Advanced Topics in Random Graphs

Joshua Erde

Department of Mathematics,
TU Graz.

Contents

Preface

These notes were used to lecture a course at TU Graz for masters level students in the summer semester of 2020. The topics come piecemeal from a variety of sources: Sections 1 and 6 follow to some extent the exposition from "The Probabilistic Method" by Alon and Spencer; Section 2 follows the survey paper "Dependent Random Choice" by Fox and Sudakov; Section 3 presents results from papers of Krivelevich and Sudakov, Riordan and also Erde, Kang and Krivelevich; Section 4 owes a great deal to to Galvin's series of lectures on entropy; Section 5 follows in part the survey "The mathod of hypergraph containers" by Balogh, Morris and Samotij as well as the lecture notes on "The method of hypergraph containers" of Morris; Section 7 follows in part the book "Introduction to random graphs" of Frieze and Karoński.

1 The Lovás Local Lemma

1.1 The Local Lemma

In a typical probabilistic proof of a combinatorial result, one has to show that the probability of a certain event is positive. Many of these proofs tend to show more, that the probability is not only positive but very large, often tending to 1 as the 'dimension' of the problem considered grows.

On the other hand, there is a trivial case in which one can show that a certain event holds with positive, but very small, probability. Suppose we have n mutually independent events A_i, each of which hold with probability $p > 0$, then the probability that they all hold simultaneously is at least p^n, which is positive, but may be exponentially small in n.

It is natural to expect that something similar will be true if the events are not entirely independent, but only 'mostly independent', for some sensible definition of 'mostly independent'. One way to define this is as follows.

Definition. Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. A directed graph $D = ([n], E)$is called a *dependency digraph* for the events $A_1, A_2, \ldots, A_n$ if for all i the event A_i is mutually independent of all the events $\{A_j : (i,j) \notin D\}$.

So for example when $A_1, A_2, \ldots, A_n$ are all mutually independent a dependency diggraph is the empty graph E_n. Note that we are not simply insisting that A_i is independent of A_j if $(i,j) \notin E$ (in particular since this is a symmetric property, and so we could use an undirected graph), the property we are checking is stronger, and so it's not sufficient to simply put an edge in D between every pair of dependent events.

We might expect that there are some natural conditions which tell us that when the dependency digraph is sparse enough, there is some positive probability that all the events hold. In the following, to follow standard notations, we will think of the events which all happen with small probability as being the negation of a set of events A_i, which we will denote by $\overline{A_i}$.

Lemma 1.1 (The Lovás Local Lemma)**.** *Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. Suppose that $D = ([n], E)$ is a dependency digraph for the events $\{A_i : i \in [n]\}$ and there exists $x_1, x_2, \ldots x_n \in [0, 1)$ such that*

$$\mathbb{P}(A_i) \leq x_i \prod_{(i,j)\in E} (1 - x_j)$$

for all $i \in [n]$. Then

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) \geq \prod_{i=1}^{n} (1 - x_i).$$

In particular, with positive probability no event A_i holds.

One way to think of these weightings x_i is as pessimistic estimates for the probability $\mathbb{P}(A_i)$. Indeed, if the conditions of the lemma are satisfied, then we always have that $\mathbb{P}(A_i) \leq x_i$, however we need to choose these upper bounds with some slack, since we need to account for the extra term $\prod(1-x_j)$. The larger we choose each individual x_i to be, the larger its contribution as $(1-x_i)$ will be in the products it appears in. So, in order to choose appropriate x_i we have to balance out these two competing concerns.

Often in application the sets A_i satisfy certain symmetric conditions which allow us to simplify the (rather complicated looking) conditions in Lemma 1.1.

Corollary 1.2. *[Symmetric Local Lemma] Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. Suppose that each event A_i is mutually independent of a set of all but at most d of the other A_j (equivalently there is a dependency digraph with all outdegrees less than d), and that $\mathbb{P}(A_i) \leq p$ for all i. If $ep(d+1) \leq 1$ then*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) > 0.$$

Proof. If $d = 0$ then the events are mutually independent and the result follows trivially. Otherwise let $x_i = 1/(d+1) < 1$. There is a dependency digraph $D = (V, E)$ such that all outdegrees are less than d and so

$$x_i \prod_{(i,j)\in E} (1-x_j) \geq \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d.$$

Note that it is a simple check that

$$\left(1 - \frac{1}{d+1}\right)^d = \left(1 + \frac{1}{d}\right)^{-d} > e^{-1}$$

and so

$$x_i \prod_{(i,j)\in E} (1-x_j) > \frac{1}{e(d+1)} \geq p \geq \mathbb{P}(A_i).$$

Therefore by Lemma 1.1 the conclusion holds. □

1.2 The Linear Aboricity of Graphs

Definition. Given a graph G the *aboricity* of G is the minimum number of forests into which the edge set $E(G)$ can be partitioned. A *linear forest* is a forest in which every component is a path, and the *linear aboricity* of a graph, which we denote by $\mathrm{la}(G)$, is the minimum number of linear forests into which the edge set $E(G)$ can be partitioned.

The following simple conjecture is longstanding.

Conjecture 1.3 (The Linear Aboricity Conjecture)**.** *Let G be a d-regular graph. Then*

$$la(G) = \left\lceil \frac{d+1}{2} \right\rceil.$$

Note that since every d-regular graph on n vertices has $nd/2$ edges and every linear forest has at most $n-1$ edges we have that $\text{la}(G) > d/2$, and so the content of the conjecture is to show that every d-regular graph can indeed by decomposed into a small number of forests. Also, since every graph of maximum degree Δ can be embedded into a Δ-regular graph, the conjecture is equivalent to the statement that every G with satisfies $\text{la}(G) \leq \lceil (\Delta(G)+1)/2 \rceil$.

Much work has been done towards the conjecture and the best known bound without a probabilistic argument was that $\text{la}(G) \lesssim 3\Delta(G)/5$. .

It will be convenient to work instead with directed graphs. A *d-regular digragph* is a directed graph in which the indegree and outdegree of every vertex is precisely d. A *linear directed forest* is a directed graph in which every connected component is a directed path and the *dilinear aboricity* of a directed graph D, which we denote by $\text{dla}(D)$, is the minimum number of linear directed forests into which the edge set $E(G)$ can be partitioned. We then have the directed version of the Linear Aboricity Conjecture.

Conjecture 1.4. *Let D be a d-regular digraph. Then*

$$dla(D) = d + 1.$$

Note that since the edges of any connected undirected $2d$-regular graph can be oriented along an Euler cycle, so the the resulting digraph is d-regular, Conjecture 1.4 for d implies Conjecture 1.3 for $2d$.

It is a simple exercise to show that a graph G contains an independent set of size at least $n/(\Delta(G)+1)$. We will require for our proof a lemma that tells us that, at the price of decreasing the size by a constant factor, we can find a large independent set with additional structure

Lemma 1.5. *Let $H = (V, E)$ be a graph with maximum degree Δ, and let $V = V_1 \cup V_2 \cup \ldots \cup V_r$ be a partition of V into r pairwise disjoint sets. Suppose that $|V_i| \geq 2e\Delta$ for each $i \in [r]$. Then there is an independent set $W \subset V$ that contains a vertex from each V_i.*

Proof. Without loss of generality we may assume that $|V_i| = \lceil 2e\Delta \rceil = g$ for each i. We pick a single vertex from each V_i independently and uniformly at random and let W be the union of these vertices. We will show that with positive probability W is independent.

For each edge $f \in E(H)$ let A_f be the event that both ends of f are contained in W. Clearly $\mathbb{P}(A_f) \leq \frac{1}{g^2}$.

As is common in applications of the local lemma, there is a set of mutually independent underlying variables, here given by the vertex v_i chosen in each V_i, such that each of our events is determined by some subset of these variables. In this case a natural choice for a dependency digraph is to join a pair of events if they depend on a common variable (so in fact we end up with a symmetric digraph). It is easy to check that this in fact determines a dependency digraph.

What does this mean in the current application? Well, if the endpoints of f lie in V_i and V_j then A_f depends only on the value of v_i and v_j. Hence in our dependency digraph f will be joined to all the edges which have an endpoint in V_i or V_j

There are at most $\Delta|V_i \cup V_j| \leq 2g\Delta$ edges meeting $V_i \cup V_j$ and hence there is a dependency digraph for the events A_f in which the maximum degree is $< 2g\Delta$ (a strict inequality as f meets $V_i \cup V_j$). Since $e.2g\Delta.(1/g^2) = 2e\Delta/g < 1$ we have by Corollary 1.2 that with positive probability none of the events A_f hold. However this means that W is an independent set containing a vertex from each V_i. □

We note at this point that Lemma 1.5 enables us to prove Conjecture 1.4 for digraphs with no short directed cycle. The *directed girth* of a graph is the minimum length of a directed cycle in that graph.

Theorem 1.6. *Let $D = (V, E)$ be a d-regular directed graph with directed girth $g \geq 8ed$. Then*

$$dla(D) = d + 1.$$

Proof. We first use Hall's theorem to decompose D into d-pairwise disjoint 1-regular spanning subgraphs, $D_1, D_2, \ldots, D_d$. Strictly we form a bipartite graph on (A, B) where $A = B = V$ and let (a, b) be an edge if and only if (a, b) is a directed edge in D. Then this bipartite graph is d-regular and so we can decompose it into d pefect matchings, each of which correspond to a 1-regular spanning directed subgraph.

Each D_i is a union of vertex disjoint directed cycles, $C_{i_1}, C_{i_2}, \ldots C_{i_{r_i}}$. Let $E_1, E_2, \ldots, E_r$ the edge sets of each of these cycles, taken over each $1 \leq i \leq d$. We have that $\{E_i \colon i \in [r]\}$ is a partition of the edge set of D and by the girth condition each $|E_i| \geq g \geq 8ed$.

We consider the (undirected) line graph L of D, that is the graph whose vertex set is E and two edges are adjacent if and only if they share a vertex. Note that L is $4d - 2$ regular and $\{E_i \colon i \in [r]\}$ is now a partition of the vertex set of L. Since $|E_i| \geq 8ed \geq 2e(4d-2)$ we can apply Lemma 1.5 to L to find an independent set in L containing an element of each E_i. However this corresponds to a matching M in D containing at least one edge from each cycle C_{i_j}.

Hence if we consider the subgraphs $D_1 \setminus M, D_2 \setminus M, \ldots, D_d \setminus M, M$ we see that each $D_i \setminus M$ is a linear directed forest, and M is a matching, and between them they cover the edges of D. Hence

$$\mathrm{dla}(D) \leq d + 1.$$

Finally we note that, as before, D has $|V|d$ edges and each directed linear forest can have at most $|V| - 1$ edges, and so

$$\mathrm{dla}(D) \geq \frac{|V|d}{|V| - 1} > d.$$

Therefore $\mathrm{dla}(D) = d + 1$ as claimed.

□

In order to prove a result for general digraphs we show that we can decompose almost all of the edges of a regular digraph into a relatively small number of almost regular digraphs with large girth, together with a small remainder. To do so we need the following technical lemma, which also uses the Local Lemma in its proof.

Lemma 1.7. *Let $D = (V, E)$ be a d-regular directed graph, where d is sufficiently large, and let p be an integer such that $10\sqrt{d} \leq p \leq 20\sqrt{d}$. Then there is a p-colouring of V, $f : V \to [p]$, such that, for each $v \in V$ and each $i \in [p]$ the numbers*

$$N^+(v,i) = |\{u \in V \,:\, (v,u) \in E \text{ and } f(u) = i\}|$$

and

$$N^-(v,i) = |\{u \in V \,:\, (u,v) \in E \text{ and } f(u) = i\}|$$

satisfy

$$\left|N^+(v,i) - \frac{d}{p}\right|, \left|N^-(v,i) - \frac{d}{p}\right| \leq 3\sqrt{\frac{d}{p}\log(d)}.$$

Proof. We pick a random p-colouring $f : V \to [p]$ by choosing $f(v)$ for each $v \in V$ independently and uniformly at random from $[p]$. For each $v \in V$ and $i \in [p]$ let $A^+_{v,i}$ be the event that

$$\left|N^+(v,i) - \frac{d}{p}\right| > 3\sqrt{\frac{d}{p}\log(d)}$$

and similarly for $A^-_{v,i}$. We have that $N^+(v,i)$ is a binomial random variable with expectation d/p, so if we let $t = 3\sqrt{\frac{d}{p}\log(d)}$ then, by the Chernoff bounds we have that

$$\mathbb{P}(A^+_{v,i}) < 2e^{-\frac{t^2}{2(\frac{d}{p}+\frac{t}{3})}} \leq 2e^{-\frac{9\frac{d}{p}\log(d)}{3\frac{d}{p}}} \leq d^{-3}$$

and similarly for $A^-_{v,i}$. As before we have a set of mutually independent variables, the colour $f(v)$ of each vertex, which determine our events. Clear $A^\pm_{v,i}$ is determined by the set of variables $\{f(w) \colon w \in N^\pm(v)\}$. Since the in and out degree of each vertex is d, it follows that $A^\pm_{v,i}$ will share variables with at most $2d^2$ many other $A^\pm_{u,j}$ for any fixed j. Therefore there is a dependency digraph for these events with maximum degree $\leq 2d^2p$. Since

$$e\frac{1}{d^3}(2d^2p + 1) \leq 1$$

we have that by Corollary 1.2 there is a non-zero probability that none of the events $A^+_{v,i}, A^-_{v,i}$ happen. Therefore there is a colouring f satisfying the required properties.

□

We are now ready to argue the general case.

Theorem 1.8. *There exists a constant $c > 0$ such that for every d-regular digraph D*

$$dla(D) \leq d + cd^{\frac{3}{4}}(\log(d))^{\frac{1}{2}}.$$

Proof. Let $D = (V, E)$ be an arbitrary d-regular digraph. Let p be a prime satisfying $10\sqrt{d} \leq p \leq 20\sqrt{d}$ (which exists by Bertrand's postulate). By Lemma 1.7 there is a p-colouring f of V

such that the conclusions of the lemma hold. For each $i \in [p]$ let $D_i = (V, E_i)$ be the spanning subgraph of D defined by

$$E_i = \{(u,v) \in E : , f(v) \equiv f(u) + i \bmod(p)\}.$$

By assumption we have that the maximum outdegree Δ_i^+ and the maximum indegree Δ_i^- of G_i are at most

$$\frac{d}{p} + 3\sqrt{\frac{d}{p}\log(d)}.$$

Moreover, for each $i \neq p$, the length of every directed cycle in G_i is divisible by p, and so each G_i has directed girth $g_i \geq p$. It is a simple exercise to see that G_i can be completed, by adding vertices and edges, to a Δ_i-regular digraph with $\Delta_i = \max(\Delta_i^+, \Delta_i^-)$, which has the same directed girth g_i. Since $g_i > 8e\Delta_i$ (for all sufficiently large d) we have that by Theorem 1.6 that, for each $i \neq p$

$$\mathrm{dla}(G_i) \leq \Delta_i + 1 \leq \frac{d}{p} + 3\sqrt{\frac{d}{p}\log(d)} + 1.$$

To bound the size of G_p we note that, G_p can also be completed to a Δ_p-regular digraph, which can then be partitioned into Δ_p disjoint 1-regular spanning subgraphs as before using Hall's theorem. By splitting each of these into two matchings we see that

$$\mathrm{dla}(G_p) \leq 2\Delta_p \leq 2\frac{d}{p} + 6\sqrt{\frac{d}{p}\log(d)}.$$

These last two inequalities together with the fact that $10\sqrt{d} \leq p \leq 20\sqrt{d}$ imply that

$$\begin{aligned}
\mathrm{dla}(G) &\leq (p-1)\left(\frac{d}{p} + 3\sqrt{\frac{d}{p}\log(d)} + 1\right) + 2\frac{d}{p} + 6\sqrt{\frac{d}{p}\log(d)} \\
&= d + \frac{d}{p} + (p-1) + 3(p-1)\sqrt{\frac{d}{p}\log(d)} + 6\sqrt{\frac{d}{p}\log(d)} \\
&\leq d + cp\sqrt{\frac{d}{p}\log(d)} \\
&\leq d + cd^{\frac{3}{4}}(\log(d))^{\frac{1}{2}}
\end{aligned}$$

□

Since any $2d$-regular graph G can be oriented so the resulting digraph is d-regular (and since every $(2d-1)$-regular G is a subgraph of a $2d$-regular graph), we have as an immediate corollary of Theorem 1.8.

Corollary 1.9. *There exists a constant $c > 0$ such that for every d-regular graph G*

$$la(G) \leq \frac{d}{2} + cd^{\frac{3}{4}}(\log(d))^{\frac{1}{2}}.$$

1.3 Acyclic edge-chromatic number

The Local Lemma has been quite successful in applications to graph colouring. Indeed, often in colouring problems we would like to take a random colouring, but avoid the 'local' events that particular substructures are coloured 'incorrectly'. For example, if we wish to find a proper colouring the bad events are adjacent vertices which receive the same colour. Structural properties of the graph can then be used to bound the dependencies between these bad events.

As an example let's consider the following notion of colouring for a graph G: We say an edge colouring $\chi : E(G) \to [k]$ is *proper* if no two adjacent edges receive the same colour and is *acyclic* if there is no 2-coloured cycle. The *acyclic edge-chromatic number* of G, denoted by $a(G)$ is the least number of colours in a proper acyclic edge-colouring of G.

Since every proper edge-colouring of G must use at least $\Delta(G)$ many colours there is a trivial bound that $a(G) \geq \Delta(G)$. However it was conjectured by Alon, Sudakov and Zaks that this bound is not far from the truth.

Conjecture 1.10. *For any graph G, $a(G) \leq \Delta(G) + 2$.*

Alon, Macdiarmid and Reed showed, using the local lemma, that $a(G) \leq 60\Delta(G)$ and whilst this constant has been improved, the best known bound is still $a(G) \leq 16\Delta(G)$ which follows from a closer analysis of their proof.

Theorem 1.11. *For any graph G, $a(G) \leq 100\Delta(G)$*

Proof. Let us define a random colouring $\chi : E(G) \to [100\Delta]$ by choosing the colour of each vertex uniformly and independently. We wish to show that with positive probability χ is a proper acyclic colouring. There are two types of bad events that we wish to avoid:

- The event A_B that a pair of adjacent edges B is monochromatic;
- The event A_C that a cycle C of length $2k$ is properly two-coloured.

Let us call the first a *type 1* event and the second a *type k* event. Clearly neither event is likely to happen: The first event happens with probability $\frac{1}{100\Delta}$ and the second happens with probability at most $2\binom{\Delta}{2}\left(\frac{1}{100\Delta}\right)^{2k} \leq \left(\frac{1}{100\Delta}\right)^{2k-2}$.

Furthermore, note that each edge e lies in at most Δ^{2k-2} different cycles of length $2k$. Indeed, this can been seen by considering the number possibilities for the tth edge f_t in the cycle as $t = 1, \ldots, 2k$. The first edge is fixed by our choice of $e = f_1$. We then have at most Δ many choices for f_2, as it must be adjacent to f_1 and then, having chosen f_2, at most Δ many choices for f_3 and so on. However, once we've chosen f_{2k-1} then f_k is fixed, since it must join f_{2k-1} and f_1. Hence we have at most Δ^{2k-2} many possible choices during this proces.

As before, there is a set of mutually independent variables, here given by the colour of each edge $\{\chi(e) \colon e \in E(G)\}$, such that each of our events is determined by some subset of these variables. We then form a dependency digraph by joining a pair of events if they depend on a shared variable.

By our observation in this digraph each Type 1 event is connected to at most 4Δ type 1 events and at most $2\Delta^{2k-2}$ type k events and each type ℓ event is connected to at most $4\ell\Delta$ type 1 events and at most $2\ell\Delta^{2k-2}$ type k events.

To choose our estimators x_B and x_C let us take $x_B = x_1 := \frac{1}{50\Delta}$ for each event of type 1 and $x_C = x_k := \left(\frac{1}{50\Delta}\right)^{2k-2}$ for each event of type k. Then for events of type 1

$$\begin{aligned}
x_i \prod_{(i,j)\in E} (1-x_j) &= x_1(1-x_1)^{4\Delta} \prod_k (1-x_k)^{2\Delta^{2k-1}} \\
&= \frac{1}{50\Delta}\left(1-\frac{1}{50\Delta}\right)^{4\Delta} \prod_k \left(1-\left(\frac{1}{50\Delta}\right)^{2k-2}\right)^{\Delta^{2k-2}} \\
&\geq \frac{1}{50\Delta} e^{-\frac{8}{50}} \prod_k e^{-2\left(\frac{1}{50}\right)^{2k-1}} \\
&= \frac{1}{50\Delta} e^{-\frac{8}{50}} e^{-2\sum_k \left(\frac{1}{50}\right)^{2k-2}} \\
&\geq \frac{1}{50\Delta} e^{-\frac{1}{5}} \\
&\geq \frac{1}{100\Delta} = \mathbb{P}(A_B),
\end{aligned}$$

since $e^{-\frac{1}{5}} \geq \frac{1}{2}$. Similarly for events of type ℓ

$$\begin{aligned}
x_i \prod_{(i,j)\in E} (1-x_j) &= x_\ell(1-x_1)^{4\ell\Delta} \prod_k (1-x_k)^{2\ell\Delta^{2k-2}} \\
&= \left(\frac{1}{50\Delta}\right)^{2\ell-2} \left(1-\frac{1}{50\Delta}\right)^{4\ell\Delta} \prod_k \left(1-\left(\frac{1}{50\Delta}\right)^{2k-2}\right)^{2\ell\Delta^{2k-2}} \\
&\geq \left(\frac{1}{50\Delta}\right)^{2\ell-2} e^{-\frac{8\ell}{50}} \prod_k e^{-4\ell\left(\frac{1}{50}\right)^{2k-2}} \\
&\geq \left(\frac{1}{50\Delta}\right)^{2\ell-2} e^{-\frac{8\ell}{50}} e^{-4\ell\sum_k \left(\frac{1}{50}\right)^{2k-1}} \\
&\geq \left(\frac{1}{100\Delta}\right)^{2\ell-2} 2^{2\ell-2} e^{-\frac{\ell}{5}} \\
&\geq \mathbb{P}(A_C) e^{\ell(2\log 2 - \frac{1}{5})-2} \\
&\geq \mathbb{P}(A_C),
\end{aligned}$$

since $\ell \geq 4$. Hence the conditions of Lemma 1.1 hold, and so we can conclude that with positive probability χ is a proper acyclic colouring, and hence $a(G) \leq 100\Delta$. □

Alon, Sudakov and Zaks were able to show that the conjecture holds for graphs of large girth.

Theorem 1.12. *If G is a graph with maximum degree Δ and girth at least $2000\Delta\log\Delta$ then $a(G) \leq \Delta + 2$.*

Proof. Let us denote by g the girth of G, so that $g \geq 2000\Delta\log\Delta$. By Vizing's Theorem there is at least one proper colouring $\chi : E(G) \to [\Delta+1]$ using at most $\Delta+1$ colours. However, this colouring might contain monochromatic cycles.

We will randomly recolour each edge independently with a new colour $\Delta+2$, with probability $\frac{1}{32\Delta}$, and we claim that with positive probability this results in a proper acyclic colouring χ'.

As before we have two types of bad events that we wish to avoid:

- The event A_B that a pair of adjacent edges B is monochromatic in χ';
- The event A_C that a cycle C of length $2k$ is two-coloured in χ'.

Again, each event is quite unlikely to happen. Indeed, since χ is proper a pair of adjacent edges B is monochromatic in χ' only if both are recoloured $\Delta+2$, which happens with probability $\frac{1}{32^2\Delta^2}$. Furthmore a cycle of length $2k$ can be two-coloured if either it was two coloured in χ, and no edge was recoloured, with probability at most $\left(1-\frac{1}{32\Delta}\right)^{2k}$, or if half the cycle was a single colour in χ and the other half were all recoloured, with probability at most $\left(\frac{1}{32\Delta}\right)^{k}$. This suggests that there are really three types of bad events we should consider. Let us say an even cycle D is *half-coloured* if it contains a monochromatic matching and let us denote by $H(D)$ the set of those edges (note that in a properly 2-coloured cycle there are two such sets). We consider the following events:

- The event A_B that a pair of adjacent edges B is monochromatic in χ';
- The event A'_C that a cycle C of length $2k$ is two-coloured in χ and no edge is recoloured;
- The event A'_D that a cycle D of length $2k$ is half-coloured in χ and every edge not in $H(D)$ is recoloured.

Let us call these events *type I, II*(k) and *III*(k).

Note that each edge is contained in at most 2Δ pairs of adjacent edges and is contained in at most Δ many 2-coloured cycles in χ, since χ is proper. Furthermore, we can bound the number of half-coloured cycles D of length $2k$ that each edge is contained it. Indeed, given an edge e, suppose first that $e \in H(D)$. In this case, let us imagine choosing the edges of our cycle sequentially starting at e. For each edge in $H(D)$ our choice will be fixed by the colouring χ, and for each edge not in $H(D)$ we will have at most Δ choices of which edge to choose. However, having chosen all of the other edges, the last edge of the cycle is then fixed by our choice of e. Hence, there are at most Δ^{k-1} such half-coloured cycles.

On the other hand, if $e \notin H(D)$, then we have at most Δ choices for the colour of the second edge. This fixes both neighbours of e. Then, as before choosing edges sequentially around the cycle, the edges in $H(D)$ will be fixed by χ, and we will have at most Δ choices for each edge not in $H(D)$. However, having chosen all of the other edges, the second to last edge of the cycle will be fixed by our choices so far, and so again there are at most Δ^{k-1} such half-coloured cycles.

Hence each edge e is in at most $2\Delta^{k-1}$ many half-coloured cycles D. So, as before we can choose our dependency digraph such that each event which depends on k edges is connected to at most $2k\Delta$ events of type I, at most $k\Delta$ events of type II in total and at most $k\Delta^{\ell-1}$ events of type III(ℓ).

It remains to choose appropriate estimators x_i for the probabilities of each event. Recall that

- $\mathbb{P}(A_B) = \frac{1}{1024\Delta^2}$ for each event A_B of type I;
- $\mathbb{P}(A'_C) = \left(1 - \frac{1}{32\Delta}\right)^{2k} \leq e^{-\frac{k}{16\Delta}}$ for each event A'_C of type II(k);
- $\mathbb{P}(A'_D) = \left(\frac{1}{32\Delta}\right)^k$ for each event A'_D of type III(k);

note that, since by assumption the girth of G is very large, $k \geq 1000\Delta \log \Delta$ and so $\mathbb{P}(A'_C) \leq \Delta^{-20}$ for every C.

For our choice of estimators x_i let us take

- $x_B = x_1 := \frac{1}{512\Delta^2}$ for each event A_B of type I;
- $x_C = x_2 := \frac{1}{128\Delta^2}$ for each event A'_C of type II;
- $x_D = x_k := \left(\frac{1}{2\Delta}\right)^k$ for each event A'_D of type III(k).

It remains to check that the bounds in the Local Lemma hold. Firstly, for and event A_B of type I we have

$$\begin{aligned}
&x_1(1-x_1)^{4\Delta}(1-x_2)^{2\Delta}\prod_k (1-x_k)^{4\Delta^{k-1}} \\
&= \frac{1}{512\Delta^2}\left(1-\frac{1}{512\Delta^2}\right)^{4\Delta}\left(1-\frac{1}{128\Delta^2}\right)^{2\Delta}\prod_k\left(1-\left(\frac{1}{2\Delta}\right)^k\right)^{4\Delta^{k-1}} \\
&\geq \frac{1}{512\Delta^2} e^{-\frac{1}{64\Delta}} e^{-\frac{1}{32\Delta}} e^{-\frac{8}{\Delta}\sum_k 2^{-k}} \\
&\geq \frac{1}{512\Delta^2} e^{-\frac{1}{64\Delta}} e^{-\frac{1}{32\Delta}} e^{-\frac{8}{\Delta}2^{-2000\Delta\log\Delta}} \\
&\geq \mathbb{P}(A_B)
\end{aligned}$$

since $k \geq g \geq 2000\Delta \log \Delta$ and $\Delta > 2$.

For an event A'_C of type II(ℓ) we have

$$\begin{aligned}
&x_2(1-x_1)^{4\ell\Delta}(1-x_2)^{2\ell\Delta}\prod_k (1-x_k)^{4\ell\Delta^{k-1}} \\
&= \frac{1}{128\Delta^2}\left(1-\frac{1}{512\Delta^2}\right)^{4\ell\Delta}\left(1-\frac{1}{128\Delta^2}\right)^{2\ell\Delta}\prod_k\left(1-\left(\frac{1}{2\Delta}\right)^k\right)^{4\ell\Delta^{k-1}} \\
&\geq \frac{1}{128\Delta^2} e^{-\frac{\ell}{64\Delta}} e^{-\frac{\ell}{32\Delta}} e^{-\frac{8\ell}{\Delta}\sum_k 2^{-k}} \\
&\geq \frac{1}{128\Delta^2} e^{-\frac{\ell}{21\Delta}} \\
&\geq e^{-\frac{\ell}{16\Delta}} e^{\frac{\ell}{100\Delta}} \frac{1}{128\Delta^2} \\
&\geq e^{-\frac{\ell}{16\Delta}} \Delta^{10} \frac{1}{128\Delta^2} \\
&\geq e^{-\frac{\ell}{16\Delta}} \geq \mathbb{P}(A'_C),
\end{aligned}$$

again since $k, \ell \geq g \geq 2000\Delta \log \Delta$ and $\Delta > 2$.

Finally for an event A'_D of type III(ℓ) we have

$$\begin{aligned}
&x_\ell (1-x_1)^{4\ell\Delta}(1-x_2)^{2\ell\Delta}\prod_k (1-x_k)^{4\ell\Delta^{k-1}} \\
&= \left(\frac{1}{2\Delta}\right)^\ell \left(1-\frac{1}{512\Delta^2}\right)^{4\ell\Delta}\left(1-\frac{1}{128\Delta^2}\right)^{2\ell\Delta}\prod_k \left(1-\left(\frac{1}{2\Delta}\right)^k\right)^{4\ell\Delta^{k-1}} \\
&\geq \left(\frac{1}{32\Delta}\right)^\ell 2^{4\ell} e^{-\frac{\ell}{21\Delta}} \\
&\geq \left(\frac{1}{32\Delta}\right)^\ell e^{\ell(4\log 2 - \frac{1}{21\Delta})} \\
&\geq \mathbb{P}(A'_D).
\end{aligned}$$

where in the third line we have used that this is the same quantity which appears in the previous computation. □

2 Dependent random choice

Recently a technique which is based on a simple application of the alteration method has been used in various contexts, normally to do with embedding sparse graphs. In this chapter we give a short overview of the method, which is known as dependent random choice, and a few examples of applications.

The basic idea can be summarised as follows: We would like to find, in a dense graph (that is a graph with large minimum/average degree), a set of vertices U such that every small subset of U has many common neighbours. To do this, we first pick a small set of vertices T at random from the graph and let U' be the set of common neighbours of T. Intuitively, if we have some subset of G with not many common neighbours, then it is unlikely that all the members of T will lie in this set of common neighbours, and hence it is unlikely to be a subset of U'. Therefore the expected number of 'bad' subsets in U' will be small and so by removing a small number of vertices, one from each 'bad' set, we should find a set U with the desired properties.

Lemma 2.1. *Let G be a graph with $|G| = n$ and let $d = 2|E(G)|/n$ be the average degree of G. If there exist positive integers t, a, m, r 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 subset $R \subset U$ of size $|R| = r$ has at least m common neighbours.

Proof. For any set $X \subset V$ let $\Gamma(X) = \{v \in V : v \in N(x) \text{ for all } x \in X\}$ be the set of common neighbours of X. We pick a set of vertices T uniformly at random from V^t, that is with repetition. Let $A = \Gamma(T)$ be the set of common neighbours of T, and let X be the random variable with counts the size of A. for any vertex v, the probability that $v \in A$ is the probability that every element of T is a neighbour of v and so, by linearity of expectation

$$\mathbb{E}(X) = \sum_{v \in V}\left(\frac{|N(v)|}{n}\right)^t = n^{-t}\sum_{v \in V}|N(v)|^t.$$

Since the function $f(x) = x^t$ is convex on $[0, \infty)$, we can use Jensen's inequality to say

$$\mathbb{E}(X) \geq n^{-t}.n\left(\frac{\sum_{v \in V}|N(v)|}{n}\right)^t = n^{1-t}\left(\frac{2|E(G)|}{n}\right)^t = \frac{d^t}{n^{t-1}}.$$

Let Y be the random variable which counts the number of subset $R \subset A$ of size r with fewer than m common neighbours. For any $R \subset V$, let $\Gamma(R)$ be the set of common neighbours of R, then the probability that R is a subset of A is just

$$\left(\frac{|\Gamma(R)|}{n}\right)^t.$$

Therefore, if we let $\mathcal{R} = \{R \subset G : |R| = r \text{ and } |\Gamma(R)| < m\}$ be the set of 'bad' subset of V, we have that

$$\mathbb{E}(Y) = \sum_{R \in \mathcal{R}}\mathbb{P}(R \subset A) = \sum_{R \in \mathcal{R}}\left(\frac{|\Gamma(R)|}{n}\right)^t < \binom{n}{r}\left(\frac{m}{n}\right)^t.$$

Therefore we have that

$$\mathbb{E}(X-Y) \geq \frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq a.$$

Therefore there exists a choice of T for which $X - Y \geq a$. We delete one vertex from each subset R of $\Gamma(T)$ of size r with fewer than m common neighbours. Let U be the remaining subset of $\Gamma(T)$. We have that $|U| = X - Y \geq a$ and by construction every subset of U of size r has at least m common neighbours. □

Once we have a large set U such that every small subset has many common neighbours we can embed bipartite graphs in it in the following way

Lemma 2.2. *Let G be a graph, a, m, r be positive integers and suppose there exists a subset $U \subset V(G)$ of at least a vertices such that every subset $R \subset U$ of size r has at least m common neighbours.*

If H is a bipartite graph on vertex sets A and B such that $|V(H)| \leq m$, $|A| \leq a$ and every vertex in B has degree at most r, then H is a subgraph of G.

