) is a valid choice above. ■

Above bound can be improved to $b = \epsilon\sqrt{n}$ which is tight (consider $G(n, \epsilon)$).

6.2 Variant of basic method

We will now revisit the topic of Ramsey number of bounded degree graphs. For simplicity, we will only consider bipartite graphs. We want to prove the following improved bound for Ramsey numbers of bipartite graphs of bounded degree.

Theorem 1: Suppose H is an n-vertex bipartite graph of maximum degree b. If G has density ϵ and $N \geq 64bn/\epsilon^b$ vertices then $H \subseteq G$.

We note that setting $\epsilon = 1/2$ we get the if H is an n-vertex bipartite graph of maximum degree b then $r(H) = O(b2^b n)$. It is known[22] that there are n-vertex bipartite graphs of maximum degree b satisfying $r(H) \geq 2^{cb}n$ so the above above upper bound is (almost) tight. We also get for the cube Q_b that $r(Q_b) = O(b2^{2b})$ which is better than the 2^{3b} bound we proved earlier.

Suppose $\epsilon = 1/2$ (we can usually assume this since one of the colors has density 1/2). If we want to use the Basic Lemma, then we need to find a subset of size $n = \Omega(N)$ where every b-tuple of vertices have $n = \Omega(N)$ common neighbours. The basic lemma cannot give us this since we get $\Theta(n) - \Theta(n^d)\Theta(1)$. In fact, this stronger version of the basic lemma is false!

This calls for a relaxed version, in which we allow some of the subsets to violate the condition.

Lemma 1: Suppose $\epsilon > 0$ and $b \leq n$. If $N \geq 16bn/\epsilon^b$ and G is an $N \times N$ bipartite graph with ϵN^2 edges, then it contains a subset U of size at least $2n + 1$ that contains less than $\binom{U}{b}/(2b)^b$ b-sets with less than n common neighbors.

[22] The proof of this statement is very similar to the one we gave for not necessarily bipartite graphs.

Proof: Pick b vertices in one side and let U be their common neighbors in the other side. Then $\mathbb{E}[|U|] \geq \epsilon^b N$ so with probability at least $\epsilon^b/2$ we have $|U| \geq \epsilon^b N/2$. The expected number of b-sets in U with less than n common neighbors is at most $\binom{N}{b}(n/N)^b \leq n^b/b!$ so the probability to have more than $2n^b/b!\epsilon^b$ such b-sets is at most $\epsilon^b/2$. Thus with positive probability we have a set U satisfying $|U| \geq \epsilon^b N/2$ and containing less than $2n^b/b!\epsilon^b$ b-tuples with less than n common neighbors.

We now claim that above chosen U satisfies both conditions of the lemma. First condition follows from the fact that $N \geq 16bn/\epsilon^b$. For the second condition, we see that the number of b-sets in U with less than n common neighbors is at most

$$\frac{2n^b}{b!\epsilon^b} \leq \frac{(8n/\epsilon)^b}{(2b)^b} \leq \frac{(\epsilon^b N/2b)^b}{(2b)^b} \leq \frac{\binom{\epsilon^b N/2}{b}}{(2b)^b} \leq \frac{\binom{|U|}{b}}{(2b)^b} ,$$

where the first inequality follows from Sterling's approximation, the second from the assumption that $N \geq 16bn/\epsilon^{b+1}$, the third from $\binom{n}{k} \geq (n/k)^k$ and the last from $|U| \geq \epsilon^b N/2$. ■

If we want to use the relaxed lemma above then we of course need to prove a corresponding embedding lemma

Lemma 2: If a bipartite G contains a set U as in Lemma 1, then it contains a copy of every $n \times n$ bipartite graph of maximum degree b.

Proof: We call a set S of size $|S| \leq b$ *good* if there are at most $\binom{|U|}{b-|S|}/(2b)^{b-|S|}$ b-sets $S' \supseteq S$ that have less than n common neighbors. Note that the lemma's assumption means that $\emptyset$ is good. We observe that if there are t vertices for which $S \cup s$ is *not* good, then there are at least[23] $\frac{t}{b-|S|} \cdot \binom{|U|}{b-|S|-1}/(2b)^{b-|S|-1}$ ways to pick a b-set $S' \supseteq S$ with less than n common neighbors. This means that if S is good then we should have $\frac{t}{b-|S|} \cdot \binom{|U|}{b-|S|-1}/(2b)^{b-|S|-1} \leq \binom{|U|}{b-|S|}/(2b)^{b-|S|}$ implying that $t \leq |U|/(2b)$.

Fix an $n \times n$ bipartite graph H of bounded degree b, on vertex sets X and Y. Let's call them $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. We now use the above observation to find an embedding of H into G. The crucial step will be finding an embedding of the vertices of X into U so that for every $y \in Y$ the set $N(y)$ (which is a subset of U assuming we have already embedded X into U) has at least n common neighbors. Note that once we find such an embedding of X, we can greedily embed the vertices of Y in a trivial way. Observe that we do *not* need all b-subsets of the embedded X to have n-common neighbors.

To find an embedding of X as above we claim that we can inductively embed the vertices $x_1, \ldots, x_i$ so that for every $y \in Y$ the set $N(y) \cap \{u_1, \ldots, u_i\}$ has at least n common neighbours,

[23] t choices to pick the new vertex, $\binom{|U|}{b-|S|-1}/(2b)^{b-|S|-1}$ ways to pick the remaining $b - |S| - 1$ vertices (since we assume that the $(|S|+1)$-sets is not good) and dividing by $b - |S|$ because we count each new set $b - |S|$ times, once for each choice of who is the new vertex we first add.