Proof. We wish to find an embedding of H in G given by an injective function $\phi : V(H) \to V(G)$. We start by picking an injective function $\phi : A \to U$ arbitrarily, which is possible since $|U| \geq a \geq |A|$.

We label the vertices of B as $v_1, v_2, \ldots, v_b$ and try to embed them in this order one at a time. Suppose we have already defined $\phi(v_i)$ for all $i < j$ and we wish to embed v_j. Let $N_j \subset A$ be the neighbourhood of v_j, so $|N_j| \leq r$. Since $\phi(N_j)$ is a subset of U of size at most r, there are at least m vertices in G adjacent to all the vertices in $\phi(N_j)$. Since the total number of vertices embedded already is less than $|V(H)| \leq m$, there is at least one vertex $w \in G$ which has not been used in the embedding and is adjacent to all the vertices in $\phi(N_j)$. We set $\phi(v_j) = w$.

After we have embedded every v_b it follows that ϕ is the desired embedding of H as a subgraph of G. □

2.1 Turán Numbers of Bipartite Graphs

The abundance of variables in Lemma 2.1 make it difficult to understand exactly what's going on, so let's look at an example of an application. For a graph H and an integer n, the Turán number $\mathrm{ex}(n, H)$ denotes the maximum number of edges in a graph on n vertices which does not contain H as a subgraph. Turán's theorem determines this number precisely for complete graphs $H = K_r$, and the asymptotic behaviour for graphs of chromatic number at least 3 is given by the well known result of Erdős and Stone

Theorem 2.3 (The Erdős-Stone Theorem)**.** *For any graph H with $\chi(H) \geq 3$*

$$ex(n, H) = \left(1 - \frac{1}{\chi(H)-1} + o(1)\right)\binom{n}{2}$$

For bipartite graphs the situation is much more complicated, and there are relatively few non-trivial bipartite H for which the order of magnitude of $\mathrm{ex}(n, H)$ is known. The following

result gives a bound for the Turán number of bipartite graphs in which one vertex class has bounded degree.

Theorem 2.4. *Let H be a bipartite graph on vertex sets A and B such that all vertices in B have degree at most r. Then there exists some constant $c = C(H)$ such that*

$$ex(n, H) \leq cn^{2-\frac{1}{r}}.$$

Proof. Let $a = |A|$ and $b = |B|$. The idea is, given a graph G with $|V(G)| = n$ and $e(G) \geq cn^{2-1/r}$, to use Lemma 2.1 to find a subset $U \subset V(G)$ of size at least a in which all the subsets of size r have at least $a + b$ common neighbours.

So let us check that the required bound holds in Lemma 2.1. We let $m = a + b$, $t = r$ and (for reasons which will become clear) let $c = \max\left(a^{1/r}, \frac{e(a+b)}{r}\right)$, note that c depends only on H. Given a graph G with $|V(G)| = n$ and $e(G) \geq cn^{2-1/r}$, the average degree of G satisfies $d \geq 2cn^{1-1/r}$. Therefore

$$\begin{aligned}\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t &\geq (2c)^r - \left(\frac{en}{r}\right)^r \left(\frac{a+b}{n}\right)^r \\ &\geq (2c)^r - \left(\frac{e(a+b)}{r}\right)^r \geq c^r \\ &\geq a.\end{aligned}$$

Therefore by Lemma 2.1 there exists a subset U of $V(G)$ of size at least a in which all the subsets of size r have at least $a + b$ common neighbours. Hence, by Lemma 2.2 H is a subgraph of G. □

These bounds are best possible in terms of their dependence on r. Indeed it is known the the Turán number of the complete bipartite graphs $K_{t,r}$ when $t \geq (r-1)!$ is $\Omega(n^{2-\frac{1}{t}})$.

2.2 The Ramsey Number of the Cube

Definition. The *Ramsey number* of an arbitrary graph H is

$$r(H) = \min\{n : \text{ Every 2 colouring of } K_n \text{ contains a monochromatic copy of } H\}.$$

The *r-dimensional Hypercube*, $\mathcal{Q}_r$, is a graph with vertex set $\{0,1\}^r$ where two vertices are adjacent if and only if they differ in exactly one coordinate.

An old conjecture of Burr and Erdős is that the Ramsey number of the cube is linear in the number of vertices, that is there exists some constant C such that $r(\mathcal{Q}_r) \leq C2^r$. Early bounds were much worse than this, for example Beck showed that $r(\mathcal{Q}_r) \leq 2^{Cr^2}$. More recently Shi obtained the first bound which was polynomial in the number of vertices, showing that $r(\mathcal{Q}_r) \leq 2^{Cr+o(r)}$ for some $C \sim 2.618$. Lemma 2.1 easily implies a, slightly worse, polynomial bound on $r(\mathcal{Q}_r)$.

Theorem 2.5.

$$r(\mathcal{Q}_r) \leq 2^{3r}$$

Proof. Let $n = 2^{3r}$. Given any two colouring of K_n, one of the colour classes contains at least half the edges. Let G be the graph of this colour.

Since $\mathcal{Q}_r$ is a bipartite graph, with vertex sets of size 2^{r-1} and maximum degree r, we would like to use Lemma 2.1 as before to find a set U of size at least 2^{r-1} such that every set of r vertices has at least 2^r common neighbours.

So let us set $a = 2^{r-1}$, $m = 2^r$ and $r = r$. We want to choose an appropriate t such that

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

where d is the average degree of G. Note that, since G has at least half of the edges of K_n we have that

$$d \geq 2\frac{e(G)}{n} \geq \frac{1}{2}(n-1) \geq 2^{-c}n,$$

for an appropriately chosen $c > 1$. Now

$$\begin{aligned}\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t &\geq 2^{-ct}n - \frac{n^r}{r!}\frac{m^t}{n^t} \\ &= 2^{-ct}n - \frac{n^{r-t}m^t}{r!} \\ &= 2^{3r-ct} - \frac{2^{3r^2-2rt}}{r!}.\end{aligned}$$

Setting $t = \frac{3}{2}r$ makes the second term negligible and so, to make the first term large we need $3r - c\frac{3}{2}r \geq r$, so for example we can take $c = 4/3$. All together this gives us

$$\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq 2^r - \frac{1}{r!} \geq 2^{r-1} = a.$$

Hence, as before, we can use Lemmas 2.1 and 2.2 to say that $\mathcal{Q}_r$ is a subgraph of G, that is, in any 2-colouring of K_n, the largest colour class will contain a subgraph isomorphic to $\mathcal{Q}_r$. Therefore $r(\mathcal{Q}_r) \geq n = 2^{3r}$. □

2.3 Improvements

Lemma 2.1 tells us that in any sufficiently dense graph on n vertices we can find a large set of vertices U such that every small subset has many common neighbours. For many applications it would be useful to have both the size of U and the number of common neighbours to be linear in n, for example if we wished to prove that the Ramsey number of the cube was linear in the number of vertices using the same method.

However one can construct graphs with average degree just less than $n/2$ such that any linear size subset of the vertices contains a small subset (in fact even a pair of vertices) with $o(n)$ common neighbours.

However using a similar proof based on alterations one can prove that in every dense graph there exists a subset U of linear size in which almost every small subset has linearly many common neighbours.

Lemma 2.6. *Let $\epsilon > 0$, $r \leq n$ be positive integers, and G a graph on $N > 4r\epsilon^{-r}n$ vertices with at least $\epsilon\frac{N^2}{2}$ edges. Then there is a subset $U \subset V(G)$ with $|U| > 2n$ such that number of subsets $S \subset U$ with $|S| = r$ and less than n common neighbours is at most*

$$\frac{1}{(2r)^r}\binom{|U|}{r}.$$

How might this be useful? Well if we think about that proof of Lemma 2.2 given a set U such that every small subset had many common neighbours, we embedded a bipartite graph H by arbitrarily embedding the left hand side in U, and that verifying that we can always extend that to an embedding of H, using the fact that when we want to embed a vertex v on the right hand side, the image of it's neighbourhood is a small set in U, and so has many common neighbours which are all candidates for the image of v.

If we were more careful in how we embedded the left hand side of H into U at the beginning, then if sufficiently few of the small sets in U don't have many common neighbours, we could try to embed the left hand side in such a way that none of these 'bad' small sets appear as neighbourhoods of things in the right hand side of H. We could then extend this to an embedding of H as before.

Obviously this will require some slightly stronger conditions on the graphs H we consider. The specific numbers in this lemma have been chosen so that an analogy of the embedding lemma (Lemma 2.2) carries over in this way for graphs with $\Delta(H) \leq r$. Using this one can improve on the previous bound to

Theorem 2.7.

$$r(\mathcal{Q}_r) \leq r2^{2r+3} \leq 2^{2r+o(r)}.$$

2.4 Embedding 1-subdivisions of general graphs

Given a graph H the 1-*subdivision* of H is the graph obtained by subdividing each edge of H exactly once. Note that if K is the 1-subdivision of H then K is a bipartite graph, one of whose partition classes can be associated with the vertices of H, and the other with the edges of H, and the degree of every vertex in the latter is two.

As we will see on the example sheet, it is relatively easy to use dependent random choice to show that in any dense graph G, with say εn^2 many edges, you can find a subdivision of the 1-subdivision of a complete graph on $f(\varepsilon)\sqrt{n}$ many vertices. In fact, by using a slight variation on the dependent random choice lemma, we will show that such a G will contain a 1-subdivision of *every* graph with at most $f(\epsilon)n$ many edges.

Theorem 2.8. *Let H be a graph with at most N edges and vertices, and let G be a graph with n vertices and εn^2 edges such that $n \geq 128\varepsilon^{-3}N$. Then G contains a 1-subdivision of H.*

The proof will involve a slightly more involved use of the dependent random choice methods. Given a graph H and a subdivision H' of H, let us call the vertices of H' coming from vertices of H *principle vertices* and those coming from the subdivided edges *subvidision vertices.* A naive strategy would be to try and find a set U of size at least N such that every pair of vertices has

at least N common neighbours. We could then embed the principle vertices greedily in U, and use the large shared neighbourhoods to embded the subdivision vertices.

However since N could be linearly large in n, we can't hope to find such a set U. However, we note that if the number of vertices is linearly large, then there must be many vertices of quite low degree, for which we have to be much less careful when embedding. So our idea will be to list the principle vertices from largest to smallest degree and deal with them in turn.

At the start these high degree vertices might have many neighbours already embedded, so we will want to make sure we're working in some set U such that every vertex in U has many common neighbours with a very high proportion of the other vertices in U (such a high proportion that if we take away the 'bad' vertices for each of the already embedded neighbours of the vertex we're considering, there is still a candidate to embed). By the end however, when the vertex we're embedding has quite low degree, we won't need the proportion of 'bad' vertices for any vertex in U to be as high to be able to guarantee that a candidate to embded exists.

So, rather than finding a set U all of whose small subsets have many common neighbours, we will find a sequence of nested sets $A_0 \supseteq A_1 \supseteq \dots$ such that A_i is sufficiently large and as i increases the number of pairs in A_i with small common neighbourhoods will drop quickly. Let us write $c(x,y)$ for the *codegree* of x and y, the number of common neighbours.

Lemma 2.9. *Let G be a graph with $n \geq 128\varepsilon^{-3}N$ and let V_1 be the set of vertices in G with degree at least $\frac{\varepsilon n}{2}$. Then there is a nested sequence of subsets of vertices $V_1 = A_0 \supseteq A_1 \supseteq \dots$ such that, for all $i \geq 0$, $|A_{i+1}| \geq \frac{\varepsilon}{8}|A_i|$ and each vertex in A_i has codgree at least N with all but at most $\left(\frac{\varepsilon}{8}\right)^i |A_i|$ vertices in A_i.*

Proof. Suppose we have already picked $A_0 \supseteq A_1 \supseteq \dots \supseteq A_{i-1}$ satisfying the properties and we wish to find A_i. Let us choose a vertex w uniformly at random from V, let $A = N(w) \cap A_{i-1}$ and let $X = |A|$.

Since every vertex in A_0 has degree at least $\frac{\varepsilon n}{2}$, so does every vertex in A_{i-1} and so

$$\mathbb{E}(X) = \sum_{v \in A_{i-1}} \frac{|N(v)|}{n} \geq \frac{\varepsilon}{2}|A_{i-1}|.$$

Let Y be the random variable counting the number of pairs $x, y \in A$ with $c(x,y) < N$. Note that, for any pair $x, y \in A_{i-1}$ the probability that $x, y \in A$ is $\frac{c(x,y)}{n}$. Let $c_i = \left(\frac{\varepsilon}{8}\right)^i$.

Let E_{i-1} be the set of pairs $\{x,y\}$ in A_{i-1} such that $c(x,y) < N$, so that by assumption $|E_i| \leq \frac{c_{i-1}}{2}|A_{i-1}|^2$. It follows that

$$\mathbb{E}(Y) < \frac{N}{n}|E_{i-1}| \leq \frac{N}{n}\frac{c_{i-1}}{2}|A_{i-1}|^2.$$

Let us consider the random variable $Z = X^2 - \frac{\mathbb{E}(X)^2}{2\mathbb{E}(Y)}Y - \frac{\mathbb{E}(X)^2}{2}$. By the convexity of the function x^2 and Jensen's inequality we have that $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$ and hence by linearity of expectation $\mathbb{E}(Z) \geq 0$.

Hence there is a choice of w such that this expression if non-negative. It follows that

$$|X|^2 \geq \frac{\mathbb{E}(X)^2}{2} \geq \frac{\varepsilon^2}{8}|A+i-1|^2,$$

and also, since $n \geq 128\varepsilon^{-3}N$,

$$Y \leq \frac{2\mathbb{E}(Y)}{\mathbb{E}(X)^2}X^2 \leq \frac{2N}{n}\frac{c_{i-1}}{2}|A_{i-1}|^2\frac{4}{\varepsilon^2|A_{i-1}|^2}X^2 \leq \frac{\varepsilon c_{i-1}}{32}X^2.$$

From the first inequality we have that $|A| = X \geq \frac{\varepsilon}{4}|A_{i-1}|$ and the second inequality guarantess that the number of pairs of vertices in A with codegree less than N is at most $\frac{\varepsilon c_{i-1}}{32}|A|^2$.

If A contains a vertex that has codegree $< N$ with more than $\frac{\varepsilon c_{i-1}}{16}|A|$ many other vertices of A, we delete it and continue this process until there are there are no vertices left which have codegree $< N$ with more than $\frac{\varepsilon c_{i-1}}{16}|A|$ remaining vertices.

During this process we delete at most

$$\left(\frac{\varepsilon c_{i-1}}{32}|A|^2\right) / \frac{\varepsilon c_{i-1}}{16}|A| = \frac{|A|}{2}$$

many vertices and hence if we let A_i be the remaining set of vertices then $|A_i| \geq \frac{|A|}{2} \geq \frac{\varepsilon}{8}|A_{i-1}|$, and every vertex in A_i has codegree at least N with all but at most

$$\frac{\varepsilon c_{i-1}}{16}|A| \leq \frac{\varepsilon}{8}c_{i-1}|A_i| = c_i|A_i|$$

vertices of A_i. Hence the claim follows by induction. □

Theorem 2.10. *Let G be a graph with $n \geq 128\varepsilon^{-3}N$ vertices and εn^2 edges and let H be a bipartite graph on partition classes A, B with at most N vertices and edges such that every vertex in B has degree 2. Then G contains H as a subgraph.*

Proof. If we let V_1 be the set of vertices in G with degree at least $\frac{\varepsilon n}{2}$ then $|V_1| > \varepsilon^{\frac{1}{2}}n$. Indeed, the number of edges of G not meeting a vertex in V_1 is at most $n\frac{\varepsilon n}{2} < \frac{\varepsilon n^2}{2}$ and hence the number of edges with both vertices in V_1 is at least $\frac{\varepsilon n^2}{2}$, but also at most $\binom{|V_1|}{2}$, from which the claim follows.

By the previous lemma we can find nested subsets $V_1 = A_0 \supseteq A_1 \supseteq \ldots$ such that, for all $i \geq 0$, $|A_{i+1}| \geq \frac{\varepsilon}{8}|A_i|$ and each vertex in A_i has codgree at least N with all but at most $\left(\frac{\varepsilon}{8}\right)^i|A_i|$ vertices in A_i.

Let H' be the graph with vertex set A where two vertices in A are adjacent if they have a common neighbour in B in H. If we can find an embedding $\phi : A \to V_1$ such that for every edge (a, b) of H', $\phi(a)$ and $\phi(b)$ have codegree at least N in G, then we can clearly extend ϕ to an embedding of H greedily.

So, let us enumerate $A = \{a_1, \ldots, a_{|A|}\}$ such that $d_{H'}(a_1) \geq d_{H'}(a_2) \geq \ldots \geq d_{H'}(a_{|A|})$. Since $e(H') \leq N$, it follows that the degree of each a_i is at most $\frac{2N}{i}$. We will construct our embedding ϕ in order, starting at a_1. Let us define as above $c_j = \left(\frac{\varepsilon}{8}\right)^j$. We will ensure that the vertex a_i is

embedded in A_j, where j is the least positive integer such that $c_j \leq \frac{i}{4N}$. Note that, in this case $c_{j-1} \geq \frac{i}{4N}$.

Since $n \geq 128\varepsilon^{-3}N$,

$$\begin{aligned} |A_j| &\geq c_j|A_0| \geq c_j\varepsilon^{\frac{1}{2}}n \\ &\geq \frac{\varepsilon}{8}c_{j-1}\varepsilon^{\frac{1}{2}}n \\ &\geq \frac{\varepsilon}{8}\frac{i}{4N}\varepsilon^{\frac{1}{2}}n \\ &\geq 2i \end{aligned}$$

Suppose that we have already embedded all the vertices a_k with $k < i$ and we wish to embed a_i. Let $N^-(a_i)$ be the set of vertices a_k with $k < i$ which are adjacent to a_i in H'. Each vertex in A_j has codegree at least N with all but at most $c_j|A_j| \leq \frac{i}{4N}|A_j|$ other vertices in A_j where j is chosen as above for i.

Since a_i has degree at most $\frac{2N}{i}$ in H', at least $|A_j| - \frac{2N}{i}\frac{i}{4N}|A_j| = \frac{|A_j|}{2}$ vertices of A_j have codegree at least N with every vertex in $\phi\left(N^-(a_i)\right)$. Since in particular $\frac{|A_j|}{2} \geq i$, there is a vertex in $A_j \setminus \phi(\{a_1, \ldots, a_{i-1})$ that has codegree at least N with every vertex in $\phi\left(N^-(a_i)\right)$. We let this vertex be $\phi(a_i)$ and continue. By induction we can find the desired embedding, finishing the proof. □

3 Random Subgraphs

Recently interest has been shown in the following model of a random subgraph: Suppose G is an arbitary graph, with minimum degree at least k, then for $p \in [0,1]$ we define G_p to be the random subgraph of G obtained by including each each of G independently and with probability p. When $G = K_{k+1}$, the complete graph on $k+1$ vertices, this gives the model $G(k+1,p)$. We are interested in the question of which properties of $G(k+1,p)$ also hold in G_p, when $p = p(k)$ is a function of the minimum degree of G.

For certain properties there can clearly be no transference, for example connectedness, since G_p itself might be disconnected. However, sometimes 'local' versions of these properties are more natural to consider. For example, in $G(k+1,p)$ being connected is equivalent to have a component of size at least $k+1$, and it is perhaps possible to find a $p(k)$ that guarantees this in G_p.

3.1 Large components, paths and cycles

For example, let's consider the behaviour of the random graph in the supercritical phase. We know that as p goes from $\frac{1-\varepsilon}{k}$ to $\frac{1+\varepsilon}{k}$, $G(k,p)$ goes from having only small components to having a giant component, whose size in linear in k. If we take G to be a union of many disjoint copies of K_{k+1}, then we see that we can't hope to ask that G_p has only small components in the subcritical phase, and similarly we can't ask for uniqueness of any large components in the supercritical phase. However, we can still ask if it's true that the size of the largest component is at least linear in k in the supercritical regime. Indeed, it is easy to see that actually most of the proofs for $G(k,p)$ will work in this setting also.

We will follow the depth first search proof, both for ease of exposition, and also since with only a little more work we can actually prove a slightly stronger result when $\varepsilon(k) \to 0$ sufficiently slowly.

Theorem 3.1. *Let $p = \frac{1+\varepsilon}{k}$ where $\varepsilon k = \omega\left(\left(\frac{\log k}{k}\right)^{\frac{1}{3}}\right)$, let G be a graph with $\delta(G) \geq k$. Then with high probability G_p contains a connected component with at least $\frac{\varepsilon k}{2}$ vertices.*

For the proof we will need the following simple consequences of the Chernoff bound.

Lemma 3.2. *Let $X_1, \ldots, X_N$ be a sequence of $N = \frac{\varepsilon k^2}{2}$ many i.i.d Ber(p) random variables with $p = \frac{1+\varepsilon}{k}$ then with high probability*

(a) $\sum_{i=1}^{k^{\frac{7}{4}}} X_i \leq k^{\frac{5}{6}}$;

(b) For every $k^{\frac{7}{4}} \leq t \leq N$

$$\left|\sum_{i=1}^{t} X_i - (1+\varepsilon)\frac{t}{k}\right| \leq k^{\frac{2}{3}}.$$

Proof. The former is clear, since $\mathbb{E}\left(\sum_{i=1}^{n^{\frac{7}{4}}} X_i\right) = (1+\varepsilon)k^{\frac{3}{4}} = o(k^{\frac{5}{6}})$ and for the latter we have, for every $k^{\frac{7}{4}} \leq t \leq N$ the probability that

$$\left|\sum_{i=1}^{t} X_i - (1+\varepsilon)\frac{t}{k}\right| \leq k^{\frac{2}{3}}.$$

is the probability that $\mathrm{Bin}(t,p)$ is more that $k^{\frac{2}{3}}$ from it's expectation tp, which is at most $\exp\left(-\frac{k^{\frac{4}{3}}}{2(tp+\frac{k^{\frac{2}{3}}}{3})}\right)$.

Since $k^{\frac{3}{4}} \leq tp \leq k$ this is at most $\exp\left(-\frac{k^{\frac{1}{3}}}{4}\right)$ and hence by a union bound with high probability the inequality holds for all $k^{\frac{7}{4}} \leq t \leq N$. □

Proof of Theorem 3.1. We run the depth first search algorithm on G_p starting at an arbitrary root. Using the principle of deferred decisions we can think of our edge-queries as coming from a sequence of i.i.d $\mathrm{Ber}(p)$ random variables $X_1, X_2, \ldots$. By Lemma 3.2 we can assume that properties (a) and (b) hold with high probability.

Our claim is that after that first $N = \frac{\varepsilon k^2}{2}$ many queries in the DFS algorithm we are in the midst of revealing a component whose size is at least $\frac{\varepsilon k}{2}$. Recall that the depth first search algorithm keeps track of three sets of vertices the stack A of *active* vertices, a set W of *visited* vertices and a set U of *unvisited* vertices, and that at each point during the algorithm we have queried every edge between U and W.

We claim that at time N the number of visisted vertives $|W|$ is at most $\frac{k}{3}$. Indeed, if not, there was some point $t \leq N$ when $|W| = \frac{k}{3}$, and at that point

$$|A| \leq 1 + \sum_{i=1}^{t} X_i$$

If $t \leq k^{\frac{7}{4}}$ then by (b) $|A| \leq 1 + k^{\frac{5}{6}} \leq \frac{k}{3}$ and if $t \geq k^{\frac{7}{4}}$ then by (a)

$$|A| \leq 1 + (1+\varepsilon)\frac{t}{k} + k^{\frac{2}{3}} \leq \frac{k}{3}.$$

It then follows that $|U| \geq \frac{k}{3}$ (since $A \cup U \cup W = V(G)$ and clearly $|V(G)| \geq \delta(G) = k$). Furthermore, since each vertex in W has degree at least k, and $|A \cup W| \leq \frac{2k}{3}$ it follows that each vertex in W has at least $\frac{k}{3}$ neighbours in U. However, then at this point we've already queried at least $\frac{k^2}{9} > N$ many edges between U and W, a contradiction.

Hence, at time N, $|W| < \frac{k}{3}$. If $|A| > \frac{k}{3}$ then we're done, since A always forms a path, and so we may assume that U is non-empty, and so the algorithm is still running. Hence, in particular, each positive query so far resulted in a vertex moving from U to A.

It follows that for every $k^{\frac{7}{4}} \leq t \leq N$

$$|A \cup W| \geq \sum_{i=1}^{t} X_i \geq (1+\varepsilon)\frac{t}{k} - k^{\frac{2}{3}}.$$

If at any point during this time period A is empty, W is large, and so the algorithm has already queried all the edges between W and U. Since every vertex in W has degree at least k, it has at least $k - |W|$ many neighbours in U and so we've already queried at least $|W|(k - |W|)$ many edges between U and W.

However, since $(1+\varepsilon)\frac{t}{k} - k^{\frac{2}{3}} \leq |W| \leq \frac{k}{3}$, the parabola $|W|(k-|W|)$ will take its minmum at the lower end of this range, and so we must have that

$$\begin{aligned} t \geq |W|(k-|W|) &\geq \left((1+\varepsilon)\frac{t}{k} - k^{\frac{2}{3}}\right)\left(k - (1+\varepsilon)\frac{t}{k} + k^{\frac{2}{3}}\right) \\ &\geq (1+\varepsilon)t - (1+\varepsilon)^2\frac{t^2}{k^2} - 2k^{\frac{5}{3}} \\ &\geq (1+\varepsilon)\left(1 - (1+\varepsilon)\frac{\varepsilon}{2}\right)t - 2k^{\frac{5}{3}} \\ &= (1+\varepsilon)\left(1 - \frac{\varepsilon}{2} - \frac{\varepsilon^2}{2}\right)t - 2k^{\frac{5}{3}} \\ &> t, \end{aligned}$$

since $t \geq k^{\frac{7}{4}}$, which is a contradiction. It follows that A is never empty for $k^{\frac{7}{4}} \leq t \leq N$ and hence every positive query results in a vertex in the same component of G_p. By our assumptions on the sequence $X_1, X_2, \ldots$ the number of positive queries in this range is at least

$$\begin{aligned} \sum_{i=k^{\frac{7}{4}}}^{N} X_i &\geq \sum_{i=1}^{N} X_i - \sum_{i=1}^{k^{\frac{7}{4}}} X_i \\ &\geq (1+\varepsilon)\frac{\varepsilon k}{2} - (1+\varepsilon)k^{\frac{3}{4}} - 2k^{\frac{2}{3}} \\ &\geq \frac{\varepsilon k}{2}. \end{aligned}$$

□

As we saw before, a very similar proof will show that there is even a path of linear length in $G(k,p)$ with high probability, and the same proof will go through almost verbatim for the random subgraph model. Furthermore, as $\varepsilon \to \infty$ these proofs also show that the length of the largest component/longest path will tend upwards to $(1 - o(1))k$.

In $G(k,p)$ it is then relatively easy to find a long cycle, once you have a long path, using sprinkling. However, this is not longer possible in G_p, indeed it could even be that the girth of G is much larger than k, so knowing that we have a path of length linear in k doesn't mean that we must have many edges in order to sprinkle.

Using similar ideas as above, Krivelevich, Lee and Sudakov showed that you could at least find a cycle of length at least $(\frac{1}{2} - o(1))k$, and then using some complicated arguments about the possible structure of the graph G, used this to show that G_p will indeed contain a cycle of length at least $(1 - o(1))k$ when $p = \omega\left(\frac{1}{k}\right)$.

However, Riordan managed to give a much simpler proof, which we present below.

Theorem 3.3. *Let G be a graph with $\delta(G) \geq k$ and let $p = \omega\left(\frac{1}{k}\right)$. Then with high probability G_p contains a cycle of length at least $(1 - o(1))k$.*

Proof. We explore the graph, as in the previous proof, using the DFS process. Let U, W and A be as before, and let us write T for the forest that we produce via this search process. We consider T as being a rooted forest, where each component is rooted at the first vertex of the component that was added to A. Let us denote by R the set of edges of G that are not queried in G_p during the process.

For any v in T there is a unique path from v to the root of it's component, which we will imagine to be drawn vertically. Let us write $\mathcal{A}(v)$ for the set of *ancestors* of v in T, that is, the set of vertices on this path. We write $\mathcal{D}(v)$ for the set of *descendants* of v in the tree, those w such that $v \in \mathcal{A}(w)$. Given an integer t we will write $\mathcal{A}_t(v)$ and $\mathcal{D}_t(v)$ for the set of ancestors/descendants of t at distance exactly t, and $\mathcal{A}_{\leq t}(v)$ and $\mathcal{D}_{\leq t}(v)$ for those at distance at most t. The *depth* of v is its distance from the root of its component and the *height* of v is the $\max\{t \colon \mathcal{D}_t(v) \neq \emptyset\}$.

Lemma 3.4. *Ever edge $e \in R$ joins two vertices on some vertical path in T.*

Proof of Lemma. Let $e = (u, v)$, and suppose that u is placed into W before v. When u is placed into W, v cannot be in U, else we would have queried the edge (u, v), and so v must be in A. Hence, at this point both v and u are on the stack, and so there is a vertical path from v to u. □

Lemma 3.5. *With high probability, at most $\frac{2n}{p} = o(kn)$ edges are queried during the DFS process.*

Proof of Lemma. At the end of the DFS process we have built a spanning forest T of G_p, which thus has at most n edges. Each time an edge is queried it succeeds with probability p, and so there are at most n successful queries during the process. However, the probability that more than $\frac{2n}{p}$ queries made, but fewer than n successful queries is $o(1)$. It follows that with high probability there are at most $\frac{2n}{p}$ edges queried. □

From this point on let us fix some small constant $\varepsilon > 0$, where we will assume $\varepsilon \leq \frac{1}{10}$. We say that a vertex v is *full* if it is incident with at least $(1 - \varepsilon)k$ edges in R.

Lemma 3.6. *With high probability, all but $o(n)$ vertices of T are full.*

Proof of Lemma. Since $\delta(G) \geq k$, each $v \in V(T)$ which is not full is incident with at least εk many queried edges. If there are at least δn many such vertices for any $\delta > 0$, then there are at least $\frac{\delta\varepsilon}{2}kn$ many queried edges. However, with high probability this doesn't happen, by Lemma 3.5. □

Let us call a vertex *rich* if $|\mathcal{D}(v)| \geq \varepsilon k$ and *poor* otherwise.

Lemma 3.7. *Suppose that T contains $o(n)$ poor vertices. Then for any constant $C > 0$, all but $o(n)$ vertices of T are at height at least Ck.*

Proof of Lemma. For each rich vertex v, let $P(v)$ be a set of exactly εk (ignoring floor/ceiling signs for ease of presentation) descendants of v, obtained by choosing vertices in $\mathcal{D}(v)$ at maximal

distance from v. Hence, for every $w \in P(v)$ we have $\mathcal{D}(w) \subseteq P(v)$, and so in particular $\mathcal{D}(w) < \varepsilon k$, and so w is poor.

Consider the set S_1 of ordered pairs (v, w) where v is rich and $w \in P(v)$. By assumption there are $(1 - o(1))n$ many rich vertices, and each have $|P(v)| = \varepsilon k$ and so $|S_1| \geq (1 - o(1))\varepsilon kn$. However, for any vertex w we have that $|\mathcal{A}_{\leq i}(w)| \leq i$, since every vertex has a unique ancestor at each fixed distance less than it's depth. Thus, if $(v, w) \in S_1$ then w is poor, and $v \in \mathcal{A}(w)$. Hence, since there are only $o(n)$ many poor vertices, the number of pairs $(v, w) \in S_1$ with $d(v, w) \leq Ck$ is $o(Ckn)$.

It follows that, $S_1' = \{(v, w) \in S_1 \colon d(v, w) > Ck\}$ has size at least $(1 - o(1))\varepsilon kn$. Since each vertex v is the first vertex in at most εk many pairs, it follows that there are at least $(1 - o(1))n$ vertices v appearing in some pair $(v, w) \in S_1'$, and so in particular each of these vertices is at height at least Ck. □

Let us call a vertex v *light* if $|\mathcal{D}_{\leq (1-5\varepsilon)k}(v)| \leq (1 - 4\varepsilon)k$, and *heavy* otherwise. Let H be the set of heavy vertices in T.