where u_j is the vertex playing x_j. Actually, we claim that we can inductively embed the vertices $x_1, \ldots, x_i$ so that for every $y \in Y$ the set $N(y) \cap \{u_1, \ldots, u_i\}$ will be good. Indeed, suppose we found an embedding of $x_1, \ldots, x_i$ into U and suppose (for simplicity of notation) that $N(x_{i+1}) = \{y_1, \ldots, y_b\}$. Then when picking a vertex u_{i+1} to play x_{i+1} we just need to make sure that the b sets $N(y_1) \cap \{u_1, \ldots, u_{i+1}\}, \ldots, N(y_b) \cap \{u_1, \ldots, u_{i+1}\}$ will be good. Note that for all $y \notin \{y_1, \ldots, y_b\}$ we have $N(y) \cap \{x_1, \ldots, x_i\} = N(y) \cap \{x_1, \ldots, x_{i+1}\}$ so there is no need to worry about them. We know from induction that the b sets $N(y_1) \cap \{u_1, \ldots, u_i\}, \ldots, N(y_b) \cap \{u_1, \ldots, u_i\}$ are good, and by the previous paragraph, for each of them there are at most $|U|/2b$ vertices whose addition makes them non-good. Hence, at least $|U|/2 \geq n$ vertices are such that if we choose any of them as the new vertex u_{i+1}, then all sets $N(y_1) \cap \{u_1, \ldots, u_{i+1}\}, \ldots, N(y_b) \cap \{u_1, \ldots, u_{i+1}\}$ will be good. This means that we can choose a *new* vertex u_{i+1} to play x_{i+1}. Remind them (or ask them) why we know that the base of the induction holds, that is, why for every y we have $N(y) \cap \emptyset$ is good. ∎

Proof (of Theorem 1): Immediate from Lemmas 1 and 2 (check that you see why). ∎

6.3 Degenerate bipartite graphs

We wish to prove the following result

Theorem 1: If $b, s \geq 2$ and G is an $N \times N$ bipartite graph with at least $N^{2-1/(s^3b)}$ edges, then G contains every b-degenerate bipartite graph with at most $N^{1-2/s}$ vertices.

We say that a bipartite graph is (b, n)-nice if every b-tuple of vertices have at least n common neighbours in the other side. Let us restate a lemma proved in the home assignments.

Lemma 1: If a bipartite graph is (b, n)-nice then it contains a copy of every n-vertex b-degenerate bipartite graph.

Theorem 1 follows immediately from Lemma 2 and the next lemma

Lemma 2: Let $b, s \geq 2$ and let G be an $N \times N$ bipartite graph with at least $N^{2-1/(s^3b)}$ edges. Then G contains a bipartite subgraph which is $(b, N^{1-2/s})$-nice.

Proof: Suppose G has a bipartition into sets A, B. Apply the basic lemma with $d = 2N^{1-1/(s^3b)}$, $r = 2bs$, $m = N^{1-2/s}$ and $t = s^2b$. Then

$$\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq N^{1-1/s} - N^{2bs}\frac{1}{N^{2bs}} \geq N^{1-1/s} \ ,$$

So we can find a subset $A' \subseteq A$ of size $|A'| \geq N^{1-1/s}$ so that each $2bs$-subset in A' has at least $N^{1-2/s}$ common neighbors in B.

Pick a random subset T of bs vertices from A' and let B' be their common neighborhood. Note that the properties of A' guarantee that $|B'| \geq N^{1-2/s}$. The probability that some b-set from B with less than $N^{1-2/s}$ common neighbors in A' will belong to B' is at most

$$\binom{N}{b}\left(\frac{N^{1-2/s}}{|A'|}\right)^{bs} \leq \binom{N}{b}\left(\frac{N^{1-2/s}}{N^{1-1/s}}\right)^{bs} < 1 ,$$

so there must be a set T for which B' contains no b-set with less than $N^{1-2/s}$ common neighbors in A'.

We claim that the bipartite graph between A', B' satisfies the condition of the lemma. Clearly every b-set in B' satisfies the condition. Take a b-set S from A'. Letting $N^*_X(S)$ denote the set of common neighbors of S in vertex set X, we have

$$N^*_{B'}(S) = N^*_B(S) \cap N^*_B(T) = N^*_B(S \cup T) \subseteq N^*_B(T) = B' .$$

Since $S \cup T$ is a set of size $b+bs \leq 2bs$, we have (by the properties of A') that $|N^*_B(S \cup T)| \geq N^{1-2/s}$. Since $N^*_B(S \cup T) \subseteq B'$ all these $N^{1-2/s}$ vertices belong to B' implying that $N^*_{B'}(S) \geq N^{1-2/s}$. ■

We now apply Theorem 1 in order to prove a quasi-linear upper bound for the Ramsey number of degenerate bipartite graphs.

Corollary: If H is an n-vertex b-degenerate bipartite graph then $r(H) \leq n^{1+2(b/\log n)^{1/3}} = n^{1+o(1)}$.

Proof: Take a 2-coloring of K_N. Actually, it is enough to take a 2-coloring of the complete $N \times N$ bipartite graph. As usual, in one of the colors we have $\frac{1}{2}N^2$ edges. To get a copy of H via Theorem 1 (which has n vertices) we need $N^{1-2/s} \geq n$ i.e. $N \geq n^{1+\frac{1}{s-2}}$. We need to make sure that $N^{2-1/(s^3)b} \leq \frac{1}{2}N^2$, i.e. that $bs^3 \leq \log N$. So taking $s = (\frac{\log n}{b})^{1/3} \leq (\frac{\log N}{b})^{1/3}$ we see that it is enough to take $N = n^{1+\frac{1}{s-2}} < n^{1+2(b/\log n)^{1/3}}$. ■

7 Lecture 7

7.1 Discrepancy in graphs - Erdős–Goldberg–Pach–Spencer

Suppose G has $\frac{1}{2}\binom{n}{2}$ edges. We want to find a k-vertex subset where the number of edges deviates significantly from $\frac{1}{2}\binom{k}{2}$. Ramsey's theorem tells us that when $k = \frac{1}{2}\log n$ we can get largest possible deviation. So suppose we want $k = \Omega(n)$. Chernoff tell's us that deviation must be $O(n^{3/2})$. Indeed, using $\mathbb{P}[|B(m, 1/2) - m/2| > t] \leq 2e^{-2t^2/m}$ we get that the probability that a k vertex set spans a number of edges that deviates from $\binom{k}{2}$ by at least $n^{3/2}$ is $e^{-2n^3/k^2} \leq e^{-n}$ (since $k \leq n$). Hence probability that some subset has a deviation at least $n^{3/2}$ is $o(1)$.

We now prove that this is tight. Convince yourself that it is enough to find two sets A, B of size $\Theta(n)$ so that $\left|e(A, B) - \frac{1}{2}|A||B|\right| > cn^{1.5}$. To find A, B as above, we partition the graph into two sets of equal size ($= n/2$) X, Y. Pick a random subset $B \subseteq Y$ by taking each $y \in Y$ with probability $1/2$. Suppose a vertex $x \in X$ has d neighbors in Y. Then for any (positive or negative) integer t, a set B satisfies $d(x, B) - |B|/2 = t$ if for some i the set B contains exactly i neighbors of x and $i - 2t$ non-neighbors[24] of x. Hence, the number sets B satisfying this condition is

$$\sum_i \binom{d}{i}\binom{\frac{1}{2}n - d}{i - 2t} = \sum_i \binom{d}{i}\binom{\frac{1}{2}n - d}{\frac{1}{2}n - d + 2t - i} = \binom{\frac{1}{2}n}{\frac{1}{2}n - d + 2t} \le \binom{\frac{1}{2}n}{\frac{1}{4}n} \le \frac{2^{\frac{1}{2}n}}{\sqrt{n/2}}. \qquad (12)$$

Taking a union bound over all $t \in \{-\sqrt{n}/100, \ldots, \sqrt{n}/100\}$ we infer that

$$\mathbb{P}[|d(x, B) - |B|/2| \le \sqrt{n}/100] \le 1/3 \; .$$

This means that the expected number of vertices $x \in X$ which satisfy $|d(x, B) - |B|/2| \ge \sqrt{n}/100$ is at least $\frac{2}{3}|X|$, implying that with probability at least $1/3$ we should get at least $|X|/3 = n/6$ such vertices. The expected size of B is (via Chernoff) highly concentrated around $n/4$ so $\mathbb{P}[|B| < n/8] < 1/3$. Hence, with positive probability we get a set B of size at least $n/8$ so that at least $n/6$ vertices $x \in X$ satisfy $|d(x, B) - |B|/2| \ge \sqrt{n}/100$. At least half of these vertices should have a deviation with the same sign, hence picking those vertices, we get a set A of size at least $n/12$ so that $\left|e(A, B) - \frac{1}{2}|A||B|\right| > cn^{3/2}$.

7.2 Six standard deviations suffice

Suppose $S_1, \ldots, S_n \subseteq [n]$. Then a random coloring achieves[25] discrepancy $O(\sqrt{n \log n})$. A famous theorem of Spencer states that one can improve this to $O(\sqrt{n})$. To this end, we will need some basic properties of the entropy function.

Define $\mathrm{Ent}(X) = \sum_i p_i \log \frac{1}{p_i}$ where p_i is probability that X attains some value a_i. We will use the fact that

1. If X is uniform over a set S then $\mathrm{Ent}(X) = \log(|S|)$
2. $\mathrm{Ent}(X_1, \ldots, X_n) \le \sum_i \mathrm{Ent}(X_i)$ (without assuming the X_i are independent!)
3. If all $p_i \le 2^{-k}$ then $\mathrm{Ent}(X) = \sum_i p_i \log(1/p_i) \ge k \sum_i p_i = k$.

Lemma: $\sum_{k=0}^{pn} \binom{n}{k} < 2^{H(p)n}$, where $H(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$.

Proof: Use properties $(1), (2)$ above to prove this claim. ■

The following is the key step of the proof.

Lemma 1: There is a $\{-1, 0, 1\}$ valued function which gives discrepancy $O(\sqrt{n})$...(trivial since we can always assign the value 0)... but that attains the value 0 at most $\frac{3}{4}n$ times.

[24] Since in this case we indeed have $d(x, B) - \frac{1}{2}|B| = i - \frac{1}{2}(i + (i - 2t)) = t$

[25] If $x_1, \ldots, x_n$ are ± 1 then $\mathbb{P}[\sum x_i > t] \le e^{-t^2/2n}$

Proof: Pick a random $f : [n] \mapsto \{-1, 1\}$ and define (random) vector $b = (b_1, \ldots, b_n)$ where $b_i =$ nearest integer to $\frac{f(S_i)}{20\sqrt{n}}$. Then $b_i = 0$ iff $-10\sqrt{n} < f(S_i) < 10\sqrt{n}$ and $b_i = -3$ iff $-70\sqrt{n} < f(S_i) < -50\sqrt{n}$. Chernoff gives $p_0 = \mathbb{P}[b_i = 0] > 1 - 2e^{-50}$ and for $s > 0$ we have

$$p_s = \mathbb{P}[b_i = s] \leq \mathbb{P}[f(S_i) \geq (2s-1)10\sqrt{n}] < e^{-50(2s-1)^2} ,$$

and the same applies to p_{-s}.

Since $x\log(1/x)$ is decreasing when moving away from $x = 1/e$ we get from the above estimates for $\ldots, p_{-1}, p_0, p_1, \ldots$ that

$$\mathrm{Ent}(b_i) \leq (1 - 2e^{-50})\log(\frac{1}{1-2e^{-50}}) + 2\left(\sum_{s\geq 1} \frac{50(2s-1)^2}{e^{50(2s-1)^2}}\right) < 0.01 .$$

By Property (2) above, this means that $\mathrm{Ent}((b_1, \ldots, b_n)) < n/100$. By property (3) above, there is some $s = (s_1, \ldots, s_n)$ for which we obtain the string s with probability at least $2^{-0.01n}$, or equivalently, b takes the value s at least $2^{0.99n}$ times. Let C be the set of (at least $2^{0.99n}$) functions f which determine this vector s. Take some $f \in C$, and note that (by previous lemma) there are at most $2^{H(\frac{1}{4})n}$ functions $f' \in C$ which agree with f on at most $n/4$ values. Since $H(\frac{1}{4}) < 0.99$ there must be some $f' \in C$ which disagrees with f on at least $n/4$ entries. This means that the function $g = \frac{1}{2}(f - f')$ is a $\{-1, 0, 1\}$ valued functions, which takes the value 0 at most $\frac{3}{4}n$. Furthermore, since f, f' determine the *same* vector s (as $f, f' \in C$) we have $|g(S_i)| = \frac{1}{2}|f(S_i) - f'(S_i)| \leq 10\sqrt{n}$. ■

Lemma 1 cannot be iterated, so we need the following more general result

Lemma 2: If the n sets are over a universe of size r, then there is a $\{-1, 0, 1\}$ valued function which gives discrepancy $C\sqrt{r\log(n/r)}$ and attains the value 0 at most $0.75r$ times.