Lemma 3.8. *Suppose that T contains $o(n)$ poor vertices, and let $X \subseteq V(T)$ have size $|X| = o(n)$. Then, for k large enough, T contains a vertical path P of length at least $\varepsilon^{-2}k$ containing at most $\varepsilon^2 k$ vertices in $X \cup H$.*

Proof of Lemma. Let S_2 be the set of pairs (u, v) where $u \in \mathcal{A}(v)$ and $0 < d(u, v) < (1 - 5\varepsilon)k$. Since each v has at most one ancestor at each distance, $|S_2| \leq (1 - 5\varepsilon)kn$. On the other hand, by Lemma 3.7, all but $o(n)$ vertices u are at height at least k, and so appear in at least $(1 - 5\varepsilon)k$ pairs $(u, v) \in S_2$. It follows that only $o(n)$ many vertices u can be in significantly more pairs in S_2, for example in more than $(1 - 4\varepsilon)k$ many, However, since every $v \in \mathcal{D}_{\leq (1-5\varepsilon)k}(u)$ contributes to a pair $(u, v) \in S_2$, it follows that $H = o(n)$.

Let S_3 be the set of pairs (u, v) where $v \in X \cup H$, u is an ancestor of v and $d(u, v) \leq \varepsilon^{-2}k$. Since each v can appear in at most $\varepsilon^{-2}k$ pairs in S_3, we see that $|S_3| \leq \varepsilon^{-2}k|X \cup H| = o(kn)$. Hence, by double counting, only $o(n)$ vertices u appear in more than $\varepsilon^2 k$ pairs $(u, v) \in S_3$.

However, by Lemma 3.7, all but $o(n)$ vertices are at height at least $\varepsilon^{-2}k$. So there is some vertex u at height at least $\varepsilon^{-2}k$ which appears in at most $\varepsilon^2 k$ many pairs $(u, v) \in S_3$. Let P be the vertical path from u to some $v \in \mathcal{D}_{\varepsilon^{-2}k}(u)$. Then P has length $\varepsilon^{-2}k$ and every $v \in (X \cup H) \cap P$ appears in some pair $(u, v) \in S_3$, and so there are at most $\varepsilon^2 k$ many such v. Hence P satisfies the conclusion of the lemma. □

We are finally ready to conclude the proof of the theorem. Recall that we explored G_p via a DFS process and obtained a tree T and a set R of unqueried edges, and with high probability we may assume that these satisfy the conclusions of the previous lemmas. Note that the edges of R are still present in G_p independently with probability p.

Suppose that there is some vertex v such that

$$|\{u \colon (u, v) \in R, d(u, v) \geq (1 - 5\varepsilon)k\}| \geq \varepsilon k. \tag{3.1}$$

Then we can expose the edges (u, v) with $d(u, v) \geq (1 - 5\varepsilon)k$ in G_p, each of which is present with probability p. However, since $\varepsilon kp \to \infty$, it follows that with high probability one of the

exposures is successful, resulting in a cycle of length at least $(1-\varepsilon)k$. Hence we may assume that (3.1) fails for every vertex v. We note that in particular this implies that every full vertex is rich. Indeed, suppose that v is full but poor. Then v is adjacent to at least $(1-\varepsilon)k$ many edges in R, each of which goes between two vertices on a vertical path in T. However, since v has at most εk many descendants, at least $(1-2\varepsilon)k$ of these edges go to vertices $u \in \mathcal{A}(v)$. Since v has at most one ancestor at each distance, it follows that v satisfies (3.1), a contradiction.

Hence, since every full vertex is rich, and at most $o(n)$ vertices are not full, at most $o(n)$ vertices in T are poor. Let us apply Lemma 3.8 with X being the set of non-full vertices. We get a path P such that there are at most $\varepsilon^2 k$ vertices in P which are in $X \cup H$, that is, which are light or nor full. Let Z be the set of vertices in P which are full and light, so that $|V(P) \setminus Z| \leq \varepsilon^2 k$. Then, for any $v \in Z$, since v is full there are at least $(1-\varepsilon)k$ vertices $u \in \mathcal{A}(v) \cup \mathcal{D}(v)$ such that $(u,v) \in R$. Since v does not satisfy (3.1), at least $(1-2\varepsilon)k$ of these vertices are at distance at most $(1-5\varepsilon)k$ to v. Furthermore, since v is light, it has few descendants and so at least $2\varepsilon k$ of these vertices are in $\mathcal{A}(v)$.

Hence, since each vertex has at most one ancestor at each distance, we can find some set $R(v)$ of at least εk many vertices $u \in \mathcal{A}(v)$ such that $(u,v) \in R$ and $\varepsilon k \leq d(u,v) \leq (1-5\varepsilon)k \leq k$. We will use these sets $R(v)$ to find a long cycle.

Let us think of P as being oriented upwards towards the root, and let v_0 be the lowest vertex in $Z \subseteq P$. Since $|R(v_0)| \geq \varepsilon k$ and $\varepsilon kp \to \infty$, with high probability there is some edge (u_0, v_0) in G_p with $u_0 \in R(v_0)$. Let v_1 be the first vertex below u_0 on P with $v_1 \in Z$.

Note that $d(u_0, v_0) \geq \varepsilon k$ and $d(u_0, v_1) \leq 1 + |V(P) \setminus Z| \leq 2\varepsilon^2 k$, and so v_1 is above v_0 on P. We repeat the same process from v_1: find a $u_1 \in R(v_1)$ with $(u_1, v_1) \in G_p$ and let v_2 be the first vertex below u_1 on P. Since ε^{-1} is a fixed constant, we can continue doing this to find vertices $\{v_i, u_i \colon 1 \leq i \leq 2\varepsilon^{-1}\}$ such that, in the order $\preceq$ on P, $v_0 \preceq v_1 \preceq u_0 \preceq v_2 \preceq u_1 \preceq v_3 \dots$ with the overlapping chords $(u_i, v_i) \in G_p$ for each i. Note that, since $d(u_i, v_i) \leq k$, we remain within P since P has length at least $\varepsilon^{-2}k$.

However it is relatively simple to use these chords, together with P, to form a cycles of length at least $(1 - 2\varepsilon^{-1}2\varepsilon^2)k = (1-4\varepsilon)k$.

□

We note that Krivelevich and Samotij showed, with different methods, but still using the DFS process, that if $p = \frac{1+\varepsilon}{k}$ then whp G_p will contain a cycle whose length is linear in k.

3.2 Planarity

Similarly we could consider the planarity of a random subgraph. Recall that in $G(n,p)$ there is a sharp threshold for planarity at $\frac{1}{n}$; if $p = \frac{d}{n}$ for $d < 1$ then with high probability $G(n,p)$ is planar and if $p = \frac{d}{n}$ for $d > 1$ then with high probability $G(n,p)$ is non-planar.

Again, clearly we cannot hope for the latter to also hold in an arbitrary random subgraph. However if p is large enough then with high probability G_p will be non-planar; indeed by consid-

ering the average degree if $p > \frac{C}{k}$ for a large constant C then clearly G_p is with high probability non-planar. The fact that this is true for any $d > 1$ was first shown by Frieze and Krivelevich. We will give a slightly simplified version of their proof, due to Erde Kang and Krivelevich.

A rough sketch of the strategy is as follows. Let $p = \frac{1+\varepsilon}{k}$ and take $p_1 = \frac{1+\frac{\varepsilon}{2}}{k}$ and $p_2 = \frac{p-p_1}{1-p_1} \geq \frac{\varepsilon}{2k}$. We want to find a connected part of G_{p_1} which is relatively 'dense' in G, so that after sprinkling onto these edges with probability p_2 we will have with high probability a minor with large average degree.

The following lemma makes this more precise.

Lemma 3.9. *Let n, k be integers with $n \gg \sqrt{k}$ and $K, c_1, c_2 > 0$ be constants. Suppose T is a tree on n vertices with maximum degree at most $K-1$, F is a set of $c_1 kn$ many edges on the vertex set $V(T)$, and $p = \frac{c_2}{k}$. Then whp $T \cup F_p$ is non-planar.*

Proof. We first start by splitting T up into connected parts of size around $\sqrt{k}$. As long as n is sufficiently large compared to $\sqrt{k}$, this is relatively easy to do in a greedy fashion. Indeed, as long at $|T| \geq \sqrt{k}$ and $\Delta(T) \leq K-1$ there must be some vertex v such that the subtree T_v of T rooted at v satisfies $\sqrt{k} \leq |T_v| \leq (K-1)\sqrt{k}$, which can be seen by picking v to be the highest vertex with $|T_v| \geq \sqrt{k}$.

It follows that we can find connected disjoint vertex sets $A_1 \dots, A_r \subseteq V(T)$ such that

- $V(T) = \bigcup_{i=1}^{r} A_i$;
- $T[A_i]$ is connected for each i; and
- $\sqrt{k} \leq |A_i| \leq K\sqrt{k}$ for each i.

Indeed, we keep greedily choosing such a v and letting $A_i = T_v$ until the remaining tree has size $\leq \sqrt{k}$, and we add the rest to the last A_i.

Note that, since $V = \bigcup_{i=1}^{r} A_i$ and $|A_i| \geq \sqrt{k}$, $r \leq k^{-\frac{1}{2}} n$. Let F' be the set of edges in F which are not contained in any A_i. Then, since each A_i contains at most $\binom{|A_i|}{2} \leq \frac{K^2}{2} k$ edges inside it and $|F| \geq c_1 kn$, it follows that for large k,

$$|F'| \geq |F| - r\frac{K^2}{2}k \geq c_1 kn - \frac{K^2}{2}\sqrt{k}n \geq \frac{c_1}{2}kn.$$

Hence, on average each A_i meets at least $\frac{2|F'|}{r} \geq c_1 k^{\frac{3}{2}}$ many edges in F'.

We recursively delete sets A_i, and the edges in F' incident to them, which meet at most $\frac{c_1}{4}k^{\frac{3}{2}}$ edges remaining in F'; we must eventually stop this process before exhausting the A_i, since $r \leq k^{-\frac{1}{2}} n$ (i.e. there are at most $k^{-\frac{1}{2}} n$ many A_i) and

$$\frac{c_1}{4}k^{\frac{3}{2}}k^{-\frac{1}{2}}n = \frac{c_1}{4}kn \leq \frac{|F'|}{2}.$$

Hence there is some subfamily, without loss of generality, $\{A_1, \dots, A_\ell\}$ of the A_i, and some subset $F'' \subseteq F'$ of edges which lie between A_i and A_j with $i, j \in [\ell]$ such that at least $\frac{c_1}{4}k^{\frac{3}{2}}$ edges of F'' meet each A_i.

Note that $0 \leq e_{F''}(A_i, A_j) \leq K^2 k$ for each pair $i, j \in [\ell]$. For each pair $i, j \in [\ell]$ such that $e_{F''}(A_i, A_j) > k$ let us delete $e_{F''}(A_i, A_j) - k$ many edges in F'' which lie between A_i and A_j, and call the resulting set of edges $\hat{F}$. Then $0 \leq e_{\hat{F}}(A_i, A_j) \leq k$ for each $i, j \in [\ell]$ and furthermore each A_i still meets at least $\frac{c_1}{4K^2} k^{\frac{3}{2}}$ many edges of $\hat{F}$. Indeed, we deleted at most a $(1 - \frac{1}{K^2})$ proportion of the edges in F'' between each pair A_i and A_j, and hence at least a $\frac{1}{K^2}$ proportion of the edges meeting each A_i remains.

In particular we have

$$\sum_{i,j \in [\ell]} e_{\hat{F}}(A_i, A_j) \;\geq\; \ell \frac{c_1}{2K^2} k^{\frac{3}{2}}. \tag{3.2}$$

Let H be an auxilliary (random) graph on $[\ell]$ such that $i \sim j$ if and only if there is an edge between A_i and A_j in $\hat{F}_p$. The number of edges between A_i and A_j in $\hat{F}_p$ is distributed as $\mathrm{Bin}(e_{\hat{F}}(A_i, A_j), p)$. Note that if $np < 1/2$, then $\mathbb{P}\left(\mathrm{Bin}(n, p) \neq 0\right) = 1 - (1-p)^n \geq \frac{np}{2}$. Since $e_{\hat{F}}(A_i, A_j) \leq k$ and $p = \frac{c_2}{k}$, and without loss of generality we may assume that $c_2 < \frac{1}{2}$, it follows that

$$\mathbb{P}(i \sim j) \geq \frac{c_2 e_{\hat{F}}(A_i, A_j)}{2k}. \tag{3.3}$$

By (3.2) and (3.3), we have

$$\mathbb{E}(e(H)) = \frac{1}{2} \sum_{i,j \in [\ell]} \mathbb{P}(i \sim j) \geq \frac{1}{2} \sum_{i,j \in [\ell]} \frac{c_2 e_{\hat{F}}(A_i, A_j)}{2k} \geq \frac{1}{4k} \ell \frac{c_1 c_2}{2K^2} k^{\frac{3}{2}} = \frac{c_1 c_2}{8K^2} \ell k^{\frac{1}{2}}.$$

And so we expect H to have average degree $\Omega\left(k^{\frac{1}{2}}\right)$. It remains to show that $e(H)$ is well concentrated about its mean.

However, $e(H)$ is the sum of independent Bernoulli random variables, and so $\mathrm{Var}(e(H)) \leq \mathbb{E}(e(H))$. Hence, by Chebyshev's inequality

$$\mathbb{P}\left(e(H) \leq \frac{\mathbb{E}(e(H))}{2}\right) \leq \mathbb{P}\left(|e(H) - \mathbb{E}(e(H))| \geq \frac{\mathbb{E}(e(H))}{2}\right) \leq \frac{4\mathrm{Var}(e(H))}{\mathbb{E}(e(H))^2} \leq \frac{4}{\mathbb{E}(e(H))} = o(1).$$

Hence, with high probability $e(H) = \Omega\left(\ell k^{\frac{1}{2}}\right)$ and so H has average degree $\omega(1)$. It follows that H is non-planar. Finally, observe that by contracting each A_i the graph H becomes a minor of $T \cup F_p$, and so the result follows. □

We note that the conclusion of this lemma is much stronger than non-planarity, we found a minor with average degree $\Omega(\sqrt{k})$.

In order to find such a tree we will start a (restricted) breadth-first search process in the graph G_{p_1}. If every vertex always has $\approx k$ neighbours (in G) outside the current tree, then at at each stage the *frontier* (the active vertices in tree) should grow in size by at least $(1 + \varepsilon)$, and so always be at least some constant fraction of the whole tree.

However, if a large proportion of the vertices in the frontier do not have $\approx k$ neighbours outside the current tree, then they must all have $\Theta(k)$ many neighbours inside the tree. This would then give us sufficiently many edges in the tree to apply Lemma 3.9.

Unfortunately there are a few technical details with following this strategy. Firstly, we need to restrict our breath-first search process so that the tree we build has bounded maximum degree. Secondly, we need to make sure that during the process of exposing the edges incident to the frontier we don't add too many neighbours of the frontier to the tree, since then our sketch above wouldn't work. Finally, in order to make sure this all happens with high probability we need to first grow the tree 'by hand' for the first few stages, so that the tree is already large when we start, this will necessitate throwing away some failed attempts, whose vertices we will have to keep track of.

So, firstly let us give a simple bound on the expectation of a restricted binomial random variable which will be useful later.

Lemma 3.10. *Let $X \sim Bin(n,p)$ be a binomial random variable with $2enp < K$ for some constant $K > 0$. If $Y = \min\{X, K\}$, then*

$$\mathbb{E}(Y) \geq np - K2^{-K}.$$

Proof. For every $t \leq K$ we have that $\mathbb{P}(Y = t) \geq \mathbb{P}(X = t)$. Hence, by standard estimates

$$\begin{aligned}
\mathbb{E}(X) - \mathbb{E}(Y) &\leq \sum_{t>K} t\binom{n}{t} p^t (1-p)^{n-t} \\
&\leq \sum_{t>K} t \left(\frac{enp}{t}\right)^t \\
&\leq \sum_{t>K} enp \left(\frac{enp}{t}\right)^{t-1} \\
&\leq \sum_{t>K} \frac{K}{2} \left(\frac{enp}{K}\right)^{t-1} \\
&\leq \frac{K}{2} \left(\frac{enp}{K}\right)^{K-1} \\
&\leq K2^{-K},
\end{aligned}$$

since $\frac{enp}{K} < \frac{1}{2}$. □

Theorem 3.11. *Let ε be a positive constant, G be a graph with $\delta(G) \geq k$, and $p = \frac{1+\varepsilon}{k}$. Then whp G_p is non-planar.*

Proof. Our plan will be to sprinkle with $p_1 = \frac{1+\frac{\varepsilon}{2}}{k}$ and $p_2 = \frac{p-p_1}{1-p_1} \geq \frac{\varepsilon}{2k}$.

Initial Phase : We first run an initial phase in which we build a partial binary tree T_0 of size $\log\log\log k =: N$ in G_{p_1}. By a partial binary tree we mean a tree in which all vertices have degree three or one, and in which there is a leaf r such that there is some integer L such that every other leaf is at distance L or $L-1$ from r.

We will do so via a sequence of trials. In a general stage we will have a set of *discarded vertices* X which will have size $o(\log k)$, and a partial binary tree T' of size $\leq N$, such that so far we have only exposed edges in G_{p_1} which meet either X or a non-leaf vertex of T'.

We choose a leaf $v \in V(T')$ of minimal distance to the root and expose the edges between v and $V \setminus (X \cup V(T'))$ in G_{p_1}. If v has at least two neighbours, we choose two of them arbitrarily (or one if v is the root) and add them to T' as children of v. Otherwise we say that the trial *fails* and we add $V(T')$ to X and choose a new root v arbitrarily from $V \setminus X$ and set $T' = v$. If at any point $|T'| = N$ we set $T_0 := T'$ and we finish the initial phase.

Since each v has at least $k - |X \cup V(T')| \geq (1-\varepsilon)k$ many neighbours in $V \setminus (X \cup V(T'))$, the probability that a trial fails is at most

$$\begin{aligned}\mathbb{P}\left(\mathrm{Bin}\left((1-\varepsilon)k, p_1\right) < 2\right) &= (1-p_1)^{(1-\varepsilon)k} + (1-\varepsilon)kp_1(1-p_1)^{(1-\varepsilon)k-1} \\ &\leq \left(1 - p_1 + (1-\varepsilon)\left(1+\frac{\varepsilon}{2}\right)\right)\exp\left(-(1+\frac{\varepsilon}{2})(1-\varepsilon) + p_1\right) \\ &\leq 2e^{\varepsilon - 1} := 1 - \gamma < 1.\end{aligned}$$

Hence, each time we choose a new root the probability that we build a suitable T_0 before a trial fails is at least

$$\gamma^N.$$

Therefore, whp we build such a tree before we've chosen $\gamma^{-N}N$ new roots. Since we only ever discard at most N vertices, during this process the number of discarded vertices is at most

$$\gamma^{-N}N^2 = (\log\log k)^{-\log\gamma}\left(\log\log\log k\right)^2 = o\left(\log k\right).$$

Let S_0 be the set of leaves of T_0 and note that, since T_0 is a partial binary tree as defined above, $|S_0| \geq \frac{1}{4}|T_0|$. Furthermore, during this process we have only exposed edges which are incident to either a vertex in X or a vertex in $V(T_0) \setminus S_0$. In particular, we have not exposed any edges between S_0 and $V \setminus (X \cup V(T_0))$.

Tree Branching Phase : Suppose then that in a general step we have a tree T_t together with a set S_t of leaves of T_t, called the *frontier* of T_t, with the following properties:

(a) $|S_t| \geq \frac{\varepsilon}{16}|T_t|$;

(b) No edges from S_t to $V \setminus (X \cup V(T_t))$ have been exposed in G_{p_1};

(c) The maximum degree in T is at most $K+1$,

where

$$K := 4\log\frac{1}{\varepsilon}$$

is a large constant. Note that T_0 and S_0 satisfy these three properties.

Let $0 < \delta \ll \varepsilon$ and let us consider the set

$$V_0 = \{s \in S_t \colon e_G(s, T_t) \geq \delta k\}.$$

If $|V_0| \geq \delta|S_t|$, then $G[V(T_t)]$ contains a set F of at least $\frac{\delta^2}{2}|S_t|k \geq \frac{\delta^2\varepsilon}{32}|T_t|k$ many edges. In particular, note that this implies that $|T_t| = \Omega(k)$.

Since T_t has bounded degree, by Lemma 3.9 when we sprinkle onto the edges of F with probability p_2, whp we find a non-planar subgraph of G_p.

So, we may assume that $|V_0| \leq \delta|S_t|$. Let $V_1 = S_t \setminus V_0$. Since $|X| = o(k)$, every vertex $s \in V_1$ has degree at least $(1-2\delta)k$ to $V \setminus (X \cup V(T_t))$. Let us arbitrarily order the set $V_1 = \{s_1, \dots, s_r\}$.

We will build the new frontier S_{t+1} by exposing the neighbourhood of each s_i in turn. At the start of the process each s_i has at least $(1-2\delta)k$ many possible neighbours, however, as S_{t+1} grows, it may be that some s_i have a significant fraction of their neighbours inside S_{t+1}.

Let us initially set $S_{t+1}(0) = \emptyset$ and $B(0) = \emptyset$. We will show that whp we can either find a large complete minor, or construct, for each $1 \leq j \leq r$, sets $S_{t+1}(j)$ and $B(j)$, and a forest $F(j)$, such that:

1. $B(j) \subseteq \{s_i \colon i \in [j]\}$ and $|B(j)| < \delta|S_t|$;
2. Each $s \in B(j)$ has $e_G(s, S_{t+1}(j)) \geq \delta k$;
3. There is a forest $F(j)$ of maximum degree K in G_{p_1} which joins each $v \in S_{t+1}(j)$ to some $s \in \{s_i \colon i \in [j]\}$.

Clearly this is satisfied with $j = 0$. Suppose we have constructed appropriate $S_{t+1}(j-1)$ and $B(j-1)$.

If $d_G(s_j, S_{t+1}(j-1)) \geq \delta k$ then we let $B(j) = B(j-1) \cup s_j$, $S_{t+1}(j) = S_{t+1}(j-1)$ and $F(j) = F(j-1)$. If $|B(j)| \geq \delta|S_t|$ then we can apply Lemma 3.9 to the edges spanned by $V(T_t \cup F(j))$, those include the edges in $E_G(B(j), S_{t+1}(j))$.

Then, $|T_t \cup F(j)| \leq |T_t| + K|S_t| = \Theta(|T_t|)$ and

$$\big|E\big(G[V(T_t \cup F(j))]\big)\big| \geq e_G(B(j), S_{t+1}(j)) \geq \delta^2|S_t|k = \Theta(|T_t|k).$$

Hence, by Lemma 3.9 after sprinkling onto $G[V(T_t \cup F(j))]$ with probability p_2 whp we have a complete minor of order $\Omega\left(\sqrt{\frac{k}{\log k}}\right)$.

Therefore, we may assume that $|B(j)| < \delta|S_t|$ and so conditions (1)–(3) are satisfied by $B(j)$, $S_{t+1}(j)$ and $F(j)$.

So, we may assume that $d_G(s_j, S_{t+1}(j-1)) \leq \delta k$, and hence s_j has at least $(1-3\delta)k$ neighbours in $V \setminus (V(T_t) \cup S_{t+1}(j-1))$. We expose the neighbourhood $N(j)$ of s_j in $V \setminus (V(T_t) \cup S_{t+1}(j-1))$ in G_{p_1}. Let us choose an arbitrary subset $N'(j) \subseteq N(j)$ of size $\min\{N(j), K\}$ and let $F'(j)$ be the set of edges from s_j to $N'(j)$. We set $B(j) = B(j-1)$, $S_{t+1}(j) = S_{t+1}(j-1) \cup N'(j)$ and $F(j) = F(j-1) \cup F'(j)$. It is clear that these now satisfy (1)–(3).

Hence we may assume that we have constructed $S_{t+1}(r)$, $B(r)$, and $F(r)$. Let us set $S_{t+1} = S_{t+1}(r)$ and $T_{t+1} = T_t \cup F(r)$. Note that S_{t+1} is the frontier of T_{t+1}, and so property (b) is satisfied. Furthermore, since $F(r)$ has maximum degree K, so is property (c).

Finally, we note that, since $|B(r)| < \delta|S_t|$, we exposed the neighbourhood $N(j)$ of at least $(1-2\delta)|S_t|$ of the vertices in S_t. Furthermore, the size of the union of their neighbourhoods is

stochastically dominated by a sum of restricted binomial random variables. More precisely, if we let

$$Y \sim \min\left\{\mathrm{Bin}\left((1-3\delta)k, p_1\right), K\right\},$$

then the sizes of the neighbourhoods $(N'(i)\colon i \notin B(r))$ are stochastically dominated by a sequence of $r - |B(r)|$ many mutually independent copies of Y, $(Y_i\colon i \notin B(r))$. Hence, if we let $Z = \sum_{i \notin B(r)} Y_i$ then $|S_{t+1}|$ is stochastically dominated by Z.

Note that

$$1 + \frac{\varepsilon}{3} \leq (1-3\delta)kp_1 = (1-3\delta)\left(1 + \frac{\varepsilon}{2}\right) \leq 2.$$

Hence, since $K = 4\log\frac{1}{\varepsilon} \geq 2e(1-3\delta)kp_1$, Lemma 3.10 implies that

$$\begin{aligned}
\mathbb{E}(Y) &\geq \left(1 + \frac{\varepsilon}{3}\right) - K2^{-K} \\
&\geq \left(1 + \frac{\varepsilon}{3}\right) - Ke^{-\frac{K}{2}} \\
&= \left(1 + \frac{\varepsilon}{3}\right) - 4\log\left(\frac{1}{\varepsilon}\right)\varepsilon^2 \\
&\geq 1 + \frac{\varepsilon}{4},
\end{aligned}$$

as long as ε is sufficiently small.

Since $r - |B(r)| \geq (1-2\delta)|S_t|$, it follows that $\mathbb{E}(Z) \geq (1-2\delta)|S_t|\mathbb{E}(Y) \geq (1 + \frac{\varepsilon}{5})|S_t|$. We can then bound the probability that Z deviates from it's mean using for example the Azuma-Hoeffding inequality.

Indeed, since $Z = \sum_{i \notin B(r)} Y_i$, we can consider the exposure martingale of Z with respect to the sequence $(Y_i\colon i \notin B(r))$. Since the Y_i are independent, and take values in $[0, K]$ it follows that this martingale satisfies the bounded differences condition with this parameter K and hence an application of the Azuma-Hoeffding inequality gives

$$\begin{aligned}
\mathbb{P}\left(|S_{t+1}| < \left(1 + \frac{\varepsilon}{8}\right)|S_t|\right) &\leq \mathbb{P}\left(Z < \left(1 + \frac{\varepsilon}{8}\right)|S_t|\right) \\
&\leq \mathbb{P}\left(|Z - \mathbb{E}(Z)| > \frac{\varepsilon}{20}|S_t|\right) \\
&\leq 2\exp\left(-\frac{\varepsilon^2|S_t|^2}{400(r - |B(r)|)K^2}\right) \\
&= e^{-\Omega(|S_t|)}, \qquad (3.4)
\end{aligned}$$

since $r \leq |S_t|$. It follows that with probability at least $1 - e^{-\Omega(|S_t|)}$, $|S_{t+1}| \geq (1 + \frac{\varepsilon}{8})|S_t|$, and it is then a simple check that $|S_{t+1}| \geq \frac{\varepsilon}{16}|T_{t+1}|$ and hence property (a) is also satisfied.

Hence, we have shown that in the tth step we can either find a non-planar subgraph, or with probability at least $1 - e^{-\Omega(|S_t|)}$ we can continue our tree growth. However, since G is finite the tree growth cannot continue forever, and so, unless the tree growth fails at some step, we must eventually find a non-planar subgraph.

Recall that the probability of failure is $o(1)$ in the initial phase, and by (3.4) the probability that the tree growth fails at some step is at most

$$\sum_t e^{-\Omega(|S_t|)} = o(1),$$

since $|S_0| \geq \frac{1}{4} \log\log\log k$ and $|S_t| \geq (1 + \frac{\varepsilon}{8})|S_{t-1}|$. Hence the total probability of failure is $o(1)$, and so whp G_p is non-planar. □

As with the comment after Lemma 3.9 we actually get the stronger conclusion that with high probability G_p contains a minor with average degree $\sqrt{k}$. It follows from a well-known result of Kostochka, and Thomason, that with high probability G_p contains a complete minor of order $\Omega\left(\frac{\sqrt{k}}{\log k}\right)$ which is almost optimal (up to the polylogarithmic factor).

4 Entropy Methods

4.1 Basic Results

Given a discrete random variable X let us denote by $p(x) := \mathbb{P}(X = x)$ for each x in the range of X. We define the *entropy* of the random variable X to be

$$H(X) = \sum_x p(x) \log\left(\frac{1}{p(x)}\right).$$

Note that this quantity is always positive.

We want to think of entropy, at least heuristically, as a measure of the expected amount of 'surprise' we have upon discovering the value of X. We then have the following heuristic argument for why $H(X)$ should be defined as above.

If we have an event A, such as the event $X = x$ for some x, the amount of 'surprise' we have at the event A happening should just be some function$f(p)$ of $p := \mathbb{P}(A)$. There are a number of reasonable conditions we should expect f to satisfy:

- $f(1) = 0$, since a certain event is no surprise;
- f should be decreasing, since rarer events are more surprising;
- f is continuous;
- $f(pq) = f(p)+f(q)$, which can be motivated by considering independent events happening with probability p and q;
- finally, for normalisation we may as well assume $f(1/2) = 1$.

It turns out that $f(p) = \frac{1}{\log p}$ is the unique function satisfying these constraints. Then, $H(X)$ is the expected value, taken over the range of X, of the surprise of the event that X takes a certain value, and so $H(X)$ is the only 'reasonable' function representing the idea following these heuristics.

As an example, consider $X \sim \mathrm{Ber}(p)$, then

$$H(X) = p \log\left(\frac{1}{p}\right) + (1-p)\log\left(\frac{1}{1-p}\right),$$

and so as $p \to 1$ or 0, $H(X) \to 0$. Since this value will come up later in the course, we will write

$$h(p) := p \log\left(\frac{1}{p}\right) + (1-p)\log\left(\frac{1}{1-p}\right).$$

It is not hard to see that the entropy of this particular X is maximised when $p = 1/2$, when $H(X) = 1$, and in fact in general we have that:

Lemma 4.1. *Let X be a discrete random variable and let R be the range of X.*

$$H(X) \leq \log(|R|).$$

with equality if X is uniformly distributed.

Proof. We will use the following consequence of Jensen's inequality. Let f be concave on $[a,b]$, $\lambda_i \geq 0$ such that $\sum_{i=1}^n \lambda_i = 1$ and let $x_1, \dots x_n \in [a,b]$. Then if we consider a real random variable Y taking the values x_i with probability λ_i, we have that

$$\sum_{i=1}^{n} \lambda_i f(x_i) = \mathbb{E}(f(Y)) \leq f(\mathbb{E}(Y)) = f\left(\sum_{i=1}^{n} \lambda_i x_i\right).$$

We note that $f(x) = \log(x)$ is a concave function on $(0,\infty)$, which can be seen since its derivative $\frac{1}{x}$ is decreasing on $(0,\infty)$, and so

$$H(X) = \sum_{x \in R} p(x) \log\left(\frac{1}{p(x)}\right) \leq \log\left(\sum_{x \in R} \frac{p(x)}{p(x)}\right) = \log(|R|).$$

Finally it is easy to see that if X is uniformly distributed then $p(x) = \frac{1}{|R|}$ for each $x \in R$ and so $H(X) = \log(|R|)$. □

This gives a useful connection between entropy and counting. We are going to define a whole host of generalisations of the entropy function, and in order to try and give you some intuition for such things, and give some working examples of calculating entropy, we'll keep a motivating example in mind as we go through these definitions.

Consider the probability space Ω given by a sequence of N fair coin flips for N very large, and the random variable $X : \Omega \to \{0,1\}^{[N]}$ where $X_i = 1$ if the ith coin flip was heads and 0 if it was tails. For every subset $A \subset [N]$ we can consider the random variable X_A given by the restriction of X to just the coordinates in A. In this way we have a correspondence between random variables and subsets.

Since X_A is uniformly distributed on $\{0,1\}^A$, Lemma 4.1 tells us that $H(X) = \log|\{0,1\}^A| = \log 2^{|A|} = |A|$. So, in this setting there is a correspondence between the entropy of X_A and the cardinality of the set A.