Proof: Given as home assignment. ■

Proof (of Main Result): Apply Lemma 2 with $r = n$ getting a function f_1 with discrepancy at most $C\sqrt{n}$ and $r \leq 0.75n$ items assigned 0. Applying Lemma 2 again on the remaining r items (with new sets $S'_1, \ldots, S'_n$ obtained by restricting $S_1, \ldots, S_n$ to the r remaining items), we get f_2 with discrepancy at most $C\sqrt{\frac{3}{4}n\log(4/3)}$. In general, at iteration i we find a function f_i with discrepancy at most $C\sqrt{(3/4)^i n\log(4/3)^i} < C\sqrt{i(3/4)^i n} < C\sqrt{n} \cdot i(0.9)^i$, which sums up to $O(\sqrt{n})$. The bound on the total discrepancy (i.e. of $f = f_1 \cup f_2 \cup \cdots$) now follows from triangle inequality. ■

7.3 Tight examples for six-standard-deviations

We now wish to show that Spencer's theorem is tight, that is, that there is a family of sets $S_1, \ldots, S_n \subseteq [n]$ whose discrepancy is $\Omega(\sqrt{n})$. If A is the incidence matrix of $S_1, \ldots, S_n$ and

$f : [n] \mapsto \{-1, 1\}$ is a coloring then $||Af||_\infty$ is the discrepancy. Hence if $||Af||_2^2 = \Omega(n^2)$ then $||Af||_\infty = \Omega(\sqrt{n})$. The matrix A needs to be a 0/1 matrix, but we first find a ± 1 matrix A with the desired properties. If A's columns $A_1, \ldots, A_n$ are orthogonal (i.e. A is a Hadamard matrix) then $Af = \sum_i f_i A_i$ is a linear combination of orthogonal vectors, each of squared norm n, with ± 1 coefficients so $||Af||_2^2 = n^2$. Letting $B = \frac{1}{2}(A + J)$ be a 0/1 matrix, we need to estimate $||Bf||_\infty$. It will suffice to prove that $||(A + J)f||_2^2 = \Omega(n^2)$ since this will mean that $||(A + J)f||_\infty = \Omega(\sqrt{n})$ and hence $||Bf||_\infty = ||\frac{1}{2}(A + J)f||_\infty = \Omega(\sqrt{n})$. Since J has n columns equal to $\overline{1}$ we have $(A + J)f = \sum_i f_i A_i + \sum_i f_i \overline{1}$. If $\overline{1}$ was a column of A, say the first one, this would become

$$(A + J)f = (f_1 + \sum_i f_i)\overline{1} + f_2 A_2 + \ldots, f_n A_n \tag{13}$$

that is, we would again get a linear combination of orthogonal vectors, each of squared norm n, where at least $n-1$ of coefficients are ± 1 (the first one might be 0). Hence $||(A+J)f||_2^2 = (n-1)n$. Finally, in order to guarantee that $A_1 = \overline{1}$ we can just multiply every row[26] i satisfying $A(i, 1) = -1$ by -1.

It thus remains to construct a matrix as above. It is easy to see that starting with $H_2 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, and in general taking

$$H_{i+1} = \begin{pmatrix} H_i & -H_i \\ H_i & H_i \end{pmatrix}$$

we get the desired Hadamard matrix.

Alternative proof. Fix some ± 1 coloring f, and consider Af for a random 0/1 matrix A. If A_i is the i^{th} row of A then $\langle A_i, f\rangle = \chi(A)$. As in (12), for any t and f, the number of ways to pick the i^{th} row of A in a way that $\langle A_i, f\rangle = t$ is at most $\binom{n}{n/2}$, hence with probability at least 1/2 we have $\langle A_i, f\rangle \geq \sqrt{n}/100$. Since rows of A are independent, the probability that all rows will satisfy $|\langle A_i, f\rangle| \leq \sqrt{n}/100$ is smaller than 2^{-n}, and since there are only 2^n colorings f to consider we get that some A satisfies $||Af||_\infty \geq \sqrt{n}/100$.

7.4 Lindsy's lemma

The second proof suggests that a hard example should be "random like". Let's see that a Hadamard matrix is indeed random like. To this end, we compute the sum of entries within each rectangle, that it, the difference between the number of 1 and -1. If x/y are the characteristic vectors of the s rows and t columns then $||x||_2^2 = s$ and $||y||_2^2 = t$. Hence, by Cauchy-Schwartz

$$|x^T Hy|^2 \leq ||x||_2^2 \cdot ||Hy||_2^2 = s \cdot y^T H^T Hy = s \cdot y^T (nI_n)y = stn$$

In particular, the difference between the number of ± 1 in any rectangle is at most $n^{3/2}$. Observe that if s, t are of order n then this means that the fraction of $+1$ is very close to 1/2.

[26]Since this operation does not affect the "Hadamardness" of the matrix.

If we think of the matrix as defining a bipartite graph, then it defines an $n \times n$ bipartite graph where the graph spanned by any pair of large sets has density $1/2 \pm n^{-1/2}$ edges. By the result we proved earlier (regarding discrepancy in graphs), this cannot be improved. Of course, we can also construct a graph with this property by taking $G(n, n, 1/2)$ (as we did in the first item of this lecture) but the Hadamard matrix gives us an explicit example.

7.5 The eigenvalue bound

Let us return to the proof of the lower bound for the Hadamard matrix. Setting $B = \frac{1}{2}(A + J)$ we tried to estimate $||Bf||_2^2$. So suppose B is the incidence matrix of m sets over n elements, so that B is an $m \times n$ matrix. This means that

$$||Bf||_2^2 = (Bx)^T(Bx) = x^T(B^TB)x \ .$$

Since B^TB is an $n \times n$ symmetric matrix it is diagonalizable with real eigenvalues. In fact, B^TB is also PSD implying that its eigenvalues are $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$. Hence, if we write x using the eigenvectors of $v_1, \ldots, v_n$ we get

$$||Bf||_2^2 = (\sum_i \alpha_i v_i)^T(B^TB)(\sum_i \alpha_i v_i) = \sum_i \alpha_i^2 \lambda_i \geq \lambda_n \sum_i \alpha_i^2 = n\lambda_n$$

implying that $||Bf||_\infty \geq \sqrt{\frac{n}{m}\lambda_n}$. We see that proving a lower bound on the discrepancy of B can be reduced to proving a lower bound on the smallest eigenvalue of B^TB.

Exercise: Prove that this eigenvalue bound is very weak for the Hadamard matrix. That is, there is a vector for which the bound is just $\Theta(1)$. Hint: Recall that in (13) we assumed that the f_i are ± 1. Play around with them (using values other than ± 1) to get a constant bound.

7.6 Beck-Fiala Theorem

Given a set system (or equivalently, a hypergraph) $\mathcal{H}$ we denote by $\Delta(\mathcal{H})$ the maximum degree of $\mathcal{H}$, that is, the largest integer Δ so that some element of the ground set belongs to Δ sets.
Theorem: Any set system of maximum degree t (no restriction on size/number of sets) has discrepancy at most $2t - 1$.

Note that the bound is independent of the size of the sets and/or their number! A well-known conjecture states that the above bound can be improved to $O(\sqrt{t})$, which would be best possible due to the above tight examples for Spencer's Theorem. The best known bound is (roughly) $2t - \log^* t$.

Lemma 1: If a set system is such that all sets contain more than t items, yet no item belongs to more than t sets, then there are more items than sets.

Proof: Count set/item incidences. ■

Proof of Beck-Fiala: Call x_i active if $-1 < x_i < 1$. Call a set active if it contains more than t active variables. We are going to iteratively compute assignments to the x_i while maintaining the invariant that active sets have discrepancy 0 and that once a variable becomes non-active its value never changes. We will make sure that at each iteration a new variable becomes non-active.

We start with all $x_i = 0$, which clearly satisfies the condition. Now, as long as there are active sets, write a linear system with the non-active variables as constants (i.e. ± 1) and the active variables as unknowns, requiring the discrepancy of active sets to be 0. This system has a solution with all active $x_i \in [0, 1]$ (the current one) and is under-determined (by Lemma 1), hence the dimension of the (affine) subspace of solutions is at least 1. We can thus "move along a line of solutions" till we reach the boundary of the $[-1, 1]^q$ cube (where q is the number of undetermined variables). Hence there is a solution that assigns at least one new variable the value ± 1. We thus get a new assignment with one less active variable and where all active sets have 0 discrepancy[27].

If we reach a state where there are no more active sets, we just assign the remaining active variables the value 1. It is easy to see that at the end the discrepancy is strictly smaller[28] than $2t$ and hence at most $2t - 1$. ■

8 Lecture 8

8.1 Discrepancy of arithmetic progressions: Roth's $\frac{1}{4}$-Theorem

We want to compute the discrepancy of arithmetic progression within $[n]$. That is, the ground set is $[n]$ and the sets to be colored are all sets of integers that form an arithmetic progression. It is not hard to see that there are about $n^2 \log n$ sets in this set system [29]. The simple random coloring thus gives us an $\sqrt{n \log n}$ upper bound. Our goal now is to prove a result of Roth from 1964 giving an $\Omega(n^{1/4})$ lower bound for the discrepancy of this set system.

For proving a lower bound, it is clear that we cannot consider only progressions of a fixed difference. For example if we color $1, \ldots, k$ red, then $k + 1, \ldots, 2k$ black etc, it is clear that any progression of difference k will have discrepancy at most 1. We do plan, however, to prove the lower bound by considering only progressions of a fixed length $s = \sqrt{n/6}$. Hence, the ground set is now of size n, and we have $m = n \cdot s = O(n^{3/2})$ sets corresponding to the progressions

$$S(p, q) = \{p, p + q, p + 2q, \ldots, p + (s - 1)q\}, \quad 1 \le p \le n \quad \text{and} \quad 1 \le q \le 6s \; .$$

Note that these sets are arithmetic progression that might "wrap-around", but they do not contain the same element twice (since $sq \le 6s^2 \le n$), so proving a lower bound for this set system will imply the same lower bound for arithmetic progression, perhaps with a loss of 1/2. We plan on

[27] We note that at this point some active sets might become non-active.

[28] Since, at worst, x_i was very close to -1 before becoming 1.

[29] n choices for the first element, n choices for the difference, and n/d options for the length if the difference is d.

using the eigenvalue bound $\mathrm{disc}(\mathcal{A}) \geq \sqrt{\frac{n}{m}\lambda_n}$, where λ_n is the smallest eigenvalue of $A^T A$, with A's rows containing the characteristic vectors of the sets of $\mathcal{A}$. Since we have $m = n^{3/2}$ sets, we need to prove that $\lambda_n = \Omega(n)$.