Given two discrete random variables, X and Y, we define the *joint entropy* (X,Y) to be

$$H(X,Y) = \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(x,y)}\right),$$

where, as before, $p(x,y) := \mathbb{P}(X = x, Y = y)$. Note that, if X and Y are independent then, by definition $p(x,y) = p(x)p(y)$ for all $x \in X$ and $y \in Y$, and so

$$\begin{aligned}
H(X,Y) &= \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(x,y)}\right) \\
&= \sum_x \sum_y p(x)p(y) \log\left(\frac{1}{p(x)p(y)}\right) \\
&= \sum_x \sum_y p(x)p(y) \left(\log\left(\frac{1}{p(x)}\right) + \log\left(\frac{1}{p(y)}\right)\right) \\
&= \sum_x p(x) \log\left(\frac{1}{p(x)}\right) \sum_y p(y) + \sum_y p(y) \log\left(\frac{1}{p(y)}\right) \sum_x p(x) \\
&= \sum_x p(x) \log\left(\frac{1}{p(x)}\right) + \sum_y p(y) \log\left(\frac{1}{p(y)}\right) \\
&= H(X) + H(Y)
\end{aligned}$$

However, in general that will not be the case.

So, in our example if we have two subsets A and B, what will the joint entropy of X_A and X_B be? Well X_A takes values in $\{0,1\}^A$ and X_B takes values in $\{0,1\}^B$, but given $x \in \{0,1\}^A$ and $y \in \{0,1\}^B$ it's not necessarily true that $p(x,y) = p(x)p(y)$, that is, the random variables X_A and X_B are not necessarily independent. Indeed, since X_A and X_B are restrictions of the same random variable X, for every $i \in A \cap B$ we have $(X_A)_i = (X_B)_i$.

So, what will the term $p(x,y)$ look like? Well, for a fixed $x \in \{0,1\}^A$, if y disagrees with x in a coordinate $i \in A \cap B$, then $p(x,y)$ is clearly 0. Otherwise, since A and B were both uniformly distributed over their range, $p(x,y) = 2^{|A\cap B|-|A|-|B|}$ and there are exactly $2^{|B|-|A\cap B|}$ such $y \in \{0,1\}^B$ which agree with x on $\{0,1\}^{A\cap B}$. Hence we can calculate

$$\begin{aligned} H(X_A, X_B) &= \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(x,y)}\right) \\ &= \sum_x 2^{|B|-|A\cap B|} \cdot 2^{|A\cap B|-|A|-|B|} \log 2^{|A|+|B|-|A\cap B|} \\ &= 2^{|A|+|B|-|A\cap B|} \cdot 2^{|A\cap B|-|A|-|B|} \log 2^{|A|+|B|-|A\cap B|} \\ &= |A| + |B| - |A \cap B| = |A \cup B|. \end{aligned}$$

So, in this context the joint entropy corresponds to the cardinality of the union $A \cup B$.

We also define the *conditional entropy* of Y given X in the following way. Let us write, as another shorthand, $p(y|x) := \mathbb{P}(Y = y | X = x)$, and similarly $p(x|y)$. We define

$$\begin{aligned} H(Y|X) &:= \sum_x p(x) \sum_y p(y|x) \log\left(\frac{1}{p(y|x)}\right) \\ &= \sum_x p(x) H(Y|X = x) \\ &= \mathbb{E}_x\big(H(Y|X = x)\big). \end{aligned}$$

Where the first equation is a definition, and the other equalities are merely different ways to rewrite this quantity. Note the difference between $H(Y|X = x)$, which is the entropy of the random variable $(Y|X = x)$, and $H(Y|X)$, which is the expected value of the latter over all possible values of x. In particular, $(Y|X)$ is not a random variable.

Back to our example, given subsets A and B and considering $H(X_B|X_A)$, what will $p(y|x)$ be? Well, as before, given a fixed x, this term is 0 unless x and y agree on $\{0,1\}^{A\cap B}$, and if they do agree on $A \cap B$ then it is clear that $p(y|x) = 2^{-|B\setminus A|}$. Also, for each x, there are exactly $2^{|B|-|A\cap B|} = 2^{|B\setminus A|}$ such y which agree with x on $\{0,1\}^{A\cap B}$. Hence we can calculate

$$\begin{aligned} H(X_B|X_A) &:= \sum_x p(x) \sum_y p(y|x) \log\left(\frac{1}{p(y|x)}\right) \\ &= \sum_x p(x) 2^{|B\setminus A|} 2^{-|B\setminus A|} \log\left(2^{|B\setminus A|}\right) \\ &= 2^{|A|} 2^{-|A|} \log\left(2^{|B\setminus A|}\right) \\ &= |B \setminus A|. \end{aligned}$$

So, in this context the conditional entropy corresponds to the cardinality of the set difference $B \setminus A$.

We can think of the conditional entropy as being the expected surprise in learning the value of Y, given that the value of X is known. We might expect, heuristically, that having extra knowledge should only decrease how surprised we are, and indeed that turns out to be the case:

Lemma 4.2 (Dropping conditioning)**.** *Let X, Y and Z be discrete random variables. Then*

$$H(Y|X) \leq H(Y).$$

Proof. Noting that $p(y)p(x|y) = p(x)p(y|x) = p(x,y)$, we see that

$$\begin{aligned}
H(Y|X) &= \sum_x p(x) \sum_y p(y|x) \log\left(\frac{1}{p(y|x)}\right) \\
&= \sum_y p(y) \sum_x p(x|y) \log\left(\frac{1}{p(y|x)}\right) \\
&\leq \sum_y p(y) \log\left(\sum_x \frac{p(x|y)}{p(y|x)}\right) \\
&= \sum_y p(y) \log\left(\sum_x \frac{p(x)}{p(y)}\right) \\
&= \sum_y p(y) \log\left(\frac{1}{p(y)}\right) \\
&= H(Y).
\end{aligned}$$

Where in the above we make repeated use of the fact that, if we sum the probabilities that a random variable takes a specific value over its entire range, then the result is 1, and Jensen's inequality (See Lemma 4.1) in the third line. □

Using our correspondence between the set world and the random variable world, we can now use Lemma 4.2 to say something about sets. Indeed, we have that

$$|B \setminus A| = H(X_B|X_A) \leq H(X_B) = |B|.$$

In a similar fashion, any identity or inequality about entropy will specialise to a combinatorial identity or inequality about finite sets. The converse is not true, but sometimes it can give intuition about what identities may hold. For example, we know that $|A \cup B| = |A| + |B \setminus A|$. Translating this back into the language of entropy would give the statement $H(X_A, X_B) = H(X_A) + H(X_B|X_A)$, which we will see in fact holds for all pairs of random variables.

Lemma 4.3 (Chain rule)**.** *Let X and Y be discrete random variables. Then*

$$H(X,Y) = H(X) + H(Y|X).$$

Proof.

$$\begin{aligned}
H(X,Y) &= \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(x,y)}\right) \\
&= \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(x)p(y|x)}\right) \\
&= \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(x)}\right) + \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(y|x)}\right) \\
&= \sum_x p(x) \log\left(\frac{1}{p(x)}\right) + \sum_x \sum_y p(x)p(y|x) \log\left(\frac{1}{p(y|x)}\right) \\
&= H(X) + \sum_x p(x) \sum_y p(y|x) \log\left(\frac{1}{p(y|x)}\right) \\
&= H(X) + H(Y|X).
\end{aligned}$$

□

One can also define the joint entropy of a sequence of discrete random variables $X_1, X_2, \ldots, X_n$ in a similar way and by induction it follows that

$$H(X_1, X_2, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \ldots H(X_n|X_1, X_2, \ldots, X_{n-1}).$$

We shall sometimes also refer to this as the *chain rule*. Note that, by Lemma 4.2 and Lemma 4.3 we have that

$$H(X_1, X_2, \ldots, X_n) \leq \sum_i H(X_i). \tag{4.1}$$

This seemingly quite simple statement is really quite useful, since it allows us to reduce the calculation of the entropy of a single random variable, to the calculation of many, hopefully simpler, random variables. Often, using this we can turn quite 'global' calculations into 'local' ones which are much easier to deal with.

So far we have an analogue of set union and set difference, so a natural idea would be consider the entropic function corresponding to intersection. Since $|A \cap B| = |A| + |B| - |A \cup B|$ this quantity should be represented by $H(X)+H(Y)-H(X,Y)$. We call this the *mutual information of X and Y* and it is denoted by $I(X;Y)$. Note that, by Lemma 4.3

$$I(X;Y) := H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

As the name suggests, we can think of this quantity of measuring the amount of information that X and Y share, and indeed this should be the amount of information 'left' from $H(X)$ after we get rid of the information remaining in X once we know Y, $H(X|Y)$. From Lemma 4.2 if follows that $I(X;Y) \geq 0$, and in fact by analysing when we get equality in Jensen's inequality one can show that $I(X;Y) = 0$ if and only if X and Y are independent. Hence, the mutual information is in some way a measure of the dependence of the random variables X and Y.

4.2 Brégman's Theorem

The *permanent* of an $n \times n$ matrix A is

$$\mathrm{perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)}$$

where S_n is the set of permutations of $[n]$. Note that this is very close to the definition of $\det(A)$, only with the factor of $(-1)^{\mathrm{sgn}(\sigma)}$ removed. Given a 0/1 matrix A we should expect that we can bound the permanent in terms of the number of non-zero entries of A in some way. In 1963 Minc gave a very natural conjecture for a bound given the row sums.

Conjecture 4.4 (Minc's Conjecture)**.** *Let A be an $n \times n$ 0/1 matrix such that the sum of the entries of the ith row is r_i. Then*

$$perm(A) \leq \prod_{i=1}^{n} (r_i!)^{\frac{1}{r_i}}.$$

It turns out this conjecture can be very easily transformed into an equivalent conjecture about graphs. There is a natural correspondence between $n \times n$ 0/1 matrices and bipartite graphs with partition classes of size n. Given such a matrix A we can consider a graph G on vertex set (V, W) where $V = \{v_1, \ldots, v_n\}$ and $W = \{w_1, \ldots, w_n\}$ with an edge between v_i and w_j if and only if $a_{ij} = 1$.

Now, a permutation σ gives a non-zero contribution to $\mathrm{perm}(A)$ if and only if $a_{i\sigma(i)} = 1$ for all $i \in [n]$, that is, if and only if $(v_i, w_{\sigma(i)})$ is an edge for every $i \in [n]$. However, since σ is injective, $\{(v_i, w_{\sigma(i)}) \colon i \in [n]\}$ gives a perfect matching of G. Conversely, any perfect matching M of G determines a permutation σ of $[n]$ given by $\sigma(i) = j$ such that $(v_i, w_j) \in M$, and the contribution of this permutation to $\mathrm{perm}(A)$ is non-zero. Putting this together we see that if we write $\Phi(G)$ for the set of perfect matchings of G and $\phi(G) = |\Phi(G)|$ then

$$\mathrm{perm}(A) = \phi(G).$$

Since the row sums of A are precisely the degrees of vertices in V, an upper bound on the permament of A in terms of the row sums is equivalent to an upper bound on the number of perfect matchings of G in terms of the degrees of vertices in one partition class. Minc's conjecture was proved by Brégman's, and so is now known as Brégman's Theorem, but we will give a proof using entropy methods due to Radhakrishnan.

Theorem 4.5 (Brégman's Theorem)**.** *Let G be a bipartite graph on vertex classes A and B such that $|A| = |B| = n$. Then*

$$\phi(G) \leq \prod_{v \in A} (d(v)!)^{\frac{1}{d(v)}}.$$

Proof. Let M be a perfect matching of G chosen uniformly at random from $\Phi(G)$. For convenience we will associate A with the set $[n]$ in the natural way, and denote by d_i the degree of the vertex i. For each $i \in [n]$ let X_i be the neighbour of i in M and we identify M with $X = (X_1, X_2, \ldots, X_n)$. More precisely, since M determines and is determined by $(X_1, X_2, \ldots, X_n)$ it follows that $H(M) = H(X)$.

Since M is unifromly distributed over $\phi(G)$ possibilities we have that $H(M) = H(X) = \log(\phi(G))$. Hence if we can bound $H(X)$ from above, we can also bound $\phi(G)$. Note that to get the stated bound we would need to show that

$$H(X) \leq \sum_{i=1}^{n} \frac{\log(d_i!)}{d_i}.$$

A naive first approach might be the use the sub-additivity of entropy to say

$$H(X) \leq \sum_{i=1}^{n} H(X_i),$$

and since there are at most d_i possibilities for the random variable X_i we have that

$$H(X) \leq \sum_{i=1}^{n} H(X_i) \leq \sum_{i=1}^{n} \log(d_i).$$

However, by Stirling's approximation, $\log(d_i!)/d_i \sim \log(d_i/e)$, and so this bound is not enough. However perhaps we can improve this bound by using the chain rule, since we have

$$H(X) = \sum_{i=1}^{n} H(X_i | X_1, X_2, \dots X_{i-1}).$$

We can think of this as revealing the matching one edge at a time, and working out the remaining entropy at each step given what we know. Now instead of just using the naive bound for each X_i we can hopefully take into account the fact that, if we already know $X_1, , X_2, \dots X_{i-1}$ this may reduce the number of possibilities for X_i, since some of the vertices $1, 2, \dots, i-1$ may be matched to neighbours of i in M, reducing the range of X_i.

So, we hope that the expected range of X_i once we've revealed X_1 up to X_{i-1} will be significantly lower than d_i so that we can bound the expected entropy by a smaller amount. However, the amount that this range will decrease will depend very much on the ordering we chose for the vertices when we identified A with $[n]$; if i doesn't have any shared neighbours with $1, 2, \dots, i-1$ in G then the range of X_i will not be reduced at all when we reveal $X_1, \dots, X_{i-1}$. Furthermore, to work out this expected gain, we would have to know something about the distribution of the edges in the matching M.

However, for each vertex $v \in A$ there are *some* orderings of A in which we might hope that it's likely that the range of X_v is reduced by revealing the neighbours of X_w with $w < v$, so at least we might hope that the range is reduced *on average*.

That is, perhaps a sensible idea is to average the chain rule over all possible orderings of A. Explicitly, given any permutation σ of $[n]$ we can apply the chain rule with respect to the ordering given by σ to see

$$H(X) = \sum_{i=1}^{n} H(X_{\sigma(i)} | X_{\sigma(1)}, X_{\sigma(2)}, \dots X_{\sigma(i-1)}).$$

Then, averaging over all possible choices of σ

$$H(X) \leq \frac{1}{n!} \sum_{\sigma} \sum_{i=1}^{n} H(X_{\sigma(i)} | X_{\sigma(1)}, X_{\sigma(2)}, \dots X_{\sigma(i-1)}).$$

It turns out that this idea will also make it easier to work out the expected gain we get by exposing the neighbours of earlier vertices.

For each $i \in [n]$ and permutation σ let us write $J_{\sigma,i} = \{k \colon \sigma(k) < \sigma(i)\} \subseteq [n] \setminus \{i\}$. Each term in the sum above is of the form $H(X_i|X_{J_{\sigma,i}})$. So we can re-write the sum as

$$\begin{aligned} H(X) &\leq \frac{1}{n!}\sum_{\sigma}\sum_{i=1}^{n} H(X_{\sigma(i)}|X_{\sigma(1)}, X_{\sigma(2)}, \dots X_{\sigma(i-1)}) \\ &= \frac{1}{n!}\sum_{i=1}^{n}\sum_{\sigma} H(X_i|X_{J_{\sigma,i}}) \end{aligned}$$

For each of these terms, if we think about calculating them sequentially, we've reduced the range of X_i by how many of the previously exposed X_j lie in $N(i)$, the neighbourhood of i. For each fixed value of $X_{J_{\sigma,i}}$, which corresponds to some sequence of $|J_{\sigma,i}|$ many vertices in B, say C, the entropy of X_i conditioned on $X_{J_{\sigma,i}} = C$ can be bounded above by $\log(|N(i) \setminus C|)$. So, let us denote by $N_\sigma(i) = N(i) \setminus \{X_j : j \in J_{\sigma,i}\}$ the vertices in the neighbourhood of i without those already chosen by some X_j.

It follows that, for any fixed σ and i we can calculate as follows, where C ranges over sequences of vertices in $N(i)$ of length $|J_{\sigma,i}|$

$$\begin{aligned} H(X_i|X_{J_{\sigma,i}}) &= \sum_{C} \mathbb{P}(X_{J_{\sigma,i}} = C) H(X_i|X_{J_{\sigma,i}} = C) \\ &\leq \sum_{j=1}^{d_i} \mathbb{P}(|N_\sigma(i)| = j) \log j \end{aligned}$$

Where we used the definition of conditional entropy, and then Lemma 4.1. However, since we're picking a random matching, it doesn't seem like we have any control over this improvement, since we don't know how much this will reduce the range of X_i.

However, for any fixed matching M, if we pick a random permutation σ, we claim that the size of $|N_\sigma(i)|$ is in fact uniformly distributed between 1 and d_i. Indeed, for a given matching we only care about the order in which we pick i and the vertices matched in M to the neighbours of i. Since i is equally likely to be chosen in any position in this list, the claim follows. In other words, for a fixed matching M, the proportion of σ such that $|N_\sigma(i)| = k$ is $\frac{1}{d_i}$ for each $1 \leq k \leq d_i$.

Since this is true separately for each particular matching, then it is also true when we pick a random matching. So, even though we can't bound any of the terms $\mathbb{P}(|N_\sigma(i)| = j)$ for a fixed σ, we can bound their average.

That is to say, if we pick M and σ both uniformly at random then

$$\mathbb{P}_{\sigma,M}(|N_\sigma(i)| = j) = 1/d_i$$

or equivalently

$$\frac{1}{n!}\sum_{\sigma} \mathbb{P}(|N_\sigma(i)| = j) = \frac{1}{d_i}$$

Hence,

$$\begin{aligned}
H(X) &= \frac{1}{n!}\sum_{\sigma}\sum_{i=1}^{n} H(X_{\sigma(i)}|X_{\sigma(1)}, X_{\sigma(2)}, \dots X_{\sigma(i-1)}) \\
&= \sum_{i=1}^{n}\frac{1}{n!}\sum_{\sigma} H(X_i|X_{J_{\sigma,i}}) \\
&\leq \sum_{i=1}^{n}\frac{1}{n!}\sum_{\sigma}\sum_{j=1}^{d_i}\mathbb{P}(|N_\sigma(i)| = j)\log(j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{d_i}\left(\sum_{\sigma}\frac{1}{n!}\mathbb{P}(|N_\sigma(i)| = j)\right)\log(j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{d_i}\frac{\log j}{d_i} = \sum_{i=1}^{n}\frac{\log(d_i!)}{d_i}
\end{aligned}$$

giving the bound as claimed. □

Note that this bound is tight. If we take G to be $\frac{n}{d}$ copies of $K_{d,d}$ then we have that $d(v) = d$ for all $v \in A$ and every matching consists of picking one from the $d!$ possible matchings on each $K_{d,d}$. Therefore.

$$\phi(G) = \prod_{i=1}^{\frac{n}{d}} d! = \prod_{v\in A}(d(v)!)^{\frac{1}{d(v)}}.$$

A natural question to ask is what happens for a non-bipartite G? It turns out a similar bound can be given, and as we will see in the examples sheet, it can actually be derived in a clever way from Brégman's Theorem.

Theorem 4.6. *Let $G = (V, E)$ be a graph with $|V| = 2n$. Then*

$$\phi(G) \leq \prod_{v\in V}(d(v)!)^{\frac{1}{2d(v)}}.$$

4.3 Shearer's lemma and projection inequalities

4.3.1 Shearer's Lemma

Given a sequence of discrete random variables random variables $X_1, X_2, \dots, X_n$ and some subset $A \subseteq [n]$ let define $X_A := (X_i : i \in A)$.

Lemma 4.7 (Shearer's inequality)**.** *Let $X_1, X_2, \dots, X_n$ be discrete random variables and $\mathcal{A}$ a collection (not necessarily distinct) of subsets of $[n]$, such that each $i \in [n]$ is in at least m members of $\mathcal{A}$. Then*

$$H(X_1, X_2, \dots, X_n) \leq \frac{1}{m}\sum_{A\in\mathcal{A}} H(X_A).$$

Proof. Let $A = \{a_1, a_2, \ldots, a_k\}$ with $a_1 < a_2 < \ldots < a_k$. We have that

$$\begin{aligned} H(X_A) &= H(X_{a_1}) + H(X_{a_2}|X_{a_1}) + \ldots + H(X_{a_k}|X_{a_1}, X_{a_2}, \ldots, X_{a_{k-1}}) \\ &\geq H(X_{a_1}|X_{<a_1}) + H(X_{a_2}|X_{<a_2}) + \ldots + H(X_{a_k}|X_{<a_k}), \end{aligned}$$

where $X_{<i} = (X_1, X_2, \ldots X_{i-1})$. This follows from repeated applications of the chain rule, and the fact that entropy only decreases if we condition on more variables. Therefore

$$\begin{aligned} \sum_{A \in \mathcal{A}} H(X_A) &\geq m. \sum_{i \in [n]} H(X_i|X_{<i}) \\ &= m.H(X_1, X_2, \ldots, X_n) \end{aligned}$$

□

4.3.2 The Bollobás-Thomason Box Theorem

Shearer's Lemma is closely related to notions of isoperimetry, the relation between the volume of a shape and it's 'perimeter' in the following way. If we think about a shape $S \subseteq \mathbb{R}^n$ with area $|S|$ then we can think about the process of picking a random point inside of S. This determines a vector $X = (X_1, \ldots, X_n)$ where the X_i are dependent on each other, depending on what the shape S is.

Suppose we take a very fine grid approximating $\mathbb{R}^n$, we can then think of S as being a discrete subset of this grid, whose number of points is proportional to $|S|$. Since this vector $X = (X_1, \ldots, X_n)$ now has some finite range, we can relate the volume of S directly to the entropy of X. That is

$$H(X) = \log |S|.$$

How can we interpret the random variable X_A for $A \subset [n]$? Well in this case, this is relatively clear, these correspond to the projections on the shape S onto the subspace spanned by the coordinates in A. That is, if we let S_A be the projection of S onto the subspace

$$\{(x_1, \ldots, x_n) : x_i = 0 \text{ for all } i \in A\}$$

Then the range of X_A is the 'volume' (in the $n - |A|$-dimensional sense) of S_A. We will write S_j for $S_{\{j\}}$.

In this way, Shearer's inequality gives us a way to relate the volume of a shape to it's lower dimensional projections. For example, if we just consider the 1-dimensional projections, we have the famous Loomis Whitney inequality:

Theorem 4.8 (The Loomis-Whitney inequality)**.** *Let $S \subset \mathbb{Z}^n$ then,*

$$|S|^{n-1} \leq \prod_{i=1}^{n} |S_{[n]\setminus\{i\}}|$$

For example, in two dimensions this simply says that the area of a shape can be bounded above by the product of its one-dimensional projections, a relatively trivial fact. But even in

three-dimensions it is not clear what the relationship should be between the volume of a shape and its projections onto two dimensional subspaces.

Notice that, this theorem is tight when $|S|$ is a 'box', that is, a set of the form $[1, m_1] \times [1, m_2] \times \ldots \times [1, m_n]$. Indeed, the volume of $|S|$ is $\prod_{i=1}^{n} m_i$ and the volume of the projection of S onto the hyperplane where $x_i = 0$ is just $\prod_{j \neq i} m_i$. This is perhaps not surprising, as a box represents the case where the X_is are independent, where we get equality in the argument for Shearer's inequality.

In fact, we will show a more general theorem, and deduce the Loomis-Whitney theorem as a corollary. We say a collection of sets $\mathcal{C} = \{C_1, \ldots, C_m\} \subset 2^{[n]}$ is a *k-uniform cover* if each $i \in [n]$ belongs to exactly k many of the C_j.

Theorem 4.9 (Uniform covers theorem)**.** *Let $S \subset \mathbb{Z}^n$ and let $\mathcal{C} \subseteq 2^{[n]}$ be a k-uniform cover, then*

$$|S|^k \leq \prod_{C \in \mathcal{C}} |S_C|$$

Remark 4.10. *Note that $\mathcal{C} = \{[n] \setminus \{i\} : i \in [n]\}$ is an $(n-1)$-uniform cover of $[n]$, and so Theorem 4.8 follows from Theorem 4.9.*

Proof. Let us choose a point $X = (X_1, \ldots, X_n)$ uniformly at random from S. Then, $H(X) = \log |S|$. Since $\mathcal{C}$ is a k-uniform cover, ever $i \in [n]$ is in at least k many $Cin\mathcal{C}$ (in fact, in exactly k many), and so by Lemma 4.7 it follows that

$$H(X) \leq \frac{1}{k} \sum_{C \in \mathcal{C}} H(X_C).$$

However, the range of X_C is $|S_C|$ and so it follows that

$$H(X) \leq \frac{1}{k} \sum_{C \in \mathcal{C}} \log |S_C|.$$

Combining the two equatons we see that

$$\log |S| \leq \frac{1}{k} \sum_{C \in \mathcal{C}} \log |S_C|$$

and so

$$|S|^k \leq \prod_{C \in \mathcal{C}} |S_C|,$$

as claimed. □

As before, if we consider the 1-uniform cover $\{\{i\} : i \in [n]\}$, Theorem 4.9 tells us the elementary fact the volume of a shape can be bounded by the product of its one-dimensional projections.

By taking limits of finer and finer grids it is possible to show that Theorem 4.9 also holds for subsets of $\mathbb{R}^n$ with the Lebesgue measure. In fact a rather amazing strengthening of Theorem 4.9 can be shown to hold, which is known as the Bollobás-Thomason Box Theorem. In what follows we will write $|S|$ for the Lebesgue measure of a set $S \subseteq \mathbb{R}^n$.

Theorem 4.11 (Bollobás-Thomason Box Theorem)**.** *Let $S \subset \mathbb{R}^n$ be compact. Then there is a box $A \subset \mathbb{R}^n$ such that $|A| = |S|$ and $|A_I| \leq |S_I|$ for all $I \subseteq [n]$.*

That is, for any shape we can find a box of the same volume such that *every* lower dimensional projection of this box has smaller volume than the corresponding projection of S. This immediately tells us that for any upper bound we might want to prove for the volume of a set in terms of the volumes of its projection, we only have to check that it holds for boxes.

Indeed, if we know that for every box A, $|A| \leq f(A_I : I \subset [n])$ for some function f which is increasing in each coordinate, then for any S we have that $|S| = |A| \leq f(A_I : I \subset [n]) \leq f(S_I : I \subset [n])$.

It is possible to prove this theorem via a continuous version of Theorem 4.9 (which can be proven analytically using Hölder's inequality) and a careful inductive argument, but we can also do so using an entropy argument, however to do so we will need to define a notion of entropy for continuous random variables. Suppose X is a continuous random variable taking values in $\mathbb{R}^n$ with probability density function f, then a natural guess for the *entropy* of X is the following:

$$H(X) = -\int f(x) \log f(x) dx$$

where integration is with respect to the Lebesgue measure. This does not inherit every property of the discrete entropy, for example it can take negative values, however, many of useful properties of H are still true in this setting, and we will assume without proof that the following are true:

- If $\mathbb{P}(X \in S) = 1$ then $H(X) \leq \log |S|$ with equality if X is uniform on S;
- For any X and Y, $H(X|Y) \leq H(X)$.
- If we write $X = (X_1, \ldots, X_n)$ and $X_I = (X_i : i \in I)$ then
$$H(X) = \sum_{i=1}^{n} H(X_i | X_{[i-1]}).$$

Proof of Theorem 4.11. Let X be a random variable uniformly distributed on S, then $H(X) = \log(|S|)$. Let us define $a_i = 2^{H(X_i|X_{[i-1]})}$ and let $A = [0, a_1] \times [0, a_2] \times \ldots \times [0, a_n]$ be a box in $\mathbb{R}^n$.

Now, for any $I \subseteq [n]$, X_I takes values in S_I, and hence $H(X_I) \leq \log |S_I|$. On the other hand, using the chain rule we see that, if $I = \{i_1 < i_2 < \ldots < i_k\}$

$$\begin{aligned}
H(X_I) &= H(X_{i_1}) + H(X_{i_2}|X_{i_1}) + \ldots + H(X_{i_k}|X_{i_1}, X_{i_2}, \ldots, X_{i_{k-1}}) \\
&\geq \sum_{j \in I} H(X_j | X_{[j-1]}) \\
&= \sum_{j \in I} \log a_j \\
&= \log \left(\prod_{j \in I} a_j \right) \\
&= \log |A_I|.
\end{aligned}$$

Hence, $\log|S_I| \leq \log|A_I|$ and so $|S_I| \leq |A_I|$. □

Since, as mentioned above, Theorem 4.9 is in fact equivalent to Shearer's lemma we might expect there to be an entropy equivalent to the Box Theorem, and we shall show on the example sheet that this is the case.

Theorem 4.12. *Let $X = (X_1, \dots, X_n)$ be a discrete random variable. Then there are non-negative constants $h_1, \dots, h_n$ such that $H(X) = \sum_i h_i$ and*

$$\sum_{i \in I} h_i \leq H(X_I)$$

for every $I \subseteq [n]$.

4.4 Independent Sets in a Regular Bipartite Graph

Let G be a d-regular bipartite graph on $2n$ vertices with vertex classes A and B, and let $\mathcal{I}(G)$ be the set of independent subsets of $V(G)$. We would like to bound this number from above. As in the case of Brégman's Theorem, letting G be a disjoint union of $K_{d,d}$'s seems a natural guess for a best possible graph. Indeed in G it is clear that any independent set in G consists of an arbitrary subset taken from one side of each $K_{d,d}$. Therefore we have that

$$|\mathcal{I}(G)| = (2^{d+1} - 1)^{\frac{n}{d}}.$$

The following proof of a corresponding upper bound on $|\mathcal{I}(G)|$ using entropy methods is due to Kahn.

Theorem 4.13. *Let G be a d-regular bipartite graph on $2n$ vertices with vertex classes A and B, and let $\mathcal{I}(G)$ be the set of independent subsets of $V(G)$. Then*

$$|\mathcal{I}(G)| \leq (2^{d+1} - 1)^{\frac{n}{d}}$$

Proof. The basic idea of the proof is the same as in Theorem 4.5, we pick a random independent set I and estimate the entropy $H(I)$. As before we have that $H(I) = \log(|\mathcal{I}|)$.

We identify I with its characteristic vector $(X_v : v \in A \cup B)$, note that I determines and is determined by (X_A, X_B). The idea is that, rather than splitting X into X_v for each v, we can use the neighbourhoods of each $v \in A$ as a d-uniform cover of the vertices of B, and so use Shearer's Lemma to express X_B in terms of $X_{N(v)}$.

For each $v \in A$ let $N(v)$ be the neighbourhood of v in B. Each $w \in B$ is in exactly d of the sets $N(v)$ and so we have

$$\begin{aligned} H(I) &= H(X_A|X_B) + H(X_B) \\ &\leq \sum_{v \in A} H(X_v|X_B) + \frac{1}{d}\sum_{v \in A} H(X_{N(v)}) \\ &\leq \sum_{v \in A} \big(H(X_v|X_{N(v)}) + \frac{1}{d}H(X_{N(v)}\big), \end{aligned}$$

where the second line follows from Shearer's inequality, and the third since $N(v) \subset B$.

Fix some $v \in A$. Let χ_v be the indicator random variable of the event that I contains a vertex of $N(v)$, and let $p := \mathbb{P}(\chi_v = 0)$, that is the probability that $I \cap N(v) = \emptyset$. The nice thing about this random variable is that it contains all the information about $X_{N(v)}$ that we need to determine $H(X_v|X_{N(v)})$.

Hence ,

$$\begin{aligned} H(X_v|X_{N(v)}) &\leq H(X_v|\chi_v) \\ &= \mathbb{P}(\chi_v = 0)H(X_v|\chi_v = 0) + \mathbb{P}(\chi_v = 1)H(X_v|\chi_v = 1) \\ &= \mathbb{P}(\chi_v = 0)H(X_v|\chi_v = 0) \leq p, \end{aligned}$$

since the event $\chi_v = 1$ determines that $X_v = 0$, and since $H(X_v) \leq \log(|\text{range}(X_v)|) = 1$.

Also,

$$\begin{aligned} H(X_{N(v)}) &= H(X_{N(v)}, \chi_v) \\ &= H(\chi_v) + H(X_{N(v)}|\chi_v) \\ &\leq h(p) + (1-p)\log(2^d - 1), \end{aligned}$$

where $h(p) = p\log(1/p) + (1-p)\log(1/(1-p))$. Putting these inequalities together gives us

$$H(I) \leq \sum_{v \in A} (p + \frac{1}{d}\left(h(p) + (1-p)\log\left(2^d - 1\right)\right).$$

All that remains is to maximise the quantity on the right hand side according to p. It is a simple exercise to check that the function is convex, and to calculate its derivative, giving that the maximum is attained at $p = 2^d/(2^{d+1} - 1)$, and so $(1-p) = 2^d - 1/(2^{d+1} - 1)$ giving that:

$$\begin{aligned} H(I) &\leq \sum_{v \in A} (p + \frac{1}{d}\left(h(p) + (1-p)\log\left(2^d - 1\right)\right) \\ &= n\left(p + \frac{1}{d}\left(p\log(1/p) + (1-p)\log(1/(1-p)) + (1-p)\log\left(2^d - 1\right)\right)\right) \\ &= n\left(p + \frac{1}{d}\left(p\log\left(\frac{2^{d+1}-1}{2^d}\right) + (1-p)\log\left(\frac{2^{d+1}-1}{2^d - 1}\right) + (1-p)\log\left(2^d - 1\right)\right)\right) \\ &= n\left(p + \frac{1}{d}\left(p\log\left(2^{d+1} - 1\right) - pd + (1-p)\log\left(2^{d+1} - 1\right)\right)\right) \\ &= n\left(p - p + \frac{1}{d}\left((p + (1-p))\log\left(2^{d+1} - 1\right)\right)\right) \\ &= n\frac{1}{d}\log\left(2^{d+1} - 1\right) \end{aligned}$$

$$\log(|\mathcal{I}|) = H(I) \leq n.\frac{1}{d}\log\left(2^{d+1} - 1\right),$$

from which the result follows. □

As with Theorem 4.5 it is possible to deduce a bound for the number of independent sets in a general d-regular graph from Theorem 4.13, however it requires a clever idea using the bipartite double cover of G, which wasn't discovered for quite some time.