For each $1 \leq q \leq 6s$, let A_q be the $n \times n$ matrix whose p^{th} row is the characteristic vector of $A(p,q)$. So A_q looks like

$$\begin{array}{ccccccccccccccccc}
\multicolumn{5}{c}{\overbrace{\hphantom{1\ 0\ 0\ \cdots\ 0}}^{q}} & \multicolumn{5}{c}{\overbrace{\hphantom{1\ 0\ 0\ \cdots\ 0}}^{q}} & \multicolumn{5}{c}{\overbrace{\hphantom{1\ 0\ 0\ \cdots\ 0}}^{q}} & \\
1 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & \\
0 & 1 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & \cdots \\
0 & 0 & 1 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & \\
 & & \vdots & & & & & \vdots & & & & & \vdots & & &
\end{array}$$

Since A_q is circulant (i.e. $A_{i,j} = A_{i+1,j+1}$) the product of columns i and j equals the product of columns $i+1$ and $j+1$, which implies that $A^T A$ is also circulant. Let A be the $(6s \cdot n) \times n$ adjacency matrix of the progressions we consider here, where A is obtained by stacking the matrices $A_1, \ldots, A_{6s}$ vertically. That is

$$A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_{6s} \end{pmatrix}_{6sn \times n}$$

Our goal is thus to prove that $\lambda_n(A^T A) = \Omega(n)$.

We note that since $A_q^T A_q$ is circulant, then so is $A^T A = \sum_{q=1}^{6s} A_q^T A_q$. It is easy to check that for any k^{th}-root $\zeta = e^{2\pi i k/n}$ the vector $z_k = (1, \zeta, \ldots, \zeta^{n-1})$ is an eigenvalue of any circulant matrix. Furthermore, it is easy to check that these vectors are orthogonal[30] so these are indeed all the eigenvectors of $A^T A$. For each such $z = z_k$ we have $z^*(A^T A)z = z^*\lambda(z)z = n\lambda(z)$, hence it is enough to prove that for every such z we have

$$z^*(A^T A)z = \Omega(n^2) \; . \tag{14}$$

Observe that

$$z^*(A^T A)z = z^*(\sum_{q=1}^{6s} A_q^T A_q)z = \sum_{q=1}^{6s} z^* A_q^T A_q z = \sum_{q=1}^{6s} (A_q z)^* A_q z \; . \tag{15}$$

Fix a q in the RHS. Then

$$(A_q z)^* A_q z = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} A_q(i,j)\zeta^{j-1} \right)^* \left(\sum_{j=1}^{n} A_q(i,j)\zeta^{j-1} \right) = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} A_q(i,j)\zeta^{j-1} \right|^2 .$$

Consider $i = 1$ in the above sum. Since the first row of A_q has $1's$ only in columns $1, q+1, 2q+1, \ldots, (s-1)q+1$ we have $\sum_{j=1}^{n} A_q(1,j)\zeta^{j-1} = |\sum_{k=0}^{s-1} \zeta^{kq}|^2$. By the circulant properties of A_q, for

[30] $z_t^* z_k = \sum_{i=0}^{n-1} e^{2\pi i k/n - 2\pi i t/n} = \sum_{i=0}^{n-1} e^{2\pi i(k-t)/n} = 0$ since for any unit root z we have $\sum_{i=0}^{n-1} z^i = \frac{z^n - 1}{z - 1} = 0$.

any other $i > 1$ we have $\sum_{j=1}^{n} A_q(i,j)\zeta^{j-1} = |\zeta^{i-1}(\sum_{k=0}^{s-1}\zeta^{kq})|^2$. Since $|\zeta| = 1$ we deduce that the contribution of any other row of A_q is the same as the contribution of the first, implying that

$$(A_q z)^* A_q z = n \left|\sum_{k=0}^{s-1} \zeta^{qk}\right|^2 .$$

Therefore, summing over all q in the RHS of (15) gives

$$z^*(A^T A)z = n \sum_{q=1}^{6s} \left|\sum_{k=0}^{s-1} \zeta^{qk}\right|^2 .$$

Let's focus now on the expression $\sum_{q=1}^{6s} \left|\sum_{k=0}^{s-1} \zeta^{qk}\right|^2$. By pigeon-hole, we have $1 \le q_1 < q_2 \le 6s$ so that the angle between ζ^{q_1} and ζ^{q_2} is at most[31] $2\pi/6s$. Taking $1 \le q_0 = q_2 - q_1 \le 6s$ we get that the angle of ζ^{q_0} is in $(-\pi/3s, \pi/3s)$ implying that the angle of $\zeta^{q_0}, \zeta^{2q_0}, \ldots, \zeta^{(s-1)q_0}$ are all in $(-\pi/3, \pi/3)$, hence the real part of each of these s complex number is at least $1/2$. This clearly means that the real part of $\sum_{k=0}^{s-1} \zeta^{q_0 k}$ is at least $s/2$ so

$$\left|\sum_{k=0}^{s-1} \zeta^{q_0 k}\right|^2 \ge s^2/4 = n/24$$

implying that $z^*(A^T A)z \ge n^2/24$, thus establishing (14).

8.2 Discrepancy of arithmetic progressions: Beck's upper bound

We now wish to prove an $O(n^{1/4} \log^C n)$ upper bound for the discrepancy of arithmetic progressions. We will in fact prove only the key lemma/step of the proof (stated as Theorem 1 below), leaving the final coup de gras as a home assignment. This upper bound was later improved to $O(n^{1/4})$ by Spencer and Matoušek, thus matching Roth's lower bound.

Given a hypergraph $\mathcal{H}$ let $\mathcal{H}^t$ be the sub-hypergraph containing the edges of size at least t, and let $\Delta(\mathcal{H})$ denote $\mathcal{H}$'s maximum degree. The intuition behind the following statement is that we try to partition $\mathcal{H}$ into a hypergraph consisting of small edges (which has discrepancy $\tilde{O}(\sqrt{t})$ by standard random coloring) and a hypergraph consisting of edges of small "average" size (not really, see (17)), which can be handled using the pigeonhole principle as in Spencer's Theorem[32].

Theorem 1: If $\mathcal{H}$ is an n-vertex m-edge hypergraph and there is t so that $\Delta(\mathcal{H}^t) \le t$, then

$$\mathrm{disc}(\mathcal{H}) = O(\sqrt{t} \cdot \log n \cdot \sqrt{\log m})$$

[31] Recall that we are now considering an arbitrary vector $z = (1, \zeta, \ldots, \zeta^{n-1})$ implying that ζ can be any k-th root of unity.

[32] For the history buff: the idea of using partial assignments for the study of discrepancy problems was pioneered by Beck.

Lemma 1: Under the assumption of Theorem 1, there is a $\{-1, 0, 1\}$-valued function that attains the value 0 at most $9n/10$ times and induces discrepancy $O(\sqrt{t \cdot \log m})$.

Lemma 1 $\Rightarrow$ Theorem 1: Set $\mathcal{H}_1 = \mathcal{H}$, $H_1 = V(\mathcal{H})$ and $E_1 = E(\mathcal{H})$. Find $f_1 : H_1 \mapsto \{-1, 0, 1\}$ via Lemma 1 satisfying $\text{disc}(S) = O(\sqrt{t \log m})$ for every edge $S \in E_1$. Set $H_2 = \{v \in H_1 : f(v) = 0\}$, $E_2 = \{S \cap H_2 : S \in E_1\}$ and $\mathcal{H}_2 = (H_2, E_2)$. Clearly $\Delta(\mathcal{H}_2^t) \leq \Delta(\mathcal{H}_1^t) \leq t$ so we can apply Lemma 1 again on $\mathcal{H}_2$. Repeat till all vertices are colored. Since $|H_i| \leq (9/10)^{i-1}|H_1|$ we finish after $\log(|H_1|) = \log n$ iterations. Since each iteration produces a coloring with discrepancy $O(\sqrt{t \log |E_i|}) = O(\sqrt{t \log m})$, the total discrepancy is $O(\sqrt{t \log m} \log n)$.

Proof of Lemma 1: Consider a random $-1/1$ function f. By Chernoff $\mathbb{P}[\text{disc}(S) > t] \leq 2e^{-t^2/2|S|}$. Setting $t = 10\sqrt{|S| \log m}$ this is at most $1/2m$, implying that there are 2^{n-1} functions f satisfying

$$f(S) \leq 10\sqrt{|S| \log m} \qquad \forall S \in E(\mathcal{H}) \tag{16}$$

Fix an f satisfying (16). Then it already satisfies the condition of the lemma when $|S| \leq t$. We thus need to take care of edges of size at least t. Denote these edges $S_1, \ldots, S_k$. Recall that $S_1, \ldots, S_k$ form a hypergraph of maximum degree t. Hence

$$\sum_{i=1}^{k} |S_i| \leq tn \tag{17}$$

Given f as above, define vector $v = v(f) \in \mathbb{R}^k$ as follows. For every $1 \leq i \leq k$ we set v_i to be the nearest integer to $\frac{f(S_i)}{20\sqrt{t \log m}}$. Since we assume that $-10\sqrt{|S_i| \log m} \leq f(S_i) \leq 10\sqrt{|S_i| \log m}$ we infer that each entry v_i takes at most $\sqrt{|S_i|/t}$ values. By (17), the number of possible v's is at most[33]

$$\prod_{i=1}^{k} (|S_i|/t)^{1/2} \leq \prod_{i=1}^{k} 2^{|S_i|/2t} = 2^{\frac{1}{2t}\sum_{i=1}^{k}|S_i|} \leq 2^{n/2} \,. \tag{18}$$

We get that at least $2^{n-1}/2^{n/2} \geq 2^{n/3}$ functions f satisfy (16) and have the same vector v. Pick one such f. There are at most $2^{H(1/10)n}$ functions that agree with f on at most $n/10$ coordinates. Since $H(1/10) < 1/3$, there is an f' that disagrees with f on at least $n/10$ coordinates, satisfies (16) and has the same corresponding v. Then $g = \frac{1}{2}(f - f')$ is a $\{-1, 0, 1\}$ valued function that attains the value 0 at most $9n/10$ times. Furthermore, since both f, f' satisfy (16) we infer that for any $|S| \leq t$ we have $g(S) \leq 10\sqrt{t \log m}$. For the sets $S_1, \ldots, S_k$ of size at least t, we derive from the fact $(v(f))_i = (v(f'))_i$ that $g(S_i) \leq 40\sqrt{t \log m}$. ■

8.3 Linear and hereditary discrepancy

Suppose $\mathcal{A}$ is a collection of n sets over $[m]$. Define

$$\text{lindisc}(\mathcal{A}) = \max_{p_i \in [0,1]} \min_{\epsilon_i \in \{0,1\}} \max_{A \in \mathcal{A}} \left| \sum_{i \in A} (\epsilon_i - p_i) \right| \tag{19}$$

[33] Note that it would have been enough to know that each v_i takes at most $|S_i|/2t$ values.

Usual discrepancy corresponds to taking $p_i = 1/2$. Having $\mathrm{disc}(\mathcal{A}) \le K$ means that we can pick $S \subseteq [m]$ so that $||S \cap A| - |A|/2| \le K/2$ for[34] all $A \in \mathcal{A}$ (this is just the set of variables that attain the value 1). But having $\mathrm{lindisc}(\mathcal{A}) \le K$ means that we have this and we can also (say) find $S \subseteq [m]$ so that $||S \cap A| - |A|/3| \le K$ for all $A \in \mathcal{A}$ (by taking all $p_i = 1/3$).

Lemma 1: Suppose $m \ge n$ and $\mathrm{lindisc}(\mathcal{A}|_X) \le K$ for all $X \subseteq [m]$ of size at most n. Then $\mathrm{lindisc}(\mathcal{A}) \le K$.

Proof: Take any $p_1, \ldots, p_m$. The idea is that if all $p_i = 0/1$ then we can just take $\epsilon_i = p_i$. To "reduce" to this case, we find a new assignment p_i^*, so that all but at most n of them are 0/1, and so that $\sum_{i \in A} p_i = \sum_{i \in A} p_i^*$. We then find an assignment for the set F of (at most n) remaining variables using the assumption that $\mathrm{lindisc}(\mathcal{A}|_F) \le K$. It is then easy to see that the combined assignment satisfies $\mathrm{lindisc}(\mathcal{A}) \le K$.