There's another way to approach this proof, starting from the application of Shearer's Lemma to see that

$$\sum_{v\in A}\big(H(X_v|X_{N(v)}) + \frac{1}{d}H(X_{N(v)}\big),$$

which is slightly more conceptually difficult, but is easier to generalise.

The idea is to fix some $v \in A$ and look at the expression

$$dH(X_v|X_{N(v)}) + H(X_{N(v)}).$$

Recall that $X_{N(v)}$ is some vector in $\{0,1\}^{N(v)} \sim \{0,1\}^d$ and X_v is just some $\mathrm{Ber}(p)$ random variable for some (unknown) p. Furthermore, if we condition on $X_{N(v)}$ being non-zero, then X_v has to be 0.

The idea is to replace X_v with d independent, identically distributed random variables $X_v^1, \dots, X_v^d$, each of which has the same joint distribution with $X_{N(v)}$ as X_v does. We then have that

$$\begin{aligned} dH(X_v|X_{N(v)}) + H(X_{N(v)} &= \sum_{i=1}^{d} H(X_v^i|X_{N(v)}) + H(X_{N(v)} \\ &= H(X_v^1, X_v^2, \dots, X_v^d|X_{N(v)}) + H(X_{N(v)} \\ &= H(X_{N(v)}, X_v^1, X_v^2, \dots, X_v^d), \end{aligned}$$

since the X_v^i are independent, and then using the chain rule in the last line. However $Y = (X_{N(v)}, X_v^1, X_v^2, \dots, X_v^d)$ is distributed on $\{0,1\}^{2d}$, and is such that if there is a 1 in the first d co-ordinates then there last d co-ordinates are all 0.

Hence we can think of Y as being distributed on the set of independent sets of $K_{d,d}$. It might not be uniformly distributed, but it follows that

$$H(Y) \le \log|\mathcal{I}(K_{d,d})| = \log\left(2^{d+1} - 1\right).$$

This proof generalises quite well to counting *homomorphisms* from G to a fixed (multi-)graph H. A graph homomorphism is a map $f\colon V(G) \to V(H)$ such that every edge in G is mapped to an edge in H. Indeed, then if we consider H to be a graph on $\{0,1\}$ with an edge between 0 and 1 and a loop at 1, then a map $f : G \to \{0,1\}$ can be seen to be a homomorphism if and only if $f^{-1}(0)$ is an independent set. Similarly you can count proper k-colourings of G as the number of homomorphism from G to K_k. If we write $\mathrm{Hom}(G,H)$ for the set of homomorphisms from G to H then essentially the same proof as above shows the following.

Theorem 4.14. *Let G be a d-regular bipartite graph on $2n$ vertices, then for any multigraph H*

$$|Hom(G,H) \le |Hom(K_{d,d},H)|^{\frac{n}{d}}.$$

Unfortunately the bipartite double cover trick doesn't work here to extend this result to arbitrary graphs, and in fact the result *isn't true* in general, which can be seen by considering for example two disjoint loops.

5 The Container Method

5.1 Triangle-free graphs

Many well-known problems in combinatorics concern families of discrete objects which avoid certain forbidden configurations, for example H-free graphs, or sets of integers containing no arithmetic progression. Often one is interested in the size and structure of extremal examples e.g. Turán's Theorem or Szemerédi's Theorem. Recently, there has been increasing interest in what can be said about the typical structure of members of these families (i.e a uniformly chosen example), and to extremal questions in (sparse) random graphs.

One recent method that has proved useful in considering such problems is to consider them as example of a more general question about independent sets in hypergraphs. As an illustrative example we will focus on the family $\mathcal{F}_n(K_3)$ of triangles free graphs with vertex set $[n]$.

Now, since every bipartite graph is triangle free, it is not hard to see that there are least $2^{\frac{n^2}{4}}$ triangle free graphs on n vertices. However, it turns out that there is a much smaller family $\mathcal{G}_n$ of graphs on n vertices, where $|\mathcal{G}_n| = n^{O\left(n^{\frac{3}{2}}\right)}$, that forms a set of *containers* for $\mathcal{F}_n(K_3)$, which means that for every $H \in \mathcal{F}_n(K_3)$ there exists some $G \in \mathcal{G}_n$ such that $H \subseteq G$. Obviously we can find such a family if we don't restrict the structure of the graphs in $\mathcal{G}_n$ in any way, since the graph K_n by itself would suffice, however the useful thing is that we can find such a collection where every $G \in \mathcal{G}_n$ is 'almost triangle-free', in the sense that it contains 'few' triangles.

Why might this be useful? Well suppose we're interested in the maximum number of edges in a triangle-free subgraph of $G(n,p)$. Clearly by taking the intersection of $G(n,p)$ with $K_{\frac{n}{2},\frac{n}{2}}$ we can whp find such a subgraph with about half the number of edges in $G(n,p)$. How might we prove this to be optimal?

Well, an obvious approach would be to try to show that the number of triangle-free subgraphs of $G(n,p)$ with at least m edges is 0 for all m much larger than this, and a first attempt to show that would be via the first moment method. However, if we let X_m be the number of such subgraphs then again just by considering the subgraphs of $K_{\frac{n}{2},\frac{n}{2}}$ with at least m edges we have (as long as $m << n$)

$$\mathbb{E}(X_m) \geq \binom{\frac{n^2}{4}}{m} p^m = \left((1+o(1))\frac{epn^2}{4m}\right)^m$$

which will be very large when m is around $\frac{p}{2}\binom{n}{2}$.

However, this collection of containers tells us that the set of triangle-free graphs with m edges are 'clustered' together, which is creating a strong positive correlation between the events encoding their appearances in $G(n,p)$. By understanding this clustering we can group them into related parts and deal with each part in one go. More precisely, if $G(n,p)$ contained a triangle-free graph with m edges then it must contain many of the edges in one of our containers $G \in \mathcal{G}_n$. Then, not only is this very unlikely to happen, $\mathcal{G}_n$ is a small enough collection that we can use the union bound to conclude that whp $G(n,p)$ doesn't contain many edges in any of the $G \in \mathcal{G}_n$.

So, how does this relate to independent sets in hypergraphs? Let us consider the 3-uniform

hypergraph $\mathcal{H}$ with vertex set $V(\mathcal{H}) = E(K_n)$ and edge set

$$E(\mathcal{H}) = \{(e_1, e_2, e_3) \subseteq E(K_n) \colon e_1, e_2, e_3 \text{ form a triangle}\}.$$

Then, an independent set in $\mathcal{H}$ is a set of vertices of $\mathcal{H}$, and so corresponds to a set of edges in K_n, or in other words a graph on n vertices, which doesn't contain any hyperedge, or in other words in graph doesn't contain any triangle. Hence there is a natural bijection between $\mathcal{I}(\mathcal{H})$, the family of independent sets of $\mathcal{H}$, and $\mathcal{F}_n(K_3)$. Hence, our previous statement about the existence of our family of containers $\mathcal{G}_n$ is equivalent to the existence of a 'small' family $\mathcal{C}$ of subsets of $V(\mathcal{H})$, each containing only 'few' edges of $\mathcal{H}$, such that every independent set $I \in \mathcal{I}(\mathcal{H})$ is contained in some member of $\mathcal{C}$. In other words we can find a 'small' collection of almost independent sets which, between them, 'contain' all of the independent sets in $\mathcal{H}$.

It turns out that the only properties of the hypergraph $\mathcal{H}$ that are necessary to prove such a result are some assumptions on the maximum degree and codegree of vertices in $\mathcal{H}$. If $\mathcal{H}$ is a k-uniform hypergraph and $A \subseteq V(\mathcal{H})$ we write $d_{\mathcal{H}}(A) = |\{e \in E(\mathcal{H}) \colon A \subseteq e\}|$ and for $\ell \leq k$ we write

$$\Delta_\ell(\mathcal{H}) = \max\{d_{\mathcal{H}}(A) \colon |A| = \ell\}.$$

Lemma 5.1. *For every $c > 0$ there exists $\delta > 0$ such that if $\mathcal{H}$ is a 3-uniform hypergraph with average degree $d \geq \delta^{-1}$ and*

$$\Delta_1(\mathcal{H}) \leq cd \quad \text{and} \quad \Delta_2(\mathcal{H}) \leq c\sqrt{d},$$

then there exists a collection $\mathcal{C}$ of subsets of $V(\mathcal{H})$ with

$$|\mathcal{C}| \leq \binom{v(\mathcal{H})}{\frac{v(H)}{\sqrt{d}}}$$

such that

(a) for every $I \in \mathcal{I}(H)$, there exists $C \in \mathcal{C}$ such that $I \subseteq C$; and

(b) $|C| \leq (1-\delta)v(\mathcal{H})$ for every $C \in \mathcal{C}$.

Now, if $\mathcal{H}$ is the triangle containment hypergraph, then we have that $v(\mathcal{H}) = \binom{n}{2}$, $d = \Delta_1(\mathcal{H}) = n - 2$, since each edge in contained in exactly $n-2$ triangles, and $\Delta_2(\mathcal{H}) = 1$, since a pair of edges is contained in at most one triangle. Hence we can apply Lemma 5.1 with $c = 1$ to find a collection $\mathcal{C}$ of subgraphs of K_n with

$$|\mathcal{C}| \leq \binom{\binom{n}{2}}{\frac{\binom{n}{2}}{\sqrt{n}}} = n^{O\left(n^{\frac{3}{2}}\right)}$$

such that

(a) every triangle-free graph is a subgraph of some $C \in \mathcal{C}$; and

(b) Each $C \in \mathcal{C}$ has at most $(1-\delta)\binom{n}{2}$ edges.

However, this second condition isn't quite what we claimed, that these containers would be almost triangle free. In order to get this stronger conclusion we will have to iteratively apply the lemma.

Theorem 5.2. *For each $\varepsilon > 0$ there exist $C > 0$ such that for each $n \in \mathbb{N}$ there exists a collection $\mathcal{G}_n$ of graphs on n vertices with*

$$|\mathcal{G}_n| \leq n^{Cn^{\frac{3}{2}}},$$

such that

(a) every triangle-free graph on n vertices is a subgraph of some $G \in \mathcal{G}_n$; and

(b) Each $G \in \mathcal{G}_n$ contains at most εn^3 triangles.

Proof. We apply Lemma 5.1 with $c = 1$ to the triangle-free hypergraph to obtain a family $\mathcal{C}$ which satisfies property (a) by the conclusion of the lemma. Suppose then that $\mathcal{C}$ doesn't satisfy property (b), and so there are some $C \in \mathcal{C}$ containing at least ϵn^3 triangles. We will apply Lemma 5.1 again to the subhypergraph $\mathcal{H}[C]$ of $\mathcal{H}$ induced by C for each such C.

Note that, by assumption the average degree of $\mathcal{H}[C]$ is at least $6\varepsilon n$ since each triangle in C gives an edge in $\mathcal{H}[C]$ and $v(\mathcal{H}[C]) \leq \binom{n}{2}$ (in fact, less than $(1-\delta)\binom{n}{2}$). Furthermore Δ_1 and Δ_2 can not increase, and so we can apply the lemma with $c = \frac{1}{\varepsilon}$ and replace C in $\mathcal{C}$ with the collection of containers for $\mathcal{I}[(\mathcal{H}[C])$ given by the lemma.

Since each time we iterate this process the size of the new containers shrinks by at least $(1-\delta)$, eventually the containers we produce will have size size at most εn^2, and so contain at most εn^3 triangles (since $\Delta_1(\mathcal{H}) \leq n$). Furthermore, the total number of iterations will be bounded as a function of ϵ and δ.

Since each application of the lemma produces at most $n^{O\left(n^{\frac{3}{2}}\right)}$ many new containers, the total number of containers in the final collection is still $n^{O\left(n^{\frac{3}{2}}\right)}$. □

5.2 Applications of Theorem 5.2

Let us demonstrate the strength of Theorem 5.2 with some applications. The first concerns a well-known result of Mantel in extremal graph theory, which is a precursor to Turán's theorem.

Theorem 5.3. *Let G be a graph on n vertices with at least $\frac{n^2}{4}$ edges, then G contains a triangle.*

In other words, every subgraph of K_n with at least $\frac{1}{2}e(K_n)$ many edges contains a triangle. The corresponding problem in the random graph $G(n,p)$ was first considered by Frankl and Rödl, who prove the following:

Theorem 5.4. *For every $\delta > 0$ there exists $C > 0$ such that if $p \geq \frac{C}{\sqrt{n}}$ then whp every subgraph $G \subseteq G(n,p)$ with*

$$e(G) \geq \left(\frac{1}{2} + \alpha\right) p \binom{n}{2}$$

contains a triangle.

Note that, for p much smaller the expected number of triangles is $O(n^3p^3)$, which is asymptotically smaller than the expected number of edges $\binom{n}{2}p$, and so, since both are well concentrated, whp by removing an edge from each triangle we can find a triangle-free subgraph with $(1+o(1))p\binom{n}{2}$ many edges.

In order to prove this we will need the following so called *supersaturation* result for triangles.

Lemma 5.5. *For every $\delta > 0$ there exists $\varepsilon > 0$ such that if G is a graph on n vertices with*

$$e(G) \geq \left(\frac{1}{2} + \delta\right)\binom{n}{2},$$

then G contains at least εn^3 triangles.

Proof. If we calculate the average number of edges contained in a induced subgraph $G[U]$ on $|U| = N$ vertices, where N is large and constant, which we denote by D we see that

$$\begin{aligned} D &= \frac{\sum_U e(G[U])}{\binom{n}{N}} \\ &= e(G)\frac{\binom{n-2}{N-2}}{\binom{n}{N}} \\ &\geq \left(\frac{1}{2} + \delta\right)\binom{N}{2} \end{aligned}$$

if N is large enough. Since $e(G[U]) \leq \binom{N}{2}$ it follows that at least a $\frac{\delta}{2}$ proportion of the U satisfy

$$e(G[U]) \geq \left(\frac{1}{2} + \frac{\delta}{2}\right)\binom{N}{2}$$

since otherwise

$$\begin{aligned} D &= \frac{\sum_U e(G[U])}{\binom{n}{N}} \\ &< \frac{\frac{\delta}{2}\binom{n}{N}\binom{N}{2} + \left(1-\frac{\delta}{2}\right)\binom{n}{N}\left(\frac{1}{2}+\frac{\delta}{2}\right)\binom{N}{2}}{\binom{n}{N}} \\ &\leq \left(\frac{1}{2} + \frac{\delta}{2}\right)\binom{N}{2}. \end{aligned}$$

It follows from Theorem 5.3 that at least a $\frac{\delta}{2}$ proportion of the U contain a triangle and hence, if we write $T(G)$ for the number of triangles in G

$$\begin{aligned} T(G) &= \frac{\sum_U T(G[U])}{\binom{n-3}{N-3}} \\ &\geq \frac{\frac{\delta}{2}\binom{n}{N}}{\binom{n-3}{N-3}} \\ &\geq \frac{\delta}{4N^3}n^3. \end{aligned}$$

Hence we can take $\varepsilon = \frac{\delta}{4N^3}$. □

Using this we can give a proof of Theorem 5.4 under the slightly stronger assumption that $p = \omega\left(\frac{\log n}{\sqrt{n}}\right)$. Later on we will show how to get rid of this extra log term.

Proof of Theorem 5.4 for large p. By Lemma 5.5 with $\delta = \frac{\alpha}{2}$ there is some ε such that any graph G on n vertices with $e(G) \geq \left(\frac{1}{2} + \frac{\alpha}{2}\right)\binom{n}{2}$ contains at least εn^3 many triangles. Let us apply Theorem 5.2 with this ε to obtain a family $\mathcal{G}_n$ of containers such that each $G \in \mathcal{G}_n$ contaisn fewer than εn^3 many edges, and hence has at most $\left(\frac{1}{2} + \frac{\alpha}{2}\right)\binom{n}{2}$ many edges.

Then, since every triangle free graph is a subgraph of some $G \in \mathcal{G}_n$, if $G(n,p)$ contains a triangle-free graph with m edges, then in particular $e(G \cap G(n,p)) \geq m$ for some $G \in \mathcal{G}_n$.

However, $e(G \cap G(n,p)) \sim \mathrm{Bin}\left(e(G), p\right)$ and hence $e(G \cap G(n,p))$ is stochastically dominated by $\mathrm{Bin}\left(\left(\frac{1}{2} + \frac{\alpha}{2}\right)\binom{n}{2}, p\right)$ and hence by the Chernoff bounds

$$\mathbb{P}\left(e(G \cap G(n,p)) \geq \left(\frac{1}{2} + \alpha\right) p\binom{n}{2}\right) \leq \exp\left(-\beta p n^2\right)$$

for some constant $\beta(\alpha) > 0$. Since there are at most $n^{O\left(n^{\frac{3}{2}}\right)}$ many containers G, it follows by the union bound that

$$\mathbb{P}\left(\text{there exists } G \in \mathcal{G}_n \text{ with } e(G \cap G(n,p)) \geq \left(\frac{1}{2} + \alpha\right) p\binom{n}{2}\right) \leq n^{O\left(n^{\frac{3}{2}}\right)} \exp\left(-\beta p n^2\right) = o(1).$$

Hence, with high probability $G(n,p)$ does not contain any triangle-free graph with $\geq \left(\frac{1}{2} + \alpha\right) p\binom{n}{2}$ many edges. □

Similarly the following result of Rödl and Ruciński has a simple proof via Theorem 5.2

Theorem 5.6. *Let $r \in \mathbb{N}$ then there exists $C > 0$ such that if $p \geq \frac{C}{\sqrt{n}}$, then who every r-colouring of the edges of $G(n,p)$ contains a monochromatic triangle.*

Again the proof uses a supersaturation result, namely:

Lemma 5.7. *For every $r \in \mathbb{N}$ there exist n_0 and $\varepsilon > 0$ such that for all $n \geq n_0$, every r-colouring of the edges of K_n contains at least $r\varepsilon n^3$ monochromatic triangles.*

The proof follows as with Lemma 5.5 by applying Ramsey's theorem to the colourings induced by subsets of $[n]$ of size N for some large N.

Using this we can again give a proof of Theorem 5.6 under the slightly stronger assumption that $p = \omega\left(\frac{\log n}{\sqrt{n}}\right)$.

Proof of Theorem 5.6 for larger p. We apply Lemma 5.7 to find an ϵ such that every $(r+1)$-colouring of the edges of K_n contains at least $(r+1)\varepsilon n^3$ monochromatic triangles, and then apply Theorem 5.2 with this ε to get a family of containers $\mathcal{G}_n$, each of which contains fewer than εn^3 many triangles.

Suppose that $G(n,p)$ does not satisfy the conclusion of the theorem, then there is an r-colouring such that each of the subgraphs H_i spanned by the colour $i \in [r]$ are triangle-free. Thus $G(n,p) = \bigcup_{i=1}^r H_i$. By assumption there exists containers G_i such that $H_i \subseteq G_i$ for each i, and so $G(n,p) \subseteq \bigcup_{i=1}^r G_i$.

We now apply Lemma 5.7 to the $(r+1)$-colouring given by assigning to each edge $e \in \bigcup_{i=1}^r G_i$ some colour i such that $e \in G_i$ and the colour $(r+1)$ to the edges in $K_n \setminus \bigcup_{i=1}^r G_i$.

Lemma 5.7 implies that this colouring has at least $(r+1)\varepsilon n^3$ many monochromatic triangle, and each G_i contains at most εn^3 many triangles, and so $K_n \setminus \bigcup_{i=1}^r G_i$ contains at least εn^3 many triangles.

However, each edge of K_n belongs to fewer than n triangles and so $e\left(K_n \setminus \bigcup_{i=1}^r G_i\right) \geq \varepsilon n^2$. Consequently, for every fixed collection $G_1, \dots, G_r \in \mathcal{G}_n$,

$$\mathbb{P}\left(G(n,p) \subseteq \bigcup_{i=1}^r G_i\right) = (1-p)^{e\left(K_n \setminus \bigcup_{i=1}^r G_i\right)} \leq (1-p)^{\varepsilon n^2} \leq e^{-\varepsilon p n^2}.$$

Since there are at most $n^{O\left(n^{\frac{3}{2}}\right)}$ many containers in $\mathcal{G}$, by the union bound the probability that $G(n,p) \subseteq \bigcup_{i=1}^r G_i$ for any collection $G_1, \dots G_r \in \mathcal{G}_n$ is at most

$$\binom{|\mathcal{G}_n|}{r} e^{-\varepsilon p n^2} \leq n^{O\left(n^{\frac{3}{2}}\right)} e^{-\varepsilon p n^2} = o(1).$$

□

As before this extra log factor can be removed with a more careful analysis which we will perform later.

Finally, a third application comes from the following theorem of Łuczak. We say that a a graph G is *t-close to bipartite* if there exists a bipartite subgraph $G' \subseteq G$ with $e(G') \geq e(G) - t$.

Theorem 5.8. *For every $\alpha > 0$, there exists a $C > 0$ such that if $m \geq Cn^{\frac{3}{2}}$, then almost all triangle-free graphs with n vertices and m edges are αm-close to bipartite.*

Again the proof goes via a supersaturation type result, although this one has a slightly more involved proof, which we will not give.

Lemma 5.9. *For every $\delta > 0$ there exists $\varepsilon > 0$ such that if G is a graph on n vertices with*

$$e(G) \geq \left(\frac{1}{2} - \varepsilon\right)\binom{n}{2},$$

then either G is δn^2 close to bipartite or G contains at least εn^3 many triangles.

Again we only give a proof for larger $m = \omega\left(n^{\frac{3}{2}} \log n\right)$.

Proof of Theorem 5.6 for larger m. We apply Lemma 5.9 with $\delta(\alpha)$ sufficiently small to find ε such that the conclusion of the lemma holds. Applying Theorem 5.2 with this ε we get a family of containers $\mathcal{G}_n$, each of which contains fewer that εn^3 many triangles, and so by Lemma 5.9 each $G \in \mathcal{G}_n$ is either δn^2 close to bipartite or satisfies

$$e(G) \leq \left(\frac{1}{2} - \varepsilon\right)\binom{n}{2}.$$

So, let us estimate the number of triangle-free graphs H with n vertices m edges which are not αm-close to bipartite; note that each such H is subgraph of some container $G \in \mathcal{G}_n$.

Suppose first that G satisfies the second condition, and so has few edges. In which case the number of subgraphs H of G with exactly m edges, triangle-free or otherwise, is at most

$$\binom{e(G)}{m} \leq \binom{\left(\frac{1}{2} - \varepsilon\right)\binom{n}{2}}{m} \approx (1-2\varepsilon)^m \binom{\frac{n^2}{4}}{m}.$$

On the other hand, if there is some bipartite $G' \subseteq G$ such that $e(G') \geq e(G) - \delta n^2$ then since $e(H \cap G') \leq (1-\alpha)m$, by our assumption on H, the number of such H is at most

$$\begin{aligned}\binom{e(G) - e(G')}{\alpha m}\binom{e(G)}{(1-\alpha)m} &\leq \binom{\delta n^2}{\alpha m}\binom{\binom{n}{2}}{(1-\alpha)m} \\ &\leq \left(e\frac{\delta^\alpha}{\alpha^\alpha(2(1-\alpha))^{1-\alpha}}\right)^m \left(\frac{n^2}{m}\right)^m \\ &\leq 2^{-m}\binom{\frac{n^2}{4}}{m}\end{aligned}$$

as long as $\delta(\alpha)$ is small enough. Hence, by summing over every possible $G \in \mathcal{G}_n$ the total number of such H at most

$$n^{O\left(n^{\frac{3}{2}}\right)}(1-2\varepsilon)^m \binom{\frac{n^2}{4}}{m} \ll \binom{\frac{n^2}{4}}{m}$$

since $m \gg n^{\frac{3}{2}} \log n$.

However, since every bipartite graph is triangle free there are clearly at least $\binom{\frac{n^2}{4}}{m}$ many triangle-free graphs H on n vertices and with m edges. □

5.3 The Container Lemma

Let us start by considering the easier case of 2-uniform hypergraphs, i.e graphs.

Lemma 5.10. *For every $c > 0$ there exists $\delta > 0$ such that if G is a graph with average degree d, maximum degree $\Delta(G) \leq cd$ and $\tau := \frac{2\delta}{d}$ then there exists a collection $\mathcal{C}$ of susbets of $V(G)$ with*

$$|\mathcal{C}| \leq \binom{v(G)}{\lceil \tau v(G) \rceil},$$

such that

(a) for every $I \in \mathcal{I}(G)$, there exists $C \in \mathcal{C}$ such that $I \subseteq C$; and

(b) $|C| \leq (1-\delta)v(G)$ for every $C \in \mathcal{C}$.

Proof. The key idea in the proof will be to 'encode' each independent set $I \in \mathcal{I}(G)$ with a subset $S(I) \subseteq I$, which we will call the *fingerprint* of I. S will be small, but also crucially will have the property that knowing that I has a fingerprint $S(I) = S$ will be sufficient to guarantee that I avoids a 'large' subset of V(G), in fact a positive proportion. Hence, there is some set $C(S)$ of size at most $(1-\delta)v(G)$ such that every I with $S(I) = S$ is contained in C, and we take these C as our containers.

We will construct the fingerprint of each I using a simple, deterministic algorithm. During the algorithm we will maintain a partition of $V(G) = A \cup S \cup X$ where A is the set of 'active' vertices, S is the current version of the fingerprint, and X is the set of 'excluded' vertices, which are not in I. We start with $A = V(G)$ and $S = X = \emptyset$.

We first define an order on $V(G)$, called the *max-degree order*, by letting v_1 be the vertex of maximum degree in G, then v_2 be the vertex of maximum degree in $G[V \setminus \{v_1\}]$, and so on, so that for every $i \in [n]$, v_i is the vertex of maximum degree in $G[\{v_i, \dots, v_n\}]$.

As long as $|X| \leq \delta v(G)$ we repeat the following steps:

1. Let v be the first vertex of I in the max-degree order on $G[A]$.
2. Move v into S.
3. Move the neighbours of v into X.
4. Move the vertices preceding v in the max-degree order on $G[A]$ into X.
5. Remove the new vertices of $S \cup X$ from A.

The algorithm will terminate when $|X| > \delta v(G)$ or $I \cap A = \emptyset$, let $A(I), S(I)$ and $X(I)$ be the final values of A, S and X. We claim that $S(I)$ will have the desired properties to be the fingerprint.

Firstly, we claim that if $I, I' \in \mathcal{I}(G)$ are such that $S(I) = S(I')$, then also $A(I) = A(I')$. Indeed, it is easy to see inductively that the vertices must have been added into $S(I)$ and $S(I')$ in the same order during the algorithm, and hence the same vertices were moved into $X(I)$ and $X(I')$ at each stage, and so also $A(I) = A(I')$.

Given any S such that there is some $I \in \mathcal{I}(G)$ with $S(I) = S$, let us define the *container* of S to be $C(S) = A(I) \cup S$ if the algorithm terminated since $|X| > \delta v(G)$ or and S otherwise. Note that by the above this gives at most two different containers for each S. Let us further define

$$\mathcal{C} = \{C(S) \colon S = S(I) \text{ for some } I \in \mathcal{I}(G)\}.$$

Let us first show that $|S(I)| \leq \lceil \tau v(G) \rceil$ for each $I \in \mathcal{I}(G)$, which will then imply the bound on $|\mathcal{C}|$. To see this, we claim that whenever a vertex v is added to S in the algorithm at least

$\min\{\frac{d}{2}, \delta V(G)\}$, which we will assume to be $\frac{d}{2}$ for simplicity, vertices are added to X. Note that in this case, after $\lceil \tau v(G) \rceil$ many vertices have been added to S,

$$|X| \geq \lceil \tau v(G) \rceil \frac{d}{2} \geq \delta v(G).$$

Hence, the algorithm must stop before $\lceil \tau v(G) \rceil$ vertices have been added to S.

So, let us show that we always add at least $\frac{d}{2}$ vertices to X.

Suppose that $|S| \leq \tau v(G)$, $|X| \leq \delta v(G)$ and we just added a vertex v to S. If the set of vertices which preceded v in the max-degree order on $G[A]$, which we call W is large, then we are happy, so we may assume that $|W| \leq \frac{d}{2}$. However, then

$$\begin{aligned} e(G[A \setminus W]) &\geq e(G) - \big((\tau + \delta)v(G) + \frac{d}{2}\big)\Delta(G) \\ &\geq e(G) - \big((\tau + 2\delta)v(G)\big)cd \\ &\geq e(G)\left(1 - 4c\delta\left(1 + \frac{1}{d}\right)\right) \\ &\geq \frac{e(G)}{2} \end{aligned}$$

if δ is sufficiently small.

Then, since v is a vertex of maximum degree in $G[A \setminus W]$ it follows that v has at least $\frac{d}{2}$ neighbours in $G[A]$, which are all moved to X, as required.

Then, if the algorithm terminated when $|X| \geq \delta v(G)$, it follows that $|C(S)| = |A(I) \cup S(I)| = |V(G) \setminus X(I)| \leq (1 - \delta)v(G)$ for each S.

Conversely, if we stopped when $A \cap I = \emptyset$ then $C(S) = S(I) = I$ and by assumption $|S(I)| \leq \lceil \tau v(G) \rceil \leq (1 - \delta)v(G)$. □

The graph container lemma is originally due to Kleitman and Winston and it, as well as variants of its algorithmic proof already have a large number of interesting applications.

We can now use Lemma 5.10 in order to prove Lemma 5.1. The approach is similar, but instead of removing the neighbourhood of a vertex v when it is placed in S, we will need to keep track of the graph which is induced by edges containing v. If this graph ever gets too large, we can use Lemma 5.10 to find a container for I in this graph, otherwise we find as before that we exclude sufficiently many vertices from I in each step that we find a suitable container before the fingerprint gets too large.

Proof of Lemma 5.1. Given $I \in \mathcal{I}(\mathcal{H})$ we will again algorithmically determine a fingerprint $S(I)$. In order to do so we will maintain a set S, a 3-uniform hypergraph $\mathcal{A}$ of 'available' edges of $\mathcal{H}$ and a graph G of 'forbidden' pairs in $V(\mathcal{H})$. We start with $\mathcal{A} = \mathcal{H}$ and $S = E(G) = \emptyset$. As long as $|S| < \lfloor \frac{1}{2\sqrt{d}} v(\mathcal{H}) \rfloor$ and $V(\mathcal{A}) \cap I \neq \emptyset$ we repeat the following steps:

1. Let u be the first vertex of I in the max-degree order on $\mathcal{A}$.

2. Move u into S.
3. Move the edges $N(u) = \{vw \colon uvw \in E(\mathcal{A})\}$ into G.
4. Remove u from $V(\mathcal{A})$, and also all the vertices preceding u in the max-degree order on $\mathcal{A}$.
5. Remove from $V(\mathcal{A})$ every vertex whose degree in G is larger than $c\sqrt{d}$.
6. Remove from $E(\mathcal{A})$ every edge which contains an edge of G.

Where the max-degree order on a hypergraph is defined in the obvious way. The algorithm terminates when either $|S| \geq \lfloor \frac{1}{2\sqrt{d}} \rfloor v(\mathcal{H})$ or $V(\mathcal{A}) \cap I \neq \emptyset$ and we set $\mathcal{A}(I) = \mathcal{A}$, $S_2(I) = S$ and $G(I) = G$.

The reason we need to remove the vertices of high degree in G from $V(\mathcal{A})$ is we want to be able to apply the graph container lemma to $G(I)$, and so we will need that $\Delta(G(I))$ can be bounded above.

As before, we note that for any two $I, I' \in \mathcal{I}(G)$ if $S_2(I) = S_2(I')$ then also $G(I) = G(I')$ and $\mathcal{A}(I) = \mathcal{A}(I')$. Hence, if S is such that there exists an $I \in \mathcal{I}(G)$ with $S_2(I) = I$ we can unambiguously define $G(S) = G(I)$ and $\mathcal{A}(S) = \mathcal{A}(I)$.

Let us fix $I \in \mathcal{I}(\mathcal{H})$ and let $S_2 = S_2(I)$. Suppose first that

$$e(G(S_2)) \geq \frac{\sqrt{d}v(\mathcal{H})}{32c} \quad \text{and} \quad \Delta(G(S_2)) \leq 2c\sqrt{d}.$$

So, $G(S_2)$ has average degree $\overline{d} \geq \frac{3\sqrt{d}}{32c}$ and maximum degree $\Delta \leq \frac{64}{3}c^2\overline{d}$. Note that I is independent in $G(S_2)$, so we can apply the graph container algorithm to I, giving us some $\delta > 0$ such that we obtain a fingerprint $S_1(I)$ with $|S_1| < \frac{v(\mathcal{H})}{\sqrt{d}}$, and a container C for I with $|C| \leq (1-\delta)v(\mathcal{H})$, where C is a function only of $S_1 \cup S_2$.

We may assume then that either

$$e(G(S_2)) < \frac{\sqrt{d}v(\mathcal{H})}{32c} \quad \text{or} \quad \Delta(G(S_2)) > 2c\sqrt{d}.$$

Note however, that the second cannot hold, since the degree of a vertex in G increases by at most $\Delta(N(u)) \leq \Delta_2(\mathcal{H}) \leq c\sqrt{d}$ in each step of the algorithm, and once a vertex of G has degree larger than $c\sqrt{d}$ we remove it from $V(\mathcal{A})$ (and hence no more edges incident to it are added to G later in the algorithm). Let Y be the set of vertices we removed for this reason.

Note that at any stage of the algorithm $|Y|$ must be small. Indeed, if at any point $|Y| \geq \frac{v(\mathcal{H})}{16c}$ then

$$e(G(S_2)) \geq \frac{|Y|c\sqrt{d}}{2} \geq \frac{\sqrt{d}v(\mathcal{H})}{32} \geq \frac{\sqrt{d}v(\mathcal{H})}{8c},$$

since we may assume that $c > 4$, contradicting our assumption.

We would like to show that we remove many vertices from $\mathcal{A}$ which are not in Y during the algorithm, so that we can take $C(I) := V(\mathcal{A}) \cup S_2 \cup Y$, however if our algorithm doesn't run for very long before terminating then this won't be true. So, if the algorithm terminates because $|S| \geq \frac{v(\mathcal{H})}{2\sqrt{d}}$ we will take $C(I) = C(S_2) := V(\mathcal{A}) \cup S_2 \cup Y$. Conversely if the algorithm

terminates because $V(\mathcal{A}) \cap I = \emptyset$ we will take $C(I) = C(S_2) := S_2 \cup Y$. Let us show that these are appropriate choices for C.

Clearly if we terminate because $V(\mathcal{A}) \cap I = \emptyset$ then $|S_2| \leq \frac{v(\mathcal{H})}{2\sqrt{d}}$ and by the above comment $|Y| \leq \frac{\sqrt{d}v(\mathcal{H})}{8c}$ and hence in this case

$$|C(S_2)| \leq \frac{v(\mathcal{H})}{2\sqrt{d}} + \frac{v(\mathcal{H})}{16c} \leq (1-\delta)v(\mathcal{H}).$$

Furthermore it is clear that in this case $S_2(I) \subseteq I \subseteq Y \cup S_2$.

So, we may assume that our algorithm runs until $|S| \geq \frac{v(\mathcal{H})}{2\sqrt{d}}$. We claim that at some point in the algorithm at least $\frac{v(\mathcal{H})}{16c}$ vertices other than those in Y have been removed from $\mathcal{A}$. Since the vertices we remove from $V(\mathcal{A})$ are either in S, in Y or not in I, it follows that

$$|V(\mathcal{A}) \cup S_2 \cup Y| \leq V(\mathcal{H}) - \frac{v(\mathcal{H})}{16c} + |S_2| \leq v(\mathcal{H})\left(1 - \frac{1}{16c} + \sqrt{\delta}\right) \leq (1-\delta)v(\mathcal{H}).$$

Furthermore it is again clear that $S_2(I) \subseteq I \subseteq C(I) = V(\mathcal{A}) \cup S_2 \cup Y$.

So, let us assume for contradiction that at each stage in the algorithm at most $\frac{v(\mathcal{H})}{16c}$ vertices other than those in Y have been removed from $\mathcal{A}$. We claim that under all these assumptions we add at least $\frac{d}{4}$ edges to G in each step, which will eventually contradict our assumption that $e(G(S_2)) < \frac{\sqrt{d}v(\mathcal{H})}{32c}$.

Indeed, at each step in the algorithm we may assume that $e(G) < \frac{\sqrt{d}v(\mathcal{H})}{32c}$, $|Y| < \frac{v(\mathcal{H})}{16c}$ and at most $\frac{v(\mathcal{H})}{16c}$ other vertices have been removed from $\mathcal{A}$. In this case, since $\Delta_1(\mathcal{H}) \leq cd$ and $\Delta_2(\mathcal{H}) \leq c\sqrt{d}$, we have

$$\begin{aligned} e(\mathcal{A}) &\geq e(\mathcal{H}) - \frac{v(\mathcal{H})\Delta_1(\mathcal{H})}{16c} - |Y|\Delta_1(\mathcal{H}) - e(G)\Delta_2(\mathcal{H}) \\ &\geq e(\mathcal{H}) - \frac{v(\mathcal{H})d}{8} - \frac{d}{32}v(\mathcal{H}) \\ &\geq \frac{v(\mathcal{H})d}{6}. \end{aligned}$$

Now, if we remove a vertex u at this step in the algorithm at most $\frac{v(\mathcal{H})}{16c}$ vertices lie before it in the max-degree order on $\mathcal{A}$ and so the degree of u must be at least $\frac{d}{4}$, otherwise

$$e(\mathcal{A}) \leq \frac{1}{3}\left(\frac{v(\mathcal{H})}{16c}\Delta_1(\mathcal{H}) + v(\mathcal{H})\frac{d}{4}\right) < \frac{v(\mathcal{H})d}{6}.$$

Hence, the vertex u that we remove in the next step will have degree at least $\frac{d}{4}$, and so will add at least $\frac{d}{4}$ edges to G in this step.

However, since the algorithm will terminated when $|S| \geq \frac{v(\mathcal{H})}{2\sqrt{d}}$ it follows that

$$e(G(S_2)) \geq \frac{1}{2}\frac{v(\mathcal{H})}{2\sqrt{d}}\frac{d}{4} \geq \frac{v(\mathcal{H})\sqrt{d}}{16} \geq \frac{v(\mathcal{H})\sqrt{d}}{32c},$$

contradicting our assumption. □

We note that actually we prove a slightly stronger statement than claimed in Lemma 5.1

Lemma 5.11. *For every $c > 0$ there exists $\delta > 0$ such that if $\mathcal{H}$ is a 3-uniform hypergraph with average degree $d \geq \delta^{-1}$ and*

$$\Delta_1(\mathcal{H}) \leq cd \quad \text{and} \quad \Delta_2(\mathcal{H}) \leq c\sqrt{d},$$

then there exists a collection $\mathcal{C}$ of subsets of $V(\mathcal{H})$ and a function $f : 2^{V(\mathcal{H})} \to \mathcal{C}$ such that such that

(a) for every $I \in \mathcal{I}(H)$, there exists $S \subseteq I$ with $|S| \leq \frac{v(\mathcal{H})}{\sqrt{d}}$ and $I \subseteq f(S)$; and

(b) $|C| \leq (1-\delta)v(\mathcal{H})$ for every $C \in \mathcal{C}$.

Let us show how we can use this to remove the unnecessary logarithmic terms in our previous proofs. Firstly we can use it, in precisely the same way as before to show the following strengthened container theorem for triangle-free graphs.

Theorem 5.12. *For each $\varepsilon > 0$ there exist $C > 0$ such that for each $n \in \mathbb{N}$ there exists a collection $\mathcal{G}_n$ of graphs on n vertices and a function $f : 2^{E(K_n)} \to \mathcal{G}_n$ such that*

(a) For every triangle-free graph H on n vertices there is a subgraph $S \subseteq H$ with

$$e(S) \leq Cn^{\frac{3}{2}} \text{ and } H \subseteq f(S); \text{ and}$$

(b) Each $G \in \mathcal{G}_n$ contains at most εn^3 triangles.

Using this we can give a slight improvement to our proofs of Theorems 5.4, 5.6 and 5.8. We will just show the first and leave the others as exercises.

Proof of Theorem 5.4. We us apply Theorem 5.12, with a small $\gamma > 0$ which we will choose later, to get a family of containers $\mathcal{G}_n$. Then for any triangle-free H there is some $S \subseteq H$ with $e(S) \leq Cn^{\frac{3}{2}}$ and a graph $f(S) \in \mathcal{G}_n$ such that $H \subseteq f(S)$. Since $f(S)$ contains at most γn^3 triangles, it follow from Lemma 5.5 that if γ is small enough then $e(f(S)) \leq \left(\frac{1}{2} + \frac{\alpha}{2}\right)\binom{n}{2}$.

Let $\mathcal{S}$ be the set of fingerprints S obtained from the theorem. If there is some triangle-free $H \subseteq G(n,p)$ with $e(H) \geq \left(\frac{1}{2} + \alpha\right) p\binom{n}{2}$ then there is some $S \in \mathcal{S}$ such that $S \subseteq H \subseteq f(S)$, and so $G(n,p)$ contains at least $e(H) - e(S)$ many edges of $f(S) - S$.

However, for any fixed S the probability that this happens is at most

$$\begin{aligned}\mathbb{P}\left(\mathrm{Bin}\left(e(f(S)),p\right) \geq e(H) - e(S)\right) &\leq \mathbb{P}\left(\mathrm{Bin}\left(\left(\frac{1}{2}+\frac{\alpha}{2}\right)\binom{n}{2}, p\right) \geq \left(\frac{1}{2}+\alpha\right)p\binom{n}{2} - Cn^{\frac{3}{2}}\right) \\ &\leq e^{-\Omega(pn^2)}.\end{aligned}$$

So, if we let Y be the number of $S \in \mathcal{S}$ such that $S \subseteq G(n,p)$ and $G(n,p)$ contains at least $m - e(S)$ edges of $f(S) \setminus E(S)$, then

$$\begin{aligned}\mathbb{E}(Y) &\leq \sum_{S \in \mathcal{S}} p^{e(S)} e^{-\Omega(pn^2)} \\ &\leq \sum_{k=1}^{Cn^{\frac{3}{2}}} p^k \binom{\binom{n}{2}}{k} e^{-\Omega(pn^2)} \\ &\leq \sum_{k=1}^{Cn^{\frac{3}{2}}} \left(\frac{en^2p}{2k}\right)^k e^{-\Omega(pn^2)} \\ &= o(1).\end{aligned}$$

since $p = \omega\left(\frac{1}{\sqrt{n}}\right)$. □

5.4 A general container lemma

These methods extend to more general hypergraphs, and the following container lemma for k-uniform hypergraphs was proven independently by Balogh, Morris and Samotij, and Saxton and Thomason.

Lemma 5.13. *For every $k \in \mathbb{N}$ and $c > 0$ there exists $\delta > 0$ such that if $\mathcal{H}$ is a k-uniform hypergraph and $\tau \in (0,1)$ is such that*

$$\Delta_\ell(\mathcal{H}) \leq c\tau^{\ell-1}\frac{e(\mathcal{H})}{v(\mathcal{H})}$$

for every $1 \leq \ell \leq k$, then there exists a collection $\mathcal{C}$ of subsets of $V(\mathcal{H})$ and a function $f : 2^{V(\mathcal{H})} \to \mathcal{C}$ such that

(a) for every $I \in \mathcal{I}(H)$, there exists $S \subseteq I$ with $|S| \leq \tau v(\mathcal{H})$ and $I \subseteq f(S)$; and

(b) $|C| \leq (1-\delta)v(\mathcal{H})$ for every $C \in \mathcal{C}$.

The proof is similar to that for the 3-uniform case, but the details are much more technical. Using similar ideas to the previous section we can extend some of the ideas from the previous section to H-free graphs.

Namely, if we look at the hypergraph which encodes the edge sets of copies of H in $E(K_n)$, then one can check that it satisfies the bounded degree condition in Lemma 5.13 with $\tau = n^{-\frac{1}{m_2(H)}}$ where m_2 is the 2-*density* of H,

$$m_2(H) = \max\left\{\frac{e(F)-1}{v(F)-2} \colon F \subseteq H, v(F) \geq 3\right\}.$$

Then, using Lemma 5.13 one can prove a container theorem for H-free graphs in the vein of Theorem 5.12. Combining this with a suitable supersaturation result it is possible to give the following sparse random version of the Erdős-Stone-Simonivits theorem

Theorem 5.14. *For every graph H with $\Delta(H) \geq 2$ and every $\delta > 0$ there exists $C > 0$ such that if $p \geq Cn^{-\frac{1}{m_2(H)}}$, then whp every subgraph $G \subseteq G(n,p)$ with*

$$e(G) \geq \left(1 - \frac{1}{\chi(H)-1} + \delta\right) p\binom{n}{2}$$

contains H as a subgraph.

However, since the ideas are very similar to the previous section, we will instead give a different application of the general hypergraph container lemma, to finding arithmetic progression in random susbets of sparse subsets of the integers.

A particularly well-known result from extremal combinatorics is Szemerédi's Theorem on arithmetic progressions.

Theorem 5.15. *For every integer $k \geq 3$ and $\delta > 0$ there is an $n_0 \in \mathbb{N}$ such that if $n \geq n_0$ and $A \subseteq [n]$ is such that $|A| \geq \delta n$ then A contains a k-term arithmetic progression.*

Given $p \in [0,1]$ let us denote by $[n]_p$ the random subsets of $[n]$ given by including each integer independently with probability p. If p is very small, we shouldn't expect $[n]_p$ to contain any arithmetic progressions of length k, indeed the expected number of such progressions can be bounded above by $p^k n^2$. However, if this is much smaller than pn then by the alteration method we can easily find a subset of $[n]_p$ of very large size (in terms of pn) which doesn't contain any k-term arithmetic progression. Hence we cannot have any hope of finding a sparse random version of Szemerédi's theorem if $n^2 p^k \ll np$, or in other words we need p to be at least around $n^{-\frac{1}{k-1}}$. This however is the only obstruction, as shown by the following theorem, independently proven by Conlon and Gowers, and Schacht.

Theorem 5.16. *For every integer $k \geq$ and $\delta > 0$ there exists $C > 0$ such that if $p \geq Cn^{-\frac{1}{k-1}}$, then with high probability every subset $A \subseteq [n]_p$ of size $|A| \geq \delta pn$ contains a k-term arithmetic progression.*

We will deduce this theorem as a corollary of the following counting result.

Theorem 5.17. *For every integer $k \geq 3$ and every $\beta > 0$ there exists $C > 0$ and $n_0 \in \mathbb{N}$ such if $m \geq Cn^{1-\frac{1}{k-1}}$ and $n \geq n_0$, then there are at most*

$$\binom{\beta n}{m}$$

many m-subsets of $[n]$ which contain no k-term arithmetic progression.

Assuming this result, Theorem 5.16 follows easily.

Proof of Theorem 5.16. Given $m \in \mathbb{N}$, let Y_m denote the number of m-subsets of $[n]_p$ which contain no k-term arithmetic progression. Let $\beta = \frac{\delta}{e^2}$, and let C be the constant given by Theorem 5.17. If $p \geq \left(\frac{C}{\delta}\right) n^{-\frac{1}{k-1}}$, then $m := \delta np \geq Cn^{1-\frac{1}{k-1}}$ and so it follows from Theorem 5.17 that there are at most

$$\binom{\beta n}{\delta np}$$

many δnp-subsets of $[n]$ which contain no k-term arithmetic progression.

Hence

$$\mathbb{P}(Y_{\delta np} \geq 1) \leq \mathbb{E}(Y_{\delta np}) \leq \binom{\beta n}{\delta np} p^{\delta np} \leq \left(\frac{e\delta np}{e^2 \delta np}\right)^{\delta np} = e^{-m}.$$

In other words, with high probably $[n]_p$ doesn't contain any subsets of size δnp which don't contain any k-term arithmetic progression, and so every subset of size δnp contains a k-term arithmetic progression. □

So, it remains to prove Theorem 5.17. We will want to apply Lemma 5.13 to the k-uniform hypergraph which encodes k-term arithmetic progressions in $[n]$. More precisely we let $\mathcal{H}_k$ be the hypergraph with $V(\mathcal{H}_k) = [n]$ and

$$E(\mathcal{H}_k) = \left\{e \in \binom{[n]}{k} : e = \{a, a+d, \ldots, a+(k-1)d\} \text{ for some } a, d \in [n]\right\}.$$

We note that $e(\mathcal{H}_k) = \Theta(n^2)$, $\Delta_1(\mathcal{H}_k) = O(n)$, and $\Delta_\ell(\mathcal{H}_k) = O(1)$ for any $\ell \geq 2$, where the implicit constants might depend on k. Hence if we let $\tau = n^{-\frac{1}{k-1}}$ then we have that

- $\Delta_1(\mathcal{H}_k) = O(n) \leq c\tau^0 \frac{e(\mathcal{H}_k)}{v(\mathcal{H}_k)} = c\Theta(n)$ if c is sufficiently large;
- $\Delta_\ell(\mathcal{H}_k) = O(1) \leq c\tau^{\ell-1} \frac{e(\mathcal{H}_k)}{v(\mathcal{H}_k)} = cn^{-\frac{\ell-1}{k-1}}\Theta(n)$ for any $2 \leq \ell \leq k$ if c is sufficiently large.

Hence we can apply Lemma 5.13 with $\tau = n^{-1\frac{1}{k-1}}$. The final ingredient to the proof is again a supersaturation result for Szemerédi's Theorem, which can be deduced from the original via a simple averaging argument.

Theorem 5.18. *For every $\varepsilon > 0$ there exists $\delta > 0$ and $n_0 \in \mathbb{B}$ such that if $n \geq n_0$, then every subset $A \subseteq [n]$ with $|A| \geq \varepsilon n$ contains at least δn^2 k-term arithmetic progressions.*

Proof of Theorem 5.17. Let us set $\varepsilon = \frac{\beta}{2}$. We claim that there exist a family $\mathcal{A} \subseteq 2^{[n]}$ and a function $f : 2^{[n]} \to \mathcal{A}$ such that

(a) Each $A \in \mathcal{A}$ is of size at most εn; and

(b) For every set $B \subseteq [n]$ which does not contain a k-term arithmetic progression there exists a subset $S \subseteq B$ with
$$|S| = O\left(n^{1-\frac{1}{k-1}}\right) \quad \text{and} \quad B \subseteq f(S).$$

Indeed, let us apply Lemma 5.13 to $\mathcal{H}_k$ with $\tau = n^{-\frac{1}{k-1}}$ with a suitable value of c. We obtain a family of fingerprints and containers such that for each B which doesn't contain a k-term arithmetic progression we have $S_1 = S_1(B) \subseteq B$ and $C_1 = C_1(S) \supseteq B$ with $|S_1| \leq n^{1-\frac{1}{k-1}}$ and $|C_1| \leq (1-\delta)n$.

We then iterate: supposing we have such sets S_t and C_t for each B with $|S_t| \leq tn^{1-\frac{1}{k-1}}$ and $|C_t| \leq (1-\delta)^t n$. If $|C_t| \leq \varepsilon n$, then we place C_t in $\mathcal{A}$ and set $f(S_t) = C_t$, otherwise we apply Lemma 5.13 to $\mathcal{H}_k[C_t]$ with $\tau = n^{-\frac{1}{k-1}}$ and a suitable value of $c = c(k, \varepsilon)$.

We can do this because $e(\mathcal{H}_k[C_t]) = \Theta(n^2)$ by Theorem 5.18, since $|C_t| > \varepsilon n$ by assumption, and so our previous calculations show the assumptions of Lemma 5.13 are satisfied if c is sufficiently large.

We obtain, for every subset B of C_t which doesn't contain a k-term arithmetic progression a fingerprint $S'_{t+1} \subseteq B$ and a container $C_{t+1} \supseteq B$ with $|C_{t+1}| \leq (1-\delta)|C_t| \leq (1-\delta)^{t+1} n$. Setting $S_{t+1} = S_t \cup S'_{t+1} \subseteq B$ we have that $|S_{t+1}| \leq |S_t| + n^{1-\frac{1}{k-1}} \leq (t+1)n^{1-\frac{1}{k-1}}$.

Since δ only depends on c, ε and k after a constant number of steps we have that $(1-\delta)^t \leq \varepsilon$ and so $|C_t| \leq \varepsilon n$. This gives us the family claimed.

Let $\mathcal{S}$ be the collection of fingerprints S for each k-term arithmetic progression free set B, and let C be a sufficiently large constant so that every $S \in \mathcal{S}$ has size at most $\varepsilon C n^{1-\frac{1}{k-1}}$. Then, for every $m \geq Cn^{1-\frac{1}{k-1}}$ the number of subsets $\subseteq [n]$ of size m containing no arithmetic progression is at most

$$\begin{aligned}\sum_{S \in \mathcal{S}} \binom{|f(S)|}{m-|S|} &\leq \sum_{s \leq \varepsilon m} \binom{n}{s}\binom{\varepsilon n}{m-s} \\ &\leq \sum_{s \leq \varepsilon m} \left(\frac{en}{s}\right)^s \left(\frac{m}{\varepsilon n - m}\right)^s \binom{\varepsilon n}{m} \\ &\leq \sum_{s \leq \varepsilon m} \left(\frac{2em}{\varepsilon s}\right)^s \binom{\varepsilon n}{m}.\end{aligned}$$

Since by Szemerédi's theorem we may assume that $m = o(n)$. Then, since the function $x \mapsto \left(\frac{y}{x}\right)^x$ is increasing on $(0, \frac{y}{e})$, it follows that the right hand side is at most

$$\sum_{s \leq \varepsilon m} \left(\frac{2em}{\varepsilon s}\right)^s \binom{\varepsilon n}{m} \leq m \left(\frac{2e}{\varepsilon^2}\right)^{\varepsilon m} \binom{\varepsilon n}{m} \leq \binom{\beta n}{m}$$

where the final inequality follows since $\beta = 2\varepsilon$, and so

$$\binom{\varepsilon n}{m} \leq 2^{-m} \binom{\beta n}{m}.$$

□

6 Talagrand's Inequality

Other concentration equalities that we have considered, like the Azuma-Hoeffding inequality, gave us an exponentially small bound on the probability of deviations from the mean of suitably well behaved random variables. Our notion of suitably well behaved was basically that it was close, or bounded by, some random variable on a product space where in each co-ordinate the random variable was bounded. However the Azuma-Hoeffding inequality required that the deviations we considered be at least as large as the square root of the dimension of this product space. In many cases the random variables we are interested in will have large expectation, so we might hope to show that they are concentrated about their mean, but this expectation is much smaller than the dimension of the product space that they live in, and so the standard concentration inequalities do not give an effective bound on the tail probabilities. In this case, there is an inequality of Talagrand that can be useful for proving concentration.

We will state just a special case of the equality, without proof, in a form which is useful for applications. Let us first briefly formally introduce the notion of a product space, which we will do only for finite probability spaces for simplicity. The case for general probability spaces is similar.

Definition. Let $\{(\Omega_i, 2^{\Omega_i}, \mathbb{P}_i) : i \in [n]\}$ be finite probability spaces. We let

$$\Omega = \prod_{i=1}^{n} \Omega_i = \{(\omega_1, \omega_2, \ldots, \omega_i) : \omega_i \in \Omega_i \text{ for all } i \in [n]\}$$

be the product of the sets Ω_i and define a probability measure $\mathbb{P}$ on 2^Ω by defining the probability of elementary events to be

$$\mathbb{P}((\omega_1, \omega_2, \ldots, \omega_i)) = \prod_{i=1}^{n} \mathbb{P}_i(\omega_i)$$

and extending it to 2^Ω in the obvious way. Then the *product space* (of $\{(\Omega_i, \Sigma_i, \mathbb{P}_i) : i \in [n]\}$) is the probability space $(\Omega, 2^\Omega, \mathbb{P})$.

We note that, given a random variable on a product space, there is a a natural martingale associated with this random variable given by 'exposing' each co-ordinate in turn, the edge and vertex exposure martingales being examples of this.

Talagrand's inequality is a very general inequality about the concentration of measure in product spaces. We consider a generalisation of the *Hamming distance*. This counts the number of co-ordinates in which two $\omega, \omega' \in \Omega$ differ. In other words, the distance is $\sum_{i:\, w_i \neq w_i'} 1$. We can take a weighted version of this and consider, for any unit vector $\alpha \in \mathbb{R}^n$ the *α-Hamming distance* between ω and ω' to be

$$d_\alpha(\omega, \omega') = \sum_{i:\, \omega_i \neq \omega_i'} \alpha_i.$$

Given a set $A \subset \Omega$ and a point ω for any α we can consider the α-Hamming distance between ω and A.

$$d_\alpha(\omega, A) = \inf\{d(\omega, \omega') : \omega' \in A\}.$$

We will think of ω as being far from A if it's far in *some* α-Hamming distance, with α a unit vector. That is, we define

$$d(\omega, A) = \sup_{|\alpha|=1} d_\alpha(\omega, A).$$

Talagrand's inequality is then the following

Theorem 6.1 (Talagrand's Inequality)**.** *Let $\{(\Omega_i, \Sigma_i, \mathbb{P}_i) : i \in [n]\}$ be probability spaces and let $(\Omega, \Sigma, \mathbb{P})$ be their product. If $A, B \in \Sigma$ are such that $d(\omega, A) \geq \tau$ for all $\omega \in B$, then*

$$\mathbb{P}(A)\mathbb{P}(B) \leq e^{-\frac{\tau^2}{4}}.$$

Note that we can vary α for different points $\omega \in B$, so it doesn't even have to be 'uniformly' far from A. An equivalent formulation is to say that we are bounding $\mathbb{P}(A)\mathbb{P}(A_t)$ where $A_\tau = \{\omega : d(\omega, A) \geq \tau\}$.

This doesn't look a lot like the tail estimates we have from earlier in the course, however with not too much work one can deduce the following as a corollary, which we may sometimes refer to also as Talagrand's inequality.

First let us make a few definitions

Definition. A random variable $X : \Omega \to \mathbb{R}$ is *c-Lipschitz* if changing just one co-ordinate can change the value of X by at most c. Given some function $f : \mathbb{N} \to \mathbb{N}$ we say that X *is f-certifiable* if whenever $X(\omega_1, \omega_2, \ldots, \omega_n) \geq s$ there is a subset $I \subset [n]$ of size $|I| = f(s)$ such that X is greater than s on the entire subspace

$$\{(\omega'_1, \omega'_2, \ldots, \omega'_n) : \omega'_i = \omega_i \text{ for all } i \in I\}.$$

To put it in words, X is f-certifiable if, whenever X takes a value bigger than s, you can verify this by looking at just $f(s)$ of it's co-ordinates. Talagrand's inequality tells us that, when a c-Lipschitz random variable is f-certifiable for a suitably small f, then it is highly concentrated about its median.

Corollary 6.2. *Let X be a c-Lipschitz random variable which is f-certifiable and let m be the median of X (that is m is the unique real number such that $\mathbb{P}(X > m) \leq \frac{1}{2}$ and $\mathbb{P}(X < m) \leq \frac{1}{2}$). Then for any $t \geq 0$*

$$\mathbb{P}(X \leq m - t) \leq 2e^{-\frac{t^2}{4c^2 f(m)}} \text{ and } \mathbb{P}(X \geq m + t) \leq 2e^{-\frac{t^2}{4c^2 f(m+t)}}.$$

Proof. Let us consider the two sets

$$A = \{\omega : X(\omega) \leq m - t\} \text{ and } B = \{\omega : X(\omega) \geq m\}$$

Since $\mathbb{P}(B) \geq 1/2$ by definition, if we can show that $d(\omega, A) \geq \tau$ for all $\omega \in B$ for some τ, then we can use Theorem 6.1 to get a bound on the probability of A.

However, since X is f-certifiable, for every $\omega \in B$ there is some set $I \subset [n]$ of size $f(m)$ such that $X(\omega') \geq m$ for every ω' which agrees with ω on I. Then, since every $\omega' \in A$ has $X(\omega) \leq m - t$ and changing the value of one co-ordinate can change the value of X by at most c, it follows that every $\omega' \in A$ must disagree with ω in at least t/c of the co-ordinates in I.

Hence, if we take α to be the unit vector with $\alpha_i = 1/\sqrt{|I|}$ for all $i \in I$ and 0 otherwise, it follows that every ω has α-Hamming distance at least $t/c\sqrt{f(m)}$ to every $\omega' \in A$. Hence, $d(\omega, A) \geq t/c\sqrt{f(m)}$.

Hence, applying Theorem 6.1 tells us that

$$\mathbb{P}(X \leq m - t)\mathbb{P}(X \geq m) \leq e^{-\frac{t^2}{4c^2 f(m)}}$$

and so, since by definition $\mathbb{P}(X \geq m) \geq 1/2$ we have that

$$\mathbb{P}(X \leq m - t) \leq 2e^{-\frac{t^2}{4c^2 f(m)}}.$$

The other inequality follows in a similar fashion. □

Note that the two tail estimates are not necessarily symmetric. Looking at Corollary 6.2, we see that we can get exponentially good bounds for the tail estimates when $t >> \sqrt{f(m)}$. Most importantly this doesn't necessarily depend on the dimension of the product space we live in, and so, when $f(m)$ is small compared to n, we will get much better bounds than the Azuma-Hoeffding inequality would give us.

Also, this theorem talks about a variable being concentrated about its median rather than its mean, however, as the following lemma shows, in a lot of cases one can show the median must be close to the expectation.

Lemma 6.3. *Let $\{(\Omega_i, 2^{\Omega_i}, \mathbb{P}_i) : i \in [n]\}$ be probability spaces and let $(\Omega, 2^{\Omega}, \mathbb{P})$ be their product. Let X be a c-Lipschitz, f-certifiable random variable with $f(s) = rs$ and let m be the median of X. Then*

$$|\mathbb{E}(X) - m| \leq 20c\sqrt{rm}.$$

Proof. Clearly we have that

$$|\mathbb{E}(X) - m| \leq \mathbb{E}(|X - m|)$$

Let us split the possible values of $|X - m|$ into intervals of length $c\sqrt{rm}$. It follows that

$$\mathbb{E}(|X - m|) \leq \sum_{k=0}^{\infty} c\sqrt{rm}(k+1)\mathbb{P}(|X - m| \geq kc\sqrt{rm}).$$

We can however apply Corollary 6.2 to each summand, with $t = kc\sqrt{rm}$ to see that, since $t \leq m$

$$\mathbb{P}(X \leq m - kc\sqrt{rm}) \leq 2e^{-\frac{k^2}{4}}$$

and

$$\mathbb{P}(X \geq m + kc\sqrt{rm}) \leq 2e^{-\frac{k^2 m}{4(m+t)}} \leq 2e^{-\frac{k^2}{8}}$$

and so

$$\leq 40c\sqrt{rm}$$

Since $\sum_{k=0}^{\infty}(k+1)e^{-\frac{k^2}{8}} < 10$. □

6.1 Longest Increasing Subsequence

Suppose we pick a sequence $x_1, x_2, \ldots, x_n \in [0,1]$ independently and uniformly at random. If we put the sequence in increasing order, $x_{i_1} < x_{i_2} < \ldots < x_{i_n}$, this defines a permutation of $[n]$, and it is not hard to check that the distribution we get by picking a permutation in this way is also uniform.

Let us consider the random variable X which counts the longest increasing subsequence from this sequence. What can we say about this random variable?

Well, given some ordered subset of the x_i, $\{x_{i_1}, x_{i_2}, \ldots x_{i_k}\}$, the probability that this forms an ordered subsequence is simply $1/k!$ by symmetry. So the expected number of increasing sequences of length k is just $\binom{n}{k}/k!$ and so by Markov's inequality , if we let X_k be the number of increasing sequences of length k, we have that

$$\mathbb{P}(X \geq k) = \mathbb{P}(X_k \geq 1) \leq \mathbb{E}(X_k) = \frac{\binom{n}{k}}{k!} \leq \left(\frac{en}{k}\right)^k \left(\frac{e}{k}\right)^k = \left(\frac{e\sqrt{n}}{k}\right)^{2k}.$$

Therefore if the median is the number m such that $\mathbb{P}(X \geq m) = 1/2$, we have that $m \leq 3\sqrt{n}$, since we know that

$$\mathbb{P}(X \geq 3\sqrt{n}) \leq \left(\frac{e}{3}\right)^{6\sqrt{n}} < 1/2.$$

However, a classical theorem of Erdős and Szekeres tells us that if we let Y be the length of the largest decreasing subsequence then we always have that $XY \geq n$ (we have to be a little careful since this talks about non-decreasing and non-increasing subsequences, but with high probability no two x_i take the same value). However, by symmetry we have that X and Y have the same distribution and so, since $\mathbb{P}(X \leq 3\sqrt{n}) \geq 1/2$ we must have that $\mathbb{P}(Y \geq \frac{1}{3}\sqrt{n}) \geq 1/2$ and so the median of X must satisfy

$$\frac{1}{3}\sqrt{n} \leq m \leq 3\sqrt{n}.$$

Now, the random variable X is clearly 1-Lipschitz, changing the value of any one x_i can only change the length of the largest increasing subsequence by 1, and also we have that X is f-certifiable with $f(s) = s$. Indeed, to verify that $X \geq s$ we can simply look at the values of the x_i in an increasing subsequence of length s. Therefore by Theorem 6.1 we have

$$\mathbb{P}(X \leq m - t) \leq 2e^{-\frac{t^2}{4m}} \text{ and} \mathbb{P}(X \geq m + t) \leq 2e^{-\frac{t^2}{4(m+t)}}.$$

Since $m \sim \sqrt{n}$ we can take t to be slighty larger than $n^{1/4}$, say $t = n^{1/4}\log(n)$, to see that with high probability X must lie in the interval $[m - t, m + t]$.