Call p_j fixed if $p_j = 0/1$ otherwise call it unfixed. As long as there are more than n unfixed variables do the following: Let F be the indices of the unfixed variables. Then $|F| > n$. Consider the system of equations

$$\sum_{j \in A \cap F} y_j = 0 \quad \forall A \in \mathcal{A}$$

Since this system is under determined it has (at least) a line of solutions. Let λ be a real satisfying: (i) $0 \le p'_j = p_j + \lambda y_j \le 1$ for every $j \in F$, and (ii) $p'_j = 0/1$ for some $j \in F$, that is, at least one unfixed p_j becomes a fixed p'_j. Then

$$\sum_{j \in A} p'_j = \sum_{j \in A} p_j + \lambda \left(\sum_{j \in A \cap F} y_j \right) = \sum_{j \in A} p_j \tag{20}$$

Continue till there are at most n unfixed variables whose set of indices is F. Let $p_1^*, \ldots, p_m^*$ be the value of the p_i at the end of the process, and note that it must also satisfy (20). Then $p_j^* = 0/1$ for every $j \notin F$. Then we know that there are $\epsilon_j = 0/1$ so that

$$\left| \sum_{j \in A \cap F} (p_j^* - \epsilon_j) \right| \le K \quad \forall A \in \mathcal{A}$$

Now as our final assignment, for every $j \in F$ we take ϵ_j as above, and for $j \notin F$ we take $\epsilon_j = p_j^*$. Recalling (20), we now have for every $A \in \mathcal{A}$

$$\left| \sum_{j \in A} (p_j - \epsilon_j) \right| = \left| \sum_{j \in A \setminus F} (p_j - p_j^*) + \sum_{j \in F \cap A} (p_j - \epsilon_j) \right| = \left| \sum_{j \in A} (p_j - p_j^*) + \sum_{j \in F \cap A} (p_j^* - \epsilon_j) \right| \le K .$$

■

We now define another type of discrepancy, called hereditary discrepancy.

[34] We thus see that $\mathrm{disc}(\mathcal{A}) \le 2 \cdot \mathrm{lindisc}(\mathcal{A})$

$$\text{herdisc}(\mathcal{A}) = \max_{X \subseteq [m]} \text{disc}(\mathcal{A}|_X)$$

For example, if we take our set system to contain all intervals $\{i, \ldots, j\}$ within $[n]$. Then this set system trivially has discrepancy at most 1. But it also has hereditary discrepancy at most 1. As another example, if we take two disjoint sets on size n and taking all subsets with equal number of elements from each side, then this set system has discrepancy 0, but it's hereditary discrepancy is $n/2$ (take the induced system on one side). It also has large linear discrepancy (set p_i on one side to 0 and the other side $1/2$).

Lemma 2: $\text{lindisc}(\mathcal{A}) \leq \text{herdisc}(\mathcal{A})$.

Proof: Suppose $\text{herdisc}(\mathcal{A}) = K$. Take any $p_1, \ldots, p_m \in [0, 1]$ and suppose they have binary expansion using T digits. We will now try to "round" each p_i into an integer $p_i^0 \in \{0, 1\}$. Observe that once we've done that, we can just take $\epsilon_i = p_i^0$ in (19).

The rounding procedure will take T iterations, where each iteration will decrease the number of digits in the binary expansion of p_i. Let $J \subseteq [m]$ be the indices of the p_i whose rightmost bit is 1. By assumption, $\text{disc}(\mathcal{A}|_J) \leq K$ implying that there are $\{\delta_j \in \{-1, +1\} : j \in J\}$ so that

$$\left| \sum_{j \in A \cap J} \delta_j \right| \leq K \qquad \forall A \in \mathcal{A} \tag{21}$$

Let us set $P_j^T = p_j$. Then we define p_j^{T-1} as follows: if $j \notin J$ then $P_j^{T-1} = P_j^T$. Otherwise, $P_j^{T-1} = P_j^T + \delta_j 2^{-T}$, that is, we are "rounding" the numbers P_j^T whose expansion used T digits into a binary number with expansion of size $T - 1$ where the rounding up/down is determined by δ_j. Note that if $j \notin J$ then p_j^T already has expansion using $T - 1$ bits. We get from (21)

$$\left| \sum_{j \in A} (p_j^T - p_j^{T-1}) \right| = \left| \sum_{j \in A \cap J} 2^{-T} \delta_j \right| \leq 2^{-T} K \qquad \forall A \in \mathcal{A}$$

Continuing with this, after $T - 1$ iterations, we eventually end up with $p_j^0 \in \{0, 1\}$. We claim that we can now take $\epsilon_j = p_j^0$ in (19). Indeed, for every $A \in \mathcal{A}$ we have

$$\left| \sum_{j \in A} (p_j^0 - p_j^T) \right| \leq \sum_{i=1}^{T} \left| \sum_{j \in A} (p_j^i - p_j^{i-1}) \right| \leq \sum_{i=1}^{T} 2^{-i} K \leq K \ .$$

The result for general p_i follows from the fact that (19) is a continuous function in $p_1, \ldots, p_m$ and since given any $p_1, \ldots, p_m$ we can find collection of $\{p_1^r, \ldots, p_m^r\}_{r=1}^{\infty}$ which converge to it (namely their binary expansions with increasing accuracy). ■

Corollary: Suppose $\mathcal{A}$ is a family of n sets over $[m]$ and $m \geq n$. Suppose $\text{disc}(\mathcal{A}|_X) \leq K$ for all $X \subseteq [m]$ of size at most n. Then $\text{disc}(\mathcal{A}) \leq 2K$

Proof: Corollary's assumption means that $\mathrm{herdisc}(\mathcal{A}|_X) \leq K$ for all $X \subseteq [m]$ of size at most n. Lemma 2 then implies that $\mathrm{lindisc}(\mathcal{A}|_X) \leq K$ for all $X \subseteq [m]$ of size at most n. Lemma 2 then implies that $\mathrm{lindisc}(\mathcal{A}) \leq K$. Since $\mathrm{disc}(\mathcal{A}) \leq 2 \cdot \mathrm{lindisc}(\mathcal{A})$ the result follows. ∎

Above corollary and Spencer's Theorem now give

Corollary: Any collection of n sets has discrepancy $O(\sqrt{n})$.