Let us compare this to the bound we would get from applying Azuma-Hoeffding. Here we would be considering the exposure martingale associated with X by exposing the values of each x_i in turn. Since the function is 1-Lipschitz, and the values of the x_i are independent, it follows

by a standard argument that the associated martingale satisfies $|X_i - X_{i-1}| \leq 1$ and so by the Azuma-Hoeffding inequality we have that

$$\mathbb{P}(X \geq \mathbb{E}(X) + t) \leq e^{-\frac{t^2}{2n}} \text{ and } \mathbb{P}(X \leq \mathbb{E}(X) - t) \leq e^{-\frac{t^2}{n}}.$$

So in order to get tight concentration we would need to take $t >> n^{1/2}$, which is not only much worse than what Talagrand can give, it's not especially useful since it follows from Lemma 6.3 that $|\mathbb{E}(X) - m| = O(n^{\frac{1}{4}})$ and so $\mathbb{E}(X) \sim m \sim \sqrt{n}$.

6.2 Chromatic Number of Graph Powers

Definition. Given a graph G and $x, y \in V(G)$ let us define the *distance* between x and y, $\text{dist}_G(x, y)$, to be the length of the shortest path between them. Given a graph G the *kth power of G*, G^k is defined to be the graph with

$$V(G^k) = V(G) \text{ and } E(G^k) = \{(x, y) : \text{dist}_G(x, y) \leq k\}.$$

Suppose we have a graph G with maximum degree $\Delta(G) = d$. What can we say about the chromatic number of G^k?

A simple application of the greedy algorithm tells us that we can colour any graph H with $\Delta(H) + 1$ colours, and since $\Delta(G^k) \leq d^k$ we have that $\chi(G^k) \leq d^k + 1$. Brook's theorem tells us that for any graph which is not a cycle or complete, $\chi(H)$ beats this naive bound and $\chi(H) \leq \Delta(H)$. This result has been improved, first by Kim to show that

Theorem 6.4. *Let H be such that $g(H) \geq 5$, then*

$$\chi(H) \leq (1 + o(1))\frac{\Delta(H)}{\log(\Delta(H))}.$$

and later Johansson showed

Theorem 6.5. *Let H be such that $g(H) \geq 4$ (that is, triangle-free), then*

$$\chi(H) \leq O\left(\frac{\Delta(H)}{\log(\Delta(H))}\right).$$

Applying this to graph powers we see that, as long as G still has reasonably large girth compared to k, then we get an improvement of a log factor over the naive bound for G^k as well. The following result of Alon and Mohar tells us that if we fix $g \geq 3$ and k, and allow the maximum degree to be arbitrarily large, then there exist graphs achieving this bound.

Theorem 6.6. *Let $g \geq 3$ and k be fixed. Then for large enough d there exist graphs G with $g(G) \geq g$ and $\Delta(G) \leq d$ such that*

$$\chi(G^k) \geq \Omega\left(\frac{d^k}{\log d}\right).$$

Proof. We want to construct such a graph by picking a random graph $G(n,p)$ for suitable n and p. Let $p = \frac{d}{2n}$ so that the expected degree of each vertex is $\sim d/2$. We first want to make sure that our graph satisfies the conditions claimed on $\Delta(G)$ and $g(G)$, to do that we will use the alteration method.

It is a simple application of the chernoff bound that, for each vertex $v \in V$

$$\mathbb{P}(d(v) \geq d) < e^{-\frac{d}{100}}.$$

(where no attempt has been made to optimise the constant, none at all). So, if we let N_{bad} be the number of vertices with degree larger than d we see that

$$\mathbb{E}(N_{\text{bad}}) < ne^{-\frac{d}{100}}.$$

Hence, by Markov's inequality

$$\mathbb{P}(N_{\text{bad}} > 100ne^{-\frac{d}{100}}) \leq \frac{1}{100}.$$

Similarly if we look at the random variable $C_{<g}$ which counts the number of cycles of size $< g$, we have that

$$\mathbb{E}(C_{<g}) = \sum_{i=1}^{g-1}(np)^i = \sum_{i=1}^{g-1}(\frac{d}{2})^i < d^g.$$

So again an application of Markov's inequality tells us that

$$\mathbb{P}(C_{<g} > 100d^g) < \frac{1}{100}.$$

Combining these two estimate we see that with probability at least $98/100$, $G(n,p)$ is such that

$$N_{\text{bad}} < 100ne^{-\frac{d}{6}} \text{ and } C_{<g} < 100d^g$$

and so, since we are free to take $n \gg d \gg g$, we can remove a vertex from each small cycle and delete all vertices of degree more than d to get a graph G' with neither, which still has $(1-o(1))n$ vertices.

We want to get a bound on the chromatic number of G'^k, and we will do so by bounding above the size of the largest independent set in G'^k. That is, since

$$\chi(G'^k) \geq \frac{v(G')}{\alpha(G'^k)} \geq \frac{n}{2\alpha(G'^k)}$$

we will need to show that, with positive probability, even after the alterations we made

$$\alpha(G'^k) \leq c_k n \frac{\log(d)}{d^k}$$

for some constant c_k. For each subset U of that size we want to show that, even after we make our alterations, there is still some edge in G'^k inside U. To guarantee this we will show that with a high probability we can find many vertex disjoint paths of length k between pairs of vertices in U. Since each vertex we removed from the graph $G(n,p)$ can be in at most one of these paths, if there are sufficiently many we can conclude that U is still not independent in G'. So we will prove the following auxiliary lemma.

Lemma 6.7. *Let $G(n,p)$ be chosen with p as above, then for an appropriate choice of constant c_k the following holds: For every subset $U \subseteq V(G)$ of size*

$$|U| = c_k n \frac{\log d}{d^k} = x$$

let P be the random variable which counts the maximum size of a family of paths of length k which lie in $G(n,p)$ such that both endpoints lie in U, all the internal vertices of the paths lie outside of U, and no two paths share a vertex except in U. Then almost surely

$$P \geq \frac{c_k^2 n (\log d)^2}{2^{k+6} d^k}.$$

Proof. If we let P' be the random variable which just counts the number of paths of length k satisfying the first two conditions we see that

$$\begin{aligned}
\mathbb{E}(P') &= \binom{x}{2}(n-x)(n-x-1)\dots(n-x-k+1)p^k \\
&> c_k^2 n^2 \frac{(\log d)^2}{d^{2k}} \frac{n^{k-1}}{4} \frac{d^k}{2^k n^k} \\
&= \frac{c_k^2 n (\log d)^2}{2^{k+2} d^k}
\end{aligned}$$

If we let Q be the random variable which counts the pairs of paths which share an internal vertex outside of U it is rather tedious, but elementary, calculation to show that the largest contribution to Q comes from pairs of paths which share an endpoint and the neighbour of that endpoint. The expected number of such pairs is at most

$$\mathbb{E}(P') n^{k-2} x p^{k-1} = \mathbb{E}(P') \frac{c_k \log d}{2^{k-1} d} << \mathbb{E}(P').$$

Therefore if we pick a random set of such paths, including each one with some fixed probability q, the expected size of our collection is just $q\mathbb{E}(P')$ and the expected number of bad pairs is $q^2\mathbb{E}(Q) < q^2\mathbb{E}(P')$. Therefore by the alteration method we can find a collection such that when we remove 1 path from each bad pair we still have at least $(q-q^2)\mathbb{E}(P')$ left, and so we have that, with $q = 1/2$

$$\mathbb{E}(P) \geq \frac{1}{4}\mathbb{E}(P') > \frac{c_k^2 n (\log d)^2}{2^{k+4} d^k} := b.$$

If we define m to be the median of P we claim that $m \geq b/2$. Indeed, if not, then both m and $20\sqrt{km}$ are less than $b/2$. However, since P counts the size of the largest set of paths of length k satisfying certain conditions, including being edge disjoint, we have that P is 1-Lipschitz, that is, changing any one edge can change P by at most 1. We also note that P is f-certifiable for $f(s) = ks$. Hence, by Lemma 6.3

$$\mathbb{E}(P) \leq m + 20c\sqrt{km} < b$$

a contradiction.

Hence, since $m \geq b/2$ it follows from Talagrands inequality that

$$\mathbb{P}(P \leq b/4) \leq \mathbb{P}(P \leq m/2) \leq 2e^{-\frac{m}{16k}} \leq 2e^{-\frac{b}{32k}}.$$

Therefore,

$$\mathbb{P}(P \leq \frac{b}{4}) \leq 2\exp\left(-\frac{c_k^2 n \log(d)^2}{2^{k+9} k d^k}\right).$$

Therefore, for any fixed U, we've showed that the probability that U has less than $b/4$ appropriate paths (which we note is the claimed number in the statement of the lemma), can be bounded above by this quantity. If we can show that this quantity is small compared to the total number of U of size x, the we could conclude the result of the lemma by the union bound. Now the total number of such sets is

$$\begin{aligned}\binom{n}{x} \leq \left(\frac{en}{x}\right)^x &= \left(\frac{ed^k}{c_k \log d}\right)^{\frac{c_k n \log d}{d^k}} \\ &\leq \exp\left(\frac{c_k k n (\log d)^2}{d^k}\right).\end{aligned}$$

Therefore if we choose c_k such that

$$\frac{c_k^2}{2^{k+9}k} > 2kc_k$$

then with high probability for every such set U there will be at least the claimed number of paths. □

So, as we showed before, with probability at least 98/100, $G(n,p)$ is such that

$$N_{\text{bad}} < 100ne^{-\frac{d}{6}} \text{ and } C_{<g} < 10d^g$$

and we also know that with high probability $G(n,p)$ satisfies the conclusion of Lemma 6.7, and hence with positive probability $G(n,p)$ satisfies all three. Therefore there exists some graph G' satisfying all three conditions. From such a graph, let us delete a vertex from each cycle of length less than g and delete each vertex of degree $\geq d$. We therefore obtain a graph G such that $g(G) \geq g$ and $\Delta(G) \leq d$ and also

$$|G| \geq n - 100(ne^{-\frac{d}{6}} + d^g) \geq \frac{n}{2}.$$

We also claim that $\alpha(G^k) \leq x$. Indeed, since G' satisfied the conclusion of Lemma 6.7 we know that each subset of $V(G)$ of size x, when considered as a subset of $V(G')$, contained at least

$$\frac{c_k n \log(d)^2}{2^{k+6} d^k}$$

paths of length k, in G', between vertices of U such that no internal vertices were inside U, or in more than 1 path. Since we removed at most

$$100(ne^{-\frac{d}{6}} + d^g) < \frac{c_k n \log(d)^2}{2^{k+6} d^k}$$

vertices of G' to form G, at least one of these paths is contained in G, and hence in G^k, U is not independent.

Therefore G is a graph with $g(G) \geq g$ and $\Delta(G) \leq d$ and

$$\chi(G^k) \geq \frac{|G^k|}{\alpha(G^k)} \geq \frac{n}{2x} = \Omega\left(\frac{d^k}{\log(d)}\right).$$

□

7 Resilience

In Section 5 we considered problems of the following type: Given a p and a random graph $G(n,p)$ what is the largest subgraph not containing a triangle? Since adding edges to $G(n,p)$ only hurts this aim, another way to think of this question would be how many edges of $G(n,p)$ do we need to change to make it triangle-free. More generally, given a property $\mathcal{P}$ of graphs and a fixed graph G we can ask how far in *edit distance* is G from $\overline{\mathcal{P}}$, that is, what is the smallest r such that there exists a graph H on the same vertex set of with $e(H) = r$ and $G \triangle H \notin \mathcal{P}$). We call this the *global resilience of G w.r.t $\mathcal{P}$*. If this number is large, then G satisfies $\mathcal{P}$ in some way quite 'robustly'; changing few edges of G will not ruin the property $\mathcal{P}$.

However, for many global graph properties, such as connectedness or Hamiltonian, local changes such as removing all edges adjacent to a fixed vertex can ruin them, giving a trivial upper bound on the resilience of these properties. A more natural notion of 'robustness' to consider seems to be bounding the size of the 'local changes' one cane make. That is, we consider the smallest r such that there exists a graph H on the same vertex set of with $\Delta(H) = r$ and $G \triangle H \notin \mathcal{P}$). We call this the *local resilience of G w.r.t $\mathcal{P}$*.

As an illustrative example, consider the the case $G = K_n$, where $\mathcal{P}$ is an increasing property. In this case the global resilience of K_n w.r.t $\mathcal{P}$ tells us the extremal number of edges in a graph not in $\mathcal{P}$, whereas the local resilience of K_n tells us the largest minimum degree of a graph not in $\mathcal{P}$.

In this section we will present some results about the local resilience of the random graph $G(n,p)$ w.r.t various properties, which we will from now on just call resilience for brevity. With this is mind, given a graph property $\mathcal{P}$ we will write $\Delta_{\mathcal{P}}$ for the resilience of $G(n,p)$ w.r.t $\mathcal{P}$, where we note that $\Delta_{\mathcal{P}}$ is a function of p.

7.1 Perfect Matchings

Recall that the threshold for the existence of a perfect matching in $G(n,p)$ occurs at $\hat{p}(n) = \frac{\log n}{n}$ (assuming n is even, which we will in this section). If we let $\mathcal{M}$ be the property of containing a perfect matching, how resilient will this property be above the threshold?

Well, an obvious upper bound for the resilience comes from the following construction: Divide n into two parts X and Y of size $\frac{n}{2} - 1$ and $\frac{n}{2} + 1$ respectively. If we delete all the edges in $G(n,p)[Y]$ then clearly there can be no perfect matching in the remaining graph. However, by the Chernoff bounds, with high probability we delete $(1 + o(1))\frac{np}{2}$ many edges incident to each vertex and hence the local resilience w.r.t $\mathcal{M}$ is at most $(1 + o(1))\frac{np}{2}$. We will show that this is in fact approximately the correct value.

Theorem 7.1. *Suppose that n is even and that $p = \omega\left(\frac{\log n}{n}\right)$. Then with high probability in $G(n,p)$*

$$\left(\frac{1}{2} - \varepsilon\right) np \leq \Delta_{\mathcal{M}} \leq \left(\frac{1}{2} + \varepsilon\right) np$$

for any constant $\varepsilon > 0$.

Proof. By the comment before the proof it remains to show the lower bound. By monotonicity of $\mathcal{M}$ we may assume that ε is sufficiently small, and furthermore we only need to consider the effect of deleting edges on $\mathcal{M}$. Let H be an arbitrary subgraph of K_n of maximum degree $(\frac{1}{2}-\varepsilon)np$ and let $G := G(n,p) \setminus E(H)$.

We first note that by the Chernoff bounds with high probability the minumum degree of $G(n,p)$ is at least $(1-\delta)np$ for any small $\delta > 0$. Indeed, for each vertex v the probability that $d(v) \leq (1-\delta)np$ is at most

$$e^{-\frac{(\delta np)^2}{2np}} = e^{-\Theta(np)} = e^{-\omega(\log n)} = o(n^{-2})$$

and so by the union bound with high probability every vertex has degree at least $(1-\delta)np$. Hence with high probability the minimum degree of $G = G(n,p)\setminus E(H)$ is at least $(\frac{1}{2}-\delta+\varepsilon)np \geq (\frac{1}{2}+\frac{\varepsilon}{2})np$ if we choose δ sufficiently small.

Also we note that with high probability the number of edges in $G(n,p)$ between any two disjoint subsets $S, T \subset [n]$ of size $|S| = |T| \leq \frac{n}{4}$ is at most

$$e(S,T) \leq \left(\frac{1}{4}+\gamma\right) np|S|$$

for any small $\gamma > 0$.

Indeed, for any fixed $S, t \subseteq [n]$ with $|S| = |T| = r \leq \frac{n}{4}$ the expected number of edges between S and T is r^2p and since $r \leq \frac{n}{4}$ we have that

$$|r^2p - \left(\frac{1}{4}+\gamma\right) nrp| \geq \gamma nrp.$$

Hence by the Chernoff bounds

$$\begin{aligned}
\mathbb{P}\left(e(S,T) \geq \left(\frac{1}{4}+\gamma\right) npr\right) &\leq \mathbb{P}\left(\mathrm{Bin}(r^2,p) \geq r^2p + \gamma nrp\right) \\
&\leq \exp\left(-\frac{(\gamma nrp)^2}{2\left(r^2p + \frac{\gamma nrp}{3}\right)}\right) \\
&\leq \exp\left(-\gamma^2 nrp\right)
\end{aligned}$$

and hence the probability that there exist any such S and T is at most

$$\begin{aligned}
\sum_{r=1}^{\frac{n}{4}} \binom{n}{r}^2 e^{-\gamma^2 nrp} &\leq \sum_{r=1}^{\frac{n}{4}} \left(\frac{en}{r}\right)^{2r} e^{-\gamma^2 nrp} \\
&\leq \sum_{r=1}^{\frac{n}{4}} \left(\frac{n^2 e^{2-\gamma^2 np}}{r^2}\right)^r \\
&\leq \sum_{r=1}^{\frac{n}{4}} (o(1))^r = o(1),
\end{aligned}$$

since $p = \omega\left(\frac{\log n}{n}\right)$ 4.

Let us choose a random partition of $[n]$ into two sets of size $\frac{n}{2}$, X and Y and consider the bipartite subgraph $G[X,Y] := G'$ of G between X and Y.

We claim that with high probability in G' $d_Y(x) \geq \left(\frac{1}{4} + \frac{\varepsilon}{8}\right) np$ for all $x \in X$ and $d_X(y) \geq \left(\frac{1}{4} + \frac{\varepsilon}{8}\right) np$ for all $y \in Y$.

We note that, if we chose a bipartiton of $[n]$ uniformly at random, then the degree of a vertex $x \in X$ to Y would be distributed as $\mathrm{Bin}(d_G(v), \frac{1}{2})$ and so the probability that its degree is too small would be $o(n^{-2})$ by the Chernoff bounds, and the result would follow via the union bound. However, the degree of x to Y is not quite binomially distributed, instead it has a *hypergeometric distribution* with average $\frac{d(v)}{2}$. However one can show that the Chernoff bounds also hold for the hypergeometric distribution.

Hence there exists a partition X and Y such that the minimum degree of G' is at least $\left(\frac{1}{4} + \frac{\varepsilon}{8}\right) np$. We claim that G' satisfies Hall's condition.

Suppose we have a subset $S \subset X$ of size $> \frac{n}{4}$ such that $|N(S)| < |S|$. Then any subset $S' \subset Y \setminus N(S)$ of size $\leq \frac{n}{4}$ will also not satisfy Hall's condition. Hence it will be sufficient to show that Hall's condition holds for all subsets of size $\leq \frac{n}{4}$.

So, let $S \subset X$ be of size $\leq \frac{n}{4}$. By our assumption on $G(n,p)$ there are at most

$$\left(\frac{1}{4} + \gamma\right) np|S|$$

many edges between S and any subset $T \subset Y$ of size $|S|$ (and so also of size $\leq |S|$) in $G(n,p)$, and hence at most that many edges in G'. On the other hand, there are at least

$$e(S, N(S)) = \sum_{s \in S} d_{G'}(s) \geq \left(\frac{1}{4} + \frac{\varepsilon}{8}\right) np|S|$$

many edges from S to it's neighbourhood. If we choose $\gamma < \frac{\varepsilon}{8}$ then it follows that $|N(S)| \geq |S|$. □

7.2 Chromatic Number

In this section we want to consider the 'resilience' of the chromatic number of $G(n,p)$. Recall that for a fixed $p \in (0,1)$ we know that with high probability $\chi(G(n,p)) \approx \frac{n}{\log_b n}$ where $b = \frac{1}{1-p}$ and in fact the upper bound holds with probability

$$1 - e^{-\Omega(n^{2-o(1)})}. \tag{7.1}$$

We will show that the property that $\chi \approx \frac{n}{\log_b n}$ is very resilient to local changes, although the precise result we show will be difficult to express in terms of $\Delta_{\mathcal{P}}$ for a property $\mathcal{P}$.

Theorem 7.2. *Suppose H is a graph on $[n]$ with $\Delta(H) = n^{o(1)}$. If p is constant then with high probability*

$$\chi(G(n,p) \cup E(H)) \approx \chi(G(n,p)).$$

Remark 7.3. *Rather informally we could think about the above as saying that $\Delta_{\mathcal{P}} \geq n^{o(1)}$ where $\mathcal{P}$ is the property that $\chi \approx \frac{n}{\log_b n}$.*

Proof. Let $s = 20\Delta \log^2 n$. We randomly partition $[n]$ into s sets $V_1, \dots, V_s$ of size $\frac{n}{s}$.

We note that, by (7.1) with high probability the chromatic number of each $G(n,p)[V_i]$ is

$$\chi\left(G(n,p)[V_i]\right) \leq \frac{\frac{n}{s}}{2\log_b \frac{n}{s}} \approx \frac{1}{s}\frac{n}{2\log_b n},$$

where we used that $\Delta(H) = n^{o(1)}$.

If we just colour each of these colour classes independently, using a different set of $\approx \frac{1}{s}\frac{n}{2\log_b n}$ colours for each V_i, then clearly this is a proper colouring of $G(n,p)$.

Furthermore, any edge of H that lies between different V_i and V_j is properly coloured, so we only have to deal with the edges of H which lie inside some partition class.

Let Y denote the number of edge of H that have endpoints in the same partition class. We have that $\mathbb{E}(Y) \leq \frac{\Delta n}{2s}$ and so it follows that there exists some partition such that $Y \leq \frac{\Delta n}{s2} = \frac{n}{20\log^2 n}$.

However, if we let W be the set of endpoints of the edges in Y then $|W| \leq \frac{n}{10\log^2 n}$. Via a standard argument we will see that with high probability every set of this size spans 'few' edges in $G(n,p)$, and so has low *degeneracy*, the vertices can be linearly ordered so that each vertex has 'few' neighbours appearing before it. Since the maximum degree of Y is also small, this is also true in $G(n,p) \cup Y$, and so it is easy to recolour this subgraph using 'few' additional colours.

More precisely we claim that with high probability every set of $t \leq \frac{n}{10\log^2 n}$ vertices in $G(n,p)$ span at most $\frac{2npt}{\log^2 n}$ many edges.

Indeed, the probability there exists a set not satisfying this condition is at most

$$\begin{aligned}
&\sum_{t=1}^{\frac{n}{10\log^2 n}} \binom{n}{t} \binom{\binom{t}{2}}{\frac{2npt}{\log^2 n}} p^{\frac{2npt}{\log^2 n}} \\
&\leq \sum_{t=1}^{\frac{n}{10\log^2 n}} \left(\frac{en}{t}\right)^t \left(\frac{et^2\log^n}{2npt}\right)^{\frac{2npt}{\log^2 n}} p^{\frac{2npt}{\log^2 n}} \\
&= \sum_{t=1}^{\frac{n}{10\log^2 n}} \left(\frac{en}{t}\left(\frac{et\log^2 n}{2n}\right)^{\frac{2np}{\log^2 n}}\right)^t \\
&= \sum_{t=1}^{\frac{n}{10\log^2 n}} (o(1))^t = o(1)
\end{aligned}$$

and so it follows that every induced subgraph of $G(n,p)$ size at most $\frac{n}{10\log^2 n}$ has minimum degree at most $\frac{2np}{\log^2 n}$, and so every such subgraph has degeneracy at most $\frac{2np}{\log^2 n}$.

It follows that $G(n,p)[W] \cup Y$ has degeneracy at most $\frac{2np}{\log^2 n} + \Delta$, and so can be properly coloured using at most

$$\frac{2np}{\log^2 n} + \Delta + 1 = o\left(\frac{n}{2\log_b n}\right)$$

many new colours. Hence with high probability

$$\chi(G(n,p) \cup E(H)) \leq (1+o(1))\frac{n}{2\log_b n} \approx \chi(G(n,p)).$$

□

7.3 Hamiltonian Cycles

Recall that $\hat{p} = \frac{\log n}{n}$ is also a threshold function for the property of containing a Hamiltonian cycle, which we will denote by $\mathcal{H}$.

As before it is clear that the resilience for Hamiltonicty cannot be more than the resilience for containg a perfect matching, and so

$$\Delta_{\mathcal{H}} \leq \left(\frac{1}{2} + \varepsilon\right) np$$

for any $\varepsilon > 0$ and any $p = \omega\left(\frac{\log n}{n}\right)$. Sudakov and Vu showed that in fact this is the correct order for $\Delta_{\mathcal{H}}$ as long as p is sufficiently large, in particular $p = \omega\left(\frac{\log^4 n}{n}\right)$. Eventually this was extended to any $p = \omega\left(\frac{\log n}{n}\right)$ by Sudakov and Lee, and even hitting time versions of this result are now known.

We will prove a slightly weaker result, but one which uses a very different proof technique. The previous results all used the Pósa rotation-extension technique, whereas, in order to consider the problem in the directed setting, Ferber, Nenadov, Noever, Peter and Trujić used the *absorbing method*.

Theorem 7.4. *Let $p = \omega\left(\frac{\log^{10} n}{n}\right)$. Then with high probability in $G(n,p)$*

$$\left(\frac{1}{2} - \varepsilon\right) np \leq \Delta_{\mathcal{H}} \leq \left(\frac{1}{2} + \varepsilon\right) np$$

for any constant $\varepsilon > 0$.

Let's give a rough description of the plan, before diving into the details.

Suppose we split $[n]$ up into pieces of size $\approx \frac{n}{\log^5 n} = t$ randomly. Similar arguments as in the case of the resilience of perfect matchings will show that, with high probability, after removing H we can still find matchings between any pair of parts.

If we arrange the parts in a line, and combine the matchings, we'll cover almost all the vertices of the graph by paths of length around $\log n$. If we could join the endpoints $\{x_i, y_i\}$ of these paths together we could form a cycle on almost all the vertices of the graph.

Whilst we can't hope to do so directly, if we saved a small set of vertices at the beginning, call it U, we might hope that if U is large enough then with high probability (even after removing H) we can find disjoint paths linking the correct pairs $\{x_i, y_i\}$ together which lie inside U. Indeed, we will see that this will be likely as long as U is significantly larger than t, and so we can choose U to still be much smaller than n.

This leaves us with a large cycle which covers everything but a small set X of remaining vertices in U. In order to deal with these vertices we will find a small *absorbing path*: A path P such that for any small subset of $X \subseteq U$ there is a path P', with the same endpoints as P, whose vertex set is $V(P) \cup X$. Then, if we remove P before we find our matchings, we can use our matchings together with U to find a long path, with the same endpoints as P, which covers every vertex but $V(P) \cup X$, and then use the absorbing properties of P to complete this up to a Hamiltonian cycle.

The details of this proof will however take a low of work. Let us first show that we may assume that $G(n,p) \setminus E(H)$ is sufficiently 'pseudo-random', in a specific sense that we will make precise, which will be sufficient for the rest of the proof to function.

7.3.1 A Pseudo-Random Condition

Let us say that a graph $G = (V, E)$ with $|V| = n$ is (n, α, p)-*pseudo-random* if

(P1) $d_G(v) \geq \left(\frac{1}{2} + \alpha\right) np$ for all $v \in V$;

(P2) $e_G(S) \leq |S| \log^3 n$ for all $S \subseteq V$ with $|S| \leq \frac{10 \log^2 n}{p}$;

(P3) $e_G(S,T) \leq \left(1 + \frac{\alpha}{4}\right) |S||T|p$ for all disjoint $S, T \subseteq V$ with $|S|, |T| \geq \frac{\log^2 n}{p}$.

Lemma 7.5. *Let $\alpha > 0$ and suppose that $p \geq \frac{\log^{10} n}{n}$. Let H be a subgraph of K_n with maximum degree $\Delta(H) \leq \left(\frac{1}{2} - 3\alpha\right) np$ and let $G = G(n,p) \setminus E(H)$. Then with high probability G is (n, α, p)-pseudo-random.*

Proof. Property (P1) follows from Chernoff's inequality as before.

Furthermore, we will show that Properties (P2) and (P3) are true with high probability even in $G(n,p)$. Indeed, the probability that there exists some $S \subset [n]$ with $|S| \leq \frac{10 \log^2 n}{p}$ and $e_{G(n,p)}(S) \geq |S| \log^3 n$ is at most

$$\begin{aligned}
\sum_{s=\log n}^{\frac{10\log^2 n}{p}} \binom{n}{s} \binom{\binom{s}{2}}{s \log^3 n} p^{s \log^3 n} &\leq \sum_{s=\log n}^{\frac{10\log^2 n}{p}} \left(\frac{en}{s}\right)^s \left(\frac{eps^2}{s \log^3 n}\right)^{s \log^3 n} \\
&\leq \sum_{s=\log n}^{\frac{10\log^2 n}{p}} \left(\frac{en}{s} \left(\frac{10e}{\log n}\right)^{\log^3 n}\right)^s \\
&\leq \sum_{s=\log n}^{\frac{10\log^2 n}{p}} (o(1))^s = o(1).
\end{aligned}$$

Hence (P2) holds with high probability in $G(n,p)$ and so also in $G \subseteq G(n,p)$.

Also, for any fixed $S, T \subseteq [n]$ the Chernoff bounds imply that

$$\mathbb{P}\left(e_{G(n,p)}(S,T) \geq \left(1+\frac{\alpha}{4}\right)|S||T|p\right) \leq e^{-\frac{\alpha^2|S||T|p}{50}}.$$

Hence the probability that there exists $S, T \subseteq [n]$ with $|S|, |T| \geq \frac{\log^2 n}{p}$ and $e_G(S,T) \geq \left(1+\frac{\alpha}{4}\right)|S||T|p$ is at most

$$\begin{aligned}
\sum_{s,t \geq \frac{\log^2 n}{p}} \binom{n}{s}\binom{n}{t} e^{-\frac{\alpha^2|S||T|p}{50}} &\leq \left(\frac{en}{s}\right)^s \left(\frac{en}{t}\right)^t e^{-\frac{\alpha^2 stp}{50}} \\
&\leq \sum_{s,t \geq \frac{\log^2 n}{p}} \left(\frac{e^{1-\frac{\alpha^2 tp}{100}} n}{s}\right)^s \left(\frac{e^{1-\frac{\alpha^2 sp}{100}} n}{t}\right)^t \\
&\leq \sum_{s,t \geq \frac{\log^2 n}{p}} \left(\frac{e^{1-\frac{\alpha^2 \log^2 n}{100}} n}{s}\right)^s \left(\frac{e^{1-\frac{\alpha^2 \log^2 n}{100}} n}{t}\right)^t \\
&\leq \left(\sum_{s \geq \frac{\log^2 n}{p}} \left(\frac{e^{1-\frac{\alpha^2 \log^2 n}{100}} n}{s}\right)^s\right)^2 \\
&= o(1)
\end{aligned}$$

Hence (P3) holds with high probability in $G(n,p)$ and so also in $G \subseteq G(n,p)$. □

7.3.2 Pseudo-random implies Hamiltonian

So, for the rest of the proof we may assume that $G := G(n,p) \setminus H$ is (n,α,p)-pseudo-random. From this point this will be the only property of the graph G that we will use, so we can forgot how it was generated. We will also fix a parameter $\ell = 12\log n + 3$ for the rest of this section.

We randomly partition $[n]$ into sets $V_1 \cup V_2 \cup V_3 \cup V_4 \cup V_5$ such that

$$|V_1| = \left\lceil \frac{4\log^6 n}{p} \right\rceil \quad \text{and} \quad |V_2| = |V_3| = |V_4| = \frac{\alpha n}{6},$$

so that

$$|V_5| = \left(1-\frac{\alpha}{2}\right)n - |V_1| \approx \left(1-\frac{\alpha}{2}\right)n.$$

V_1 here will be our small set which allows us to link together our paths, and $V_2 \cup V_3 \cup V_4$ will be used to construct our absorbing path P.

For every v and every i the number of neighbours of v in V_i has a hypergeometric distribution, but will be closely approximated by a binomial random variables $\mathrm{Bin}(d_G(v), \frac{|V_i|}{n})$ and so, since np is sufficiently large and G is (n,α,p)-pseudo-random,

$$\mathbb{E}(d_{V_i}(v)) = \frac{|V_i|}{n} d_G(v) \geq \left(\frac{1}{2}+\alpha\right)|V_i|p = \omega(\log n),$$

and so by the Chernoff bounds, with high probability

$$d_{V_i}(v) \geq (1 + o(1)) \left(\frac{1}{2} + \alpha\right) |V_i| p \tag{7.2}$$

for each $v \in [n]$ and $i \in [5]$.

So, in order to follow our proof sketch we will need three parts: A *connecting lemma* which lets us use V_1 to join many pairs of vertices in G, an *absorbing lemma* which lets use build an absorbing path for the remaining vertices in V_1 using $V_2 \cup V_3 \cup V_4$ and finally a *covering lemma* that allows us to cover the remaining vertices with paths (although we won't actually give this step its own lemma).

The first is given by the following lemma, whose proof is lengthy and we will defer to the next section.

Lemma 7.6. *Let G be (n, α, p)-pseudo-random and let $\{a_i, b_i \colon i \leq t\}$ be a family of pairs of vertices of G such that $a_i \neq a_j$ and $b_i \neq b_j$ for every distinct i, j ($a_i = b_i$ is allowed), where $t \geq \frac{\log^3 n}{p}$. Let $L = \bigcup_i \{a_i, b_i\}$ and assume that $K \subseteq [n] \setminus L$ is such that*

(C1) $|K| = \omega(\ell t \log t)$;

(C2) For every $v \in K \cup L$ we have $d_K(v) \geq (1 + o(1)) \left(\frac{1}{2} + \alpha\right) |K| p$.

Then there exists a family of t internally disjoint paths $P_1, \ldots, P_t$ such that P_i connects a_i to b_i and $V(P_i) \setminus L \subseteq K$. Furthermore, each path is of length ℓ.

Using this lemma we can also prove the existence of the absorbing path.

Lemma 7.7. *There is a path P^* with $V(P^*) \subseteq V_2 \cup V_3 \cup V_4$ such that for every $W \subseteq V_1$ there is a path P^*_W such that $V(P^*_W) = V(P^*) \cup W$ and such that P^* and P^*_W have the same endpoints.*

Proof. Given a vertex $x \in V_1$ we will say a subgraph $A \ni x$ is an *absorber* for x if there are two vertices s and t in A such that A contains both an (s, t)-path P_x of length $|A| - 1$ which doesn't contain x (and so contains all other vertices in A), and an (s, t)-path P'_x of length $|A|$ which does contain x.

Let k, r be integers, we will build an absorber A of size $3 + 2kr$ for x as follows: We first take a cycle C of length $4k + 3$ which will consist of vertices

$$s, s_2, t_1, s_3, t_2, s_4, \ldots, s_{2k}, t_{2k-1}, t, t_{2k}, s_1, x, s$$

together with a set of $2k$ pairwise vertex disjoint (s_i, t_i)-paths P_i, one for each $i \in [2k]$, each of which is of length r.

Clearly $|A| = 3 + 2kr$ and we can see that

$$P_x = s, s_2, P_2, t_2, s_4, P_4, \ldots . s_{2k}, P_{2k}, t_{2k}, s_1, P_1, t_1, s_3, P_3, t_3, \ldots, s_{2k-1}, P_{2k-1}, t_{2k-1}, t$$

is a path of length $|A| - 1$ which avoids x and

$$P'_x = s, x, s_1, P_1, t_1, s_2, P_2, t_2, s_3, \ldots, s_{2k}, P_{2k}, t_{2k}, t$$

is a path of $|A|$. Hence A is an absorber for x. We will build our path P^* in two stages. Firstly we will construct an absorber A_x for each $x \in V_1$, such that $P_x \subseteq V_2 \cup V_3$ which are disjoint for different $x \in V_1$. Then, using Lemma 7.6 we will connect all of the paths P_x into one long path using vertices from V_4.

To build the absorbers A_x we will again use the connecting lemma to build first the cycle C, and then the set of paths P_i. Let us take $k = 3\lceil \log n \rceil$ and $r = \ell$, so that the absorber consists of a cycle C of length $4k + 3 = \ell$ together with $2k = 6\lceil \log n \rceil$ disjoint paths, each of length ℓ.

In order to find the cycle C^x for each $x \in V_1$ we apply Lemma 7.6 to the set V_2 (as K) with $t = |V_1|$ and a family of pairs $\{(x, x) \colon x \in V_1\}$. Note that

$$|V_2| = \Theta(n) = \omega\left(\ell |V_1| \log |V_1|\right),$$

since $|V_1| = \Theta\left(\frac{\log^3 n}{p}\right) = o\left(\frac{n}{\log^7 n}\right)$ and so (C1) is satisfied and (C2) follows from (7.2). Hence we can find a disjoint collection of cycles C^x of length ℓ inside V_2, one containing each $x \in V_1$.

Next, we use V_3 to join, for each $x \in V_1$ the pairs of designated vertices (s_i^x, t_i^x) in C^x. To do this we again apply Lemma 7.6, this time with $K = V_3$, $t = 2k|V_1|$ and the set of pairs $\{(s_i^x, t_i^x) \colon x \in V_1, 1 \leq i \leq 2k\}$. Note that, as before

$$|V_3| = \Theta(n) = \omega\left(\ell 2k |V_1| \log 2k |V_1|\right),$$

and so (C1) and (C2) are again satisfied. Hence we can find a disjoint collection of paths $\{P_i^x \colon x \in V_1, 1 \leq i \leq 2k|V_1|\}$ inside V_3.

Finally, we will use V_4 to build the path P^*. Recall that each A_x has specified vertices s^x and t^x. Let us take some arbitrary enumeration $V_1 = \{x_1, \dots, x_{|V_1|}\}$. We apply Lemma 7.6 with $K = V_4$, $t = |V_1| - 1$ and the set of pairs $\{(t^{x_i}, s^{x_{i+1}}) \colon 1 \leq i \leq |V_1| - 1\}$. Again it is a simple check that

$$|V_4| = \Theta(n) = \omega\left(\ell(|V_1| - 1) \log(|V_1| - 1)\right),$$

and so we can find a family P_i' of disjoint $(t^{x_i}, s^{x_{i+1}})$-paths in V_4. Hence we can combine the vertex sets of the absorbers A_x into a single path P^* using these paths.

It is easy to see that P^* satisfies the conclusion of the lemma. Indeed, given a subset $W \subseteq V_1$ let $\{A_w \colon w \in W\}$ be the set of absorbers constructed for vertices in W. By the definition of an absorber, there is for each w an absorbing path P_w' which starts at s^w and ends at t^w and covers all the vertices in A_w. By replacing each of the paths P_w with the paths P_w' in P^* we obtain a path P_W^* which has the same endpoints as P^* with $V(P_W^*) = V(P^*) \cup W$. □

Given these two lemmas let us finish the proof of Theorem 7.4.

Proof of Theorem 7.4. Let P^* be as in Lemma 7.7 and let $U = (V_2 \cup V_3 \cup V_4 \cup V_5) \setminus V(P^*)$. Given any $v \in U$ we have that

$$\begin{aligned} d_U(v) \geq d_{V_5}(v) &\geq (1 + o(1)) \left(\frac{1}{2} + \alpha\right) |V_5| p \\ &\leq (1 + o(1)) \left(\frac{1}{2} + \alpha\right) \left(1 - \frac{\alpha}{2}\right) np \\ &\geq \left(\frac{1}{2} + \frac{\alpha}{2}\right) |U| p. \end{aligned}$$

Let us take $s = \lceil \frac{\log^3 n}{p} \rceil$ and $k = \lfloor \frac{|U|}{s} \rfloor$ and randomly choose disjoint sets $S_1, \dots, S_k \subseteq U$ of size s. Let $S = \bigcup_{i=1}^{k} S_i$ and let $S' = U \setminus S$. Note that $|S'| \leq s$.

Note that, for any $v \in U$ and $i \in [k]$,

$$\mathbb{E}(d_{S_i}(v)) = \frac{|S_i| d_U(v)}{|U|} \geq \left(\frac{1}{2} + \frac{\alpha}{2}\right) |S_i| p = \omega(\log n). \tag{7.3}$$

It follows from the Chernoff bound that with high probability

$$d_{S_i}(v) > \left(\frac{1}{2} + \frac{\alpha}{4}\right) |S_i| p$$

for every $v \in U$ and $i \in [k]$. We claim that this condition implies there is a perfect matching M_i between each S_i and S_{i+1}. Let us fix some $i \in [k-1]$, we wish to show that $G[S_i, S_{i+1}]$ satisfies Hall's theorem. By our standard argument we only need to check Hall's condition is satisfied for sets of size at most $\frac{s}{2}$.

Let $X \subseteq S_i$ be such that $|X| \leq \frac{10 \log^2 n}{p}$, let Y be the neighbourhood of X in S_{i+1} and suppose that $|Y| < |X|$. However, by (7.3)

$$e_G(X \cup Y) > \left(\frac{1}{2} + \frac{\alpha}{4}\right) |X||S_{i+1}| p \geq |X \cup Y| \log^4 n.$$

However, this contradicts property (P2) of (n, α, p)-pseudo-random graphs.

Conversely, suppose that $\frac{10 \log^2 n}{p} \leq |X| \leq \frac{s}{2}$ and $|Y| < |X|$. In this case we have, again by (7.3), that

$$e_G(X, Y) > \left(\frac{1}{2} + \frac{\alpha}{4}\right) |X||S_{i+1}| p \geq \left(1 + \frac{\alpha}{4}\right) |X||Y| p,$$

contradicting property (P3). A similar argument holds for subsets $X \subseteq S_{i+1}$.

Hence we may assume that there exist matchings M_i between S_i and S_{i+1} for each $i \in [k-1]$. By combining these matchings, we obtain a set of s vertex disjoint paths from S_1 to S_k. If we also take a path of length 0 for each $v \in S'$ we have a set of $t' = s + |S'| \leq 2s$ many vertex disjoint paths, with endpoints say x_i, y_i for $i = 1, \dots, t'$, which cover the set U. Let $t = t' + 1$ and let x_t, y_t be the endpoints of the path P^*.

We apply Lemma 7.6 with $K = V_1$, $t = t$ and with the set of pairs $\{y_i, x_{i+1} \colon i \in [t]\}$. Since $t \leq 2s + 1$ and $s = \lfloor \frac{\log^3 n}{p} \rfloor$,

$$|V_1| = \left\lceil \frac{\log^6 n}{p} \right\rceil = \omega(\ell t \log t).$$

Indeed, $\ell t \log t = O\left(\frac{\log^4 n}{p}(\log \log n - \log p)\right)$ and since $p = \omega\left(\frac{\log^{10} n}{n}\right)$, $-\log p = O(\log n)$. Furthermore $d_{V_1}(v) \geq (1 + o(1))\left(\frac{1}{2} + \alpha\right) |V_1| p$ for every $v \in V$ by (7.2) and so (C1) and (C2) are satisfied.

Hence there exists a family of vertex disjoint $x_i - y_i$ paths in V_1. We can use these paths, together with the previously constructed path family to form a cycle C which covers $V_2 \cup V_3 \cup V_4 \cup V_5$, as well as some vertices in V_1, and contains P^* as a subpath. Letting $W = V_1 \setminus C$, we can use P^* to absorb W into a longer path P^*_W with the same endpoints as P^*, which extends C to a Hamiltonian cycle of G. □

7.3.3 Proof of the Connecting Lemma

We will need to begin with some lemmas on expansion properties of pseudo-random graphs. Firstly let us define for (not necessarily disjoint) subsets X and Y of $V(G)$ and an integer k

$$N_G^k(X,Y) = \{y \in Y : \text{ there exists an } (x,y)-\text{path } P \text{ of length } k \text{ with } V(P)\setminus\{x\} \subseteq Y\},$$

and we will write $N_G(X,Y)$ for $N_G^1(X,Y)$.

The following lemma we will need a bit later to guarantee the existence of many neighbours of a small set of vertices in a larger set, in order to find some perfect matchings.

Lemma 7.8. *Let $X, Y \subseteq V(G)$ be such that $|X| = \left\lfloor \frac{\log^2 n}{p} \right\rfloor$, $|Y| \geq \frac{50\log^3 n}{\alpha p}$ and that $|N_G(x,Y)| \geq \left(\frac{1}{2}+\frac{\alpha}{2}\right)|Y|p$ for all $x \in X$. Then*

$$|N_G(X,Y)| \geq \left(\frac{1}{2}+\frac{\alpha}{3}\right)|Y|.$$

Proof. Let $Z = X \cup N_G(X,Y)$. Since $|X|$ is small, by property (P2) it contains few edges and so

$$\begin{aligned} e_G(Z) &\geq \sum_{x\in X} N_G(x,Y) - e_G(X) \\ &\geq \left(\frac{1}{2}+\frac{\alpha}{2}\right)p|Y||X| - |X|\log^3 n \\ &\geq \left(\left(\frac{1}{2\alpha}+\frac{1}{2}\right)\frac{\alpha p|Y|}{\log^3 n} - 1\right)|X|\log^3 n \\ &\geq \left(\frac{25}{\alpha}+25-1\right)|X|\log^3 n \\ &= \left(\frac{25}{\alpha}+24\right)\frac{|X|}{|Z|}|Z|\log^3 n. \end{aligned}$$

If $|Z| < \frac{10\log^2 n}{p}$, then $\frac{|X|}{|Z|} \geq \frac{1}{10}$ and so

$$e_G(Z) \geq \frac{25}{10\alpha}|Z|\log^3 n,$$

contradicting property (P2) if α is sufficiently small. Hence $|Z| \geq \frac{10\log^2 n}{p}$, and so $|N_G(X,Y)| \geq \frac{9\log^2 n}{p}$. Let $Y' = N_G(X,Y)\setminus X$, then $\frac{8\log^2 n}{p} \leq |Y'|$, and so by property (P3) we have that $e_G(X,Y') \leq \left(1+\frac{\alpha}{4}\right)|X||Y'|p$.

However, since $N_G(x,Y) \geq \left(\frac{1}{2}+\frac{\alpha}{2}\right)|Y|p$ for all $x \in X$ it follows that

$$\begin{aligned} \left(\frac{1}{2}+\frac{\alpha}{2}\right)|X||Y|p &\leq e_G(X,Y') + 2e_G(X) \\ &\leq e_G(X,Y') + 2|X|\log^3 n \\ &\leq \left(1+\frac{\alpha}{4}\right)|X||Y'|p + 2|X|\log^3 n \end{aligned}$$

and so

$$\left(\frac{1}{2}+\frac{\alpha}{2}-\frac{\alpha}{25}\right)|X|Y|p \leq \left(\frac{1}{2}+\frac{\alpha}{2}\right)|X||Y|p - 2|X|\log^3 n \leq \left(1+\frac{\alpha}{4}\right)|X||Y'|p,$$

since $\frac{\alpha p}{25}|X||Y|p \geq 2|X|\log^3 n$. Hence

$$|Y'| \geq \frac{\left(\frac{1}{2}+\frac{\alpha}{2}-\frac{\alpha}{25}\right)}{\left(1+\frac{\alpha}{4}\right)}|Y| \geq \left(\frac{1}{2}+\frac{\alpha}{3}\right)|Y|,$$

for α sufficiently small. □

The following lemma, allows us to find many paths of a fixed length from a single vertex in a set X to a large enough set Y.

Lemma 7.9. *Let $X, Y \subseteq V(G)$ be disjoint sets such that*

(E1) $|Y| \geq \frac{130\log^3 n}{\alpha p}$*;*

(E2) $|N_G(X,Y)| \geq \frac{2\log^2 n}{p}$*;*

(E3) $|N_G(S,Y)| \geq \left(\frac{1}{2}+\frac{\alpha}{4}\right)|Y|$ *for all* $S \subseteq Y$ *with* $|S| \geq \frac{\log^2 n}{p}$*.*

Then, for any $2\log n \leq h \leq \ell$*, there exists* $x \in X$ *such that* $N_G^h(x,Y) \geq \left(\frac{1}{2}+\frac{\alpha}{8}\right)|Y|$*.*

Proof. We will show the existence of such an x using an inductive argument. More precisely, for each $i < h$ we claim that if $A \subseteq X$ is such that $|N_G^i(A,Y)| \geq \frac{2\log^2 n}{p}$ then A has a subset $A' \subseteq A$ with $|A'| \leq \left\lceil \frac{|A|}{2} \right\rceil$ such that $|N_G^{i+1}(A,Y)| \geq \frac{2\log^2 n}{p}$.

By (E2) we can apply this claim to X, and the subsequent subsets we find $h-2$ times, to find $X' \subseteq X$ such that $|X'| \leq \left\lceil \frac{|X|}{2^{h-2}} \right\rceil$ with $|N_G^{h-1}(X',Y)| \geq \frac{2\log^2 n}{p}$. However, since $h-2 \geq \log n$ it follows that $|X'| = 1$, and so $X' = \{x\}$. Let $M \subseteq N_G^{h-1}(x,Y)$ be of size $\left\lceil \frac{\log^2 n}{p} \right\rceil$, note that by definition there is a path P_w of length $h-1$ from x to each $w \in M$. Let $V^* = \left(\bigcup_{w\in M} P_w\right) \setminus \{x\}$.

Since $|M| \geq \left\lceil \frac{\log^2 n}{p} \right\rceil$, by (E3) $|N_G(M,Y)| \geq \left(\frac{1}{2}+\frac{\alpha}{4}\right)|Y|$ and hence, since $|V^*| \leq h|M|$,

$$\begin{aligned} N_G^h(x,Y) &\geq |N_G(M, Y \setminus V^*)| \\ &\geq |N_G(M,Y)| - |V^*| \\ &\geq \left(\frac{1}{2}+\frac{\alpha}{4}\right)|Y| - h|M| \\ &\geq \left(\frac{1}{2}+\frac{\alpha}{8}\right)|Y|, \end{aligned}$$

since $h|M| \leq \ell M \leq \frac{13\log^3 n}{p} \leq \frac{\alpha}{10}|Y|$.

So, it remains to prove the claim.

Note that, if $A = A_1 \cup A_2$ is a partition of A with $|A_1| = \lceil \frac{|A|}{2} \rceil$ then

$$|N_G^i(A_1, Y)| + |N_G^i(A_2, Y)| \geq |N_G^i(A, Y)| \geq \frac{2\log^2 n}{p}.$$

Hence we may assume that there is some $A' \subseteq A$ with $|A'| \leq \lceil \frac{|A|}{2} \rceil$ and $|N_G^i(A', Y)| \geq \frac{\log^2 n}{p}$.

Let us choose some subset $B \subseteq N_G^i(A', Y)$ of size $|B| = \lceil \frac{\log^2 n}{p} \rceil$. Then, by (E3)

$$|N_G(B, Y)| \geq \left(\frac{1}{2} + \frac{\alpha}{4}\right)|Y|.$$

Each $v \in B$ is the endpoint of a path P_v of length i from a vertex in A', let $V^* = \cup_{v \in B} P_v$ be the set of all the vertices in these paths. Then

$$|N_G^{i+1}(A', Y) \geq |N_G(B, Y)| - |V^*| \geq \left(\frac{1}{2} + \frac{\alpha}{4}\right)|Y| - \ell|B| \geq \frac{130\log^3 n}{2\alpha p} - \frac{13\log^3 n}{p} \geq \frac{2\log^2 n}{p},$$

if α is sufficiently small. □

Using this we can try to find appropriate paths for Lemma 7.6 in a greedy fashion. We split our set K up into two parts R_A and R_B which are both large, and we will apply Lemma 7.9 to $\{a_i : i \in [t]\}$ and R_A to see that there is some a_i that is connected to more than half of R_A by a path of length $\ell - 1/2$ in R_A and actually by repeated applications this is true for more than half of the a_i, and a similar statement holds true for the b_j and R_B. Hence there is some i where it's true for both a_i and b_i. Then, since $S = N_G^{\ell-1/2}(a_i, R_A)$ is large we will in fact find that (if we chose R_A and R_B sensibly) S must have neighbours in $N_G^{\ell-1/2}(b_i, R_B)$, allowing us to find a path of length ℓ from a_i to b_i in K.

This allows us to find one such path, and in fact we can repeat this process many times to give an approximate version of Lemma 7.6, allowing us to find half the paths. However, once the remaining set of pairs a_i, b_i is too small we can no longer apply Lemma 7.9.

Lemma 7.10. *Let $\{a_i, b_i : i \leq t\}$ be a family of pairs of vertices of G such that $a_i \neq a_j$ and $b_i \neq b_j$ for every distinct i, j, where $t \geq \frac{\log^3 n}{p}$. Let $L = \bigcup_i \{a_i, b_i\}$ and let $R_A, R_B \subseteq [n] \setminus L$ be disjoint and such that*

(D1) $|R_A|, |R_B| \geq \frac{48t\ell}{\alpha}$,

(D2) For $Z = A, B$

$$|N_G(S, R_Z)| \geq \left(\frac{1}{2} + \frac{\alpha}{4}\right)|R_Z|$$

for all $S \subseteq R_A \cup R_B \cup L$ such that $|S| \geq \frac{\log^2 n}{p}$.

Then there exists a set $I \subseteq [t]$, $|I| = \lfloor \frac{t}{2} \rfloor$ and internally disjoint (a_i, b_i)-paths P_i for each $i \in I$ such that $V(P_i) \setminus \{a_i, b_i\} \subseteq R_A \cup R_B$. Furthermore if $10\lceil \log n \rceil \leq \ell' \leq \ell$ then we can choose the paths to be of length ℓ'.

Proof. We proceed by induction. Suppose we have already found $s < \lfloor \frac{t}{2} \rfloor$ (a_i, b_i)-paths P_i for some subset $J \subseteq [t]$ of size $|J| = s$. Let us consider

$$R'_A := R_A \setminus \bigcup_{i \in J} V(P_i) \quad R'_B := R_B \setminus \bigcup_{i \in J} V(P_i).$$

Let us pick $h_A, h_B \geq 2 \log n$ such that $h_A + h_B + 1 = \ell'$. We claim that there exists some $i \in K = [t] \setminus J$ such that

$$|N_G^{h_A}(a_i, R'_A)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) |R'_A| \quad \text{and} \quad |N_G^{h_B}(b_i, R'_B)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) |R'_B|.$$

Indeed, we will find a set $I_A \subseteq K$ of size $|I_A| = \lfloor \frac{|K|}{2} \rfloor + 1$ such that

$$|N_G^{h_A}(a_i, R'_A)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) |R'_A|$$

for every $i \in I_A$, and similarly a set $I_B \subseteq K$ of size $|I_B| = \lfloor \frac{|K|}{2} \rfloor + 1$ such that

$$|N_G^{h_B}(b_i, R'_B)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) |R'_B|$$

for every $I \in I_B$. Since $I_A \cap I_B$ must then be non-empty, this guarantees the existence of such an i.

In order to show that I_A exists, suppose we have already found $v_1, \ldots, v_k \in \{a_i \colon i \in K\}$ such that

$$|N_G^{h_A}(v_i, R'_A)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) |R'_A|$$

for each $i \in [k]$. We want to apply Lemma 7.9 with $X = \{a_i \colon i \in K\} \setminus \{v_j \colon j \in [k]\}$ and $Y = R'_A$. Let us check that the conditions are satisfied.

Since $|R'_A| \geq \frac{48t\ell}{\alpha} - t\ell \geq \frac{47\ell log^3 n}{\alpha p}$ and so (E1) is satisfied. Now, by (D2) we have that for any $S \subseteq R'_A \cup \{a_i \colon i \in K\}$ of size at least $|S| \geq \frac{\log^2 n}{p}$

$$|N_G(S, R'_A)| \geq |N_G(S, R_A)| - t\ell \geq \left(\frac{1}{2} + \frac{\alpha}{4}\right) |R_A| - t\ell \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) |R_A|$$

and hence (E3) is satisfied, and also (E2), since $|X| \geq \frac{|K|}{2} \geq \frac{t}{4} \gg \frac{\log^2 n}{p}$. Hence, by the conclusion of Lemma 7.9 we can find $v_{k+1} \in X$ such that

$$|N_G^{h_A}(v_i, R'_A)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) |R'_A|.$$

Hence by induction we can find such an I_A as claimed, and a similar argument gives the existence of an I_B. It follows that there is some $i \in K = [t] \setminus J$ such that

$$|N_G^{h_A}(a_i, R'_A)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) |R'_A| \quad \text{and} \quad |N_G^{h_B}(b_i, R'_B)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) |R'_B|.$$

Let us take $S = N_G^{h_A}(a_i, R'_A)$. Then

$$|S| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) |R'_A| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right) \frac{47\ell log^3 n}{\alpha p} \gg \frac{log^2 n}{p}.$$

Hence, by (D2)

$$|N_G(S, R'_B)| \geq |N_G(S, R_B)| - s\ell \geq \left(\frac{1}{2} + \frac{\alpha}{4}\right)|R_B| - t\ell \geq \left(\frac{1}{2} + \frac{\alpha}{6}\right)|R'_B|.$$

However, also by the claim

$$|N_G^{h_B}(b_i, R'_B)| \geq \left(\frac{1}{2} + \frac{\alpha}{8}\right)|R'_B|$$

Hence there is some vertex in $N_G^{h_B}(b_i, R'_B) \cap N_G(S, R'_B)$, or in other words, some $s \in S$ adjacent to some $s' \in N_G^{h_B}(b_i, R'_B)$. However, every vertex $s \in S$ has a path of length h_A from a_i to s contained in R'_A, and similarly every $s' \in N_G^{h_B}(b_i, R'_B)$ has a path of length h_B from b_i to s' contained in R'_B. Hence we can combine these two paths to find a (a_i, b_i)-path of length $h_A + h_B + 1 = \ell'$ which is disjoint from $\bigcup_{i \in J} V(P_i)$. □

So, we can get about halfway to proving Lemma 7.6. However, once the set X of remaining a_i becomes too small (namely, smaller than $\frac{\log^2 n}{p}$), we can no longer assume that X has sufficiently many neighbours in Y to apply Lemma 7.9.

The clever idea to get around this problem is to 'blow-up' the set of a_i and b_i that we want to connect, replacing each with say a binary tree of bounded depth, so that there are sufficiently many leaves in these trees that we can apply 7.9. In this way we can keep dealing with $\frac{1}{2}$ of the remaining vertices at a time.

7.3.4 Proof of Lemma 7.6

Firstly, we will define some more parameters that we will use throughout the proof.

$$m = \lceil \log_2 t \rceil + 1, \;\; s_i = 2t \text{ for } i \in [2m] \text{ and } s_{2m+1} = s_{2m+2} = \frac{|K|}{4}, \;\; k = 2m + 2.$$

We randomly choose disjoint sets $S_i \subseteq K$ such that $|S_i| = s_i$ for all $i \in [k]$. The Chernoff bounds together with (C2) imply that for every $v \in K \cup L$ and $i \in [k]$,

$$|N_G(v, S_i)| \geq \left(\frac{1}{2} + \frac{\alpha}{2}\right) p s_i. \tag{7.4}$$

We will prove a statement like the following at the end of the section.

Lemma 7.11. *Given sets $[t] = I_1 \supseteq I_2 \supseteq \ldots \supseteq I_m$ such that $|I_j| = \lceil \frac{|I_{j-1}|}{2} \rceil$ for each j, G contains complete binary trees $T_A(i)$ and $T_B(i)$ for each $i \in [t]$ such that*

(F1) The depth of $T_A(i)$ and $T_B(i)$ is $s - 1$ for each $i \in I_s$;

(F2) $T_A(i)$ is rooted at a_i and $T_B(i)$ is rooted at b_i for each $i \in [t]$;

(F3) The vertices $T_A(i, j)$ at depth $j \in [0, m]$ in $T_A(i)$ are contained in S_j;

(F4) The vertices $T_B(i, j)$ at depth $j \in [0, m]$ in $T_B(i)$ are contained in S_{m+j};

(F5) The trees are vertex disjoint.

Supposing we can prove the lemma our plan is as follows: We first find a collection of paths $\mathcal{P}_1$ joining the pairs a_i, b_i for half of the pairs, say for each $i \in I_1$, using Lemma 7.10. We then consider the remaining pairs, $I_2 = [t] \setminus I_1$. There are trees $T_A(i)$ and $T_B(i)$ for $i \in I_2$, which are rooted at a_i and b_i and have depth 1. We can now apply Lemma 7.10 to find a family of paths $\mathcal{L}_2$ between $T_A(i,1)$ and $T_B(i,1)$ with $i \in I_2$ of length $\ell - 2$, and use these to find paths $\mathcal{P}_2$ of length ℓ between at least half of the remaining pairs, say $i \in J_2$. We let $I_3 = I_2 \setminus J_2$ and consider paths from $T_A(i,2)$ to $T_B(i,2)$ with $i \in I_3$ of length $\ell - 4$, and so on. Since each time we apply Lemma 7.10 we have a relatively large set X to apply it to, we avoid the problems we had applying it directly.

So, there are two things left to check, firstly that we actually can apply Lemma 7.10 as claimed, and secondly that Lemma 7.11 is true. Let's start with the former.

Suppose we're in 'round' s of this procedure, let us set

$$M_s = K \setminus \left(\left(\bigcup_{i=1}^{s-1} S_i \cup S_{m+i} \right) \cup \left(\bigcup_{i=1}^{s-1} \bigcup_{Q \in \mathcal{P}_i} V(Q) \right) \right)$$

that is, M_s is the set of vertices remaining in K after we remove all the paths we've constructed so far, as well as all the vertices in the trees of depth at most $s-1$. Then in round s we apply Lemma 7.10 to $\{x_j, y_j : j \in [t]\}$ where the x_j are made up of the $T_A(i,s)$ with $i \in I_s$ and similarly the y_j are made up of the $T_B(i,s)$, and with

$$R_A = S_{2m+1} \cap M_s \quad \text{and} \quad R_B = S_{2m+2} \cap M_s.$$

Since S_{2m+1}, S_{2m+2} are disjoint from S_i with $i \in [2m]$, it follows that

$$\begin{aligned} |R_A|, |R_B| &\geq \frac{|K|}{4} - O(t\ell) \\ &\geq \frac{|K|}{5} = \omega(\ell t \log t) \end{aligned}$$

and so property (D1) holds. Furthermore, given any $S \subseteq R_A \cup R_B \cup \bigcup_j \{x_j, y_j\}$ with $|S| \geq \frac{\log^2 n}{p}$ by (7.4) and Lemma 7.8 applied with $X = S$ and $Y = R_A$

$$|N_G(S, R_A)| \geq \left(\frac{1}{2} + \frac{\alpha}{3} \right) |R_A|.$$

A similar bound holds for R_B and hence (D2) holds.

It thus follows from Lemma 7.10 that there is a collection of $\lfloor \frac{t}{2} \rfloor$ many indices $j \in [t]$ such that we have a family of internally disjoint $x_j - y_j$ paths $\mathcal{L}_s$ of length $\ell - 2s$ which are contained, apart from their endpoints, in $R_A \cup R_B$. Note that since $s \leq \log n$ we have that $\ell - 2s$ is large enough to apply Lemma 7.10. Hence, for some set J_s of at least half of the $i \in I_s$ there is some $x_j \in T_A(i,s)$ joined by some path in this family to some $y_j \in T_B(i,s)$ and so we can extend this path to an $a_i - b_i$ path of length ℓ.

Hence it just remains to prove Lemma 7.11. However, we since we don't know ahead of time what the sets $I_1, I_2, \ldots$ will be, we don't really won't to prove Lemma 7.11, but rather show that we can build the trees inductively during our process. So, what we prove is that we can guarantee that the conclusions of Lemma 7.11 hold by the end of our process, where sets $I_1, \ldots$ are those that we are constructing inductively during our proof.

Proof of 'Lemma 7.11'. At the start of round s there will be $\frac{t}{2^{s-1}}$ 'active' vertices $i \in I$ which still need to be connected by paths, and so $\frac{t}{2}$ leaves in the corresponding trees $T_A(i)$, which lie in S_{s-1}, which each need to be extended by two more neighbours in S_s, and similarly $\frac{t}{2}$ leaves in $T_B(i)$, which lie in S_{m+s-1}, which each need to be extended by two more neighbours in S_{m+s}.

Since, by (7.4)

$$|N_G(v, S_j)| \geq \left(\frac{1}{2} + \frac{\alpha}{2}\right) ps_j \geq \log^3 n$$

for each $j \in [2m]$ and $v \in K$ we can reduce this to the following problem: Given a bipartite graph Γ on sets $X = \{x_1, \ldots, x_{\frac{t}{2}}\} \subseteq S_{s-1}$ and $Y = S_s = \{s_1, \ldots, s_{2t}\}$ with minimum degree at least $\log^3 n$ and satisfying (P2) and (P3) there exists a partition of B into pairs $\{z_{i,1}, z_{i,2}\}$, $i \in [t]$ such that both edges $(x_i, z_{i,1})$ and $(x_i, z_{i,2})$ are in Γ.

A standard application of Hall's Theorem reduces this to showing that Γ satisfies the following condition:

$$|N_\Gamma(S, B)| \geq 2|S| \quad \text{for all } S \subseteq X.$$

Let $T = N_\Gamma(S, B)$. We split into two cases. Firstly if $|S| \leq \frac{3\log^2 n}{p}$ then, since $e_\Gamma(S \cup T) \geq |S| \log^3 n$, it follows from (P2) that $|S \cup T| \geq \frac{10\log^2 n}{p}$ and hence

$$|T| \geq \frac{7\log^2 n}{p} \geq 2|S|.$$

Conversely if $\frac{3\log^2 n}{p} \leq |S| \leq |X| = \frac{t}{2}$ and $|T| < 2|S|$, then (P3) (applied to a superset of T of size $2|S|$) implies that

$$e_\Gamma(S, T) \leq 2\left(1 + \frac{\alpha}{4}\right)|S|^2 p.$$

However, by (7.4) we also have that

$$e_\Gamma(S, T) \geq \left(\frac{1}{2} + \frac{\alpha}{2}\right) p|S||Y| \geq 4\left(\frac{1}{2} + \frac{\alpha}{2}\right)|S|^2 p,$$

since $|Y| = 2t \geq 4|S|$, a contradiction.

□