PROBABILISTIC METHOD LECTURE NOTE

JAEHOON KIM

Abstract. This lecture note is mainly based on the following book: The probabilistic method by Alon and Spencer.

March 4, 2021

1. Basics

What is the probabilistic method? It is a proof method using probability to tackle combinatorial problems. Let's consider some examples to see what it is.

Definition 1.1. *Given a pair of integers k, s, we define $R(k,s)$ be the smallest integer n as follows: for any red/blue-edge-coloring of K_n contains a red monochromatic K_k or blue monochromatic K_s.*

For example, it is not difficult to see that $R(3,3) = 6$. One can also prove the inequality $R(k,s) \leq R(k,s-1) + R(k-1,s)$ which implies that $R(k,k) = O(\frac{2^{2k}}{\sqrt{k}})$.

To prove a lower bound inequality $R(k,k) \geq n$, one must present an edge-coloring of K_n having no monochromatic copies of K_k. Many people tried to come up with ingenious constructions of such colorings, but all those constructions fail to provide a lower bound exponential in k until Erdős proved the following. Even by now, the best 'explicit' lower bound is $2^{(\log k)^{\omega(1)}}$ by Barak, Rao, Shaltiel, Wigderson in 2012 which is far smaller than exponential.

Theorem 1.2 (Erdős, 1947)**.** $R(k,k) > \frac{k2^{k/2}}{e\sqrt{2}}$.

Proof. Consider K_n and color each edge with red/blue independently with probability $1/2$. A set of k vertices forms a monochromatic K_k with probability $2 \cdot 2^{-\binom{k}{2}}$. Since there are $\binom{n}{k}$ such events, the probability that at least one of them occurs is at most $\binom{n}{k}2^{1-\binom{k}{2}}$. If this is less than 1, then there is an 2-edge-coloring of K_n with no monochromatic clique of size k and hence $R(k,k) > n$. Since $\binom{n}{k} < \frac{1}{2}(\frac{ne}{k})^k$, it suffices to have $\frac{ne}{k} \leq 2^{(k-1)/2}$, which is equivalent to $n \leq \frac{k2^{k/2}}{e\sqrt{2}}$. □

Since the above proof, many problems have been tackled using probabilistic method. There are some philosophies here.

(1) Sometimes considering average behavior is much easier than considering optimal behavior.
(2) Sometimes average behavior is good enough to prove what we want.
(3) Sometimes probabilistic computations provides more intuition than discrete computations and is simpler as well.

In principle, many probabilistic arguments with finite sample space can be phrased as weighted counting arguments, but the tools of probability do the job more clearly and efficiently.

To make things more rigorous, we introduce the following concepts.

Definition 1.3. *A discrete probability space is a finite or countable set S together with a function $\mathbf{Pr}$ defined on the subsets of S (called events) such that*

- *If $A \subseteq S$, then $0 \leq \mathbf{Pr}(A) \leq 1$.*
- *$\mathbf{Pr}(S) = 1$ and*
- *If $A_1, A_2, \ldots$ are pairwise disjoint subsets of S, then $\mathbf{Pr}(\bigcup A_i) = \sum_{i=1}^{\infty} \mathbf{Pr}(A_i)$.*

In the above example, we implicitly construct a probability space by just explaining that we color each edge independently with probability $1/2$. The actual probability space

consists of $2^{\binom{n}{2}}$ distinct colorings of K_n with each having probability $2^{-\binom{n}{2}}$. We will describe the probability space in the former way instead of explicitly listing all elements of our probability spaces. Let's see more applications of this probabilistic method.

Definition 1.4. *A tournament on a set V of n vertices is an orientation $T = (V, E)$ of the edges of the complete graph on the set of vertices V. In other words, for every $x \neq y \in V$, either (x, y) or (y, x) is in E, but not both. If $(x, y) \in E$, we say that x is an in-neighbor of y and y is an out-neighbor of x. We say that T has the property S_k if for every $S \in \binom{V}{k}$, there is a vertex $v \in V$ which has all vertices in S as its out-neighbors.*

One natural question is that for given $k > 0$, does there always exists a tournament with the property S_k? Indeed, if n is large enough compared to k, then we can always find an n-vertex tournament with the property S_k as shown in the following theorem.

Theorem 1.5. *If $\binom{n}{k}(1-2^{-k})^{n-k} < 1$, then there is a tournament on n vertices that has the property S_k.*

Proof. Consider a random tournament on the set $V = [n] = \{1, \dots, n\}$. This means that we consider a discrete probability space consisting of $2^{\binom{n}{2}}$ tournaments on the vertex set V where each of these $2^{\binom{n}{2}}$ events has the equal probability.

For every fixed subset $K \in \binom{V}{k}$, let A_K be the event that there is no vertex that has all K as its out-neighbors. For every fixed vertex $v \in V \setminus K$, the probability that v does not have all of K as its out-neighbors is $1 - 2^{-k}$, and all $n - k$ events for distinct vertices v are independent, hence we have $\mathbf{Pr}[A_K] = (1 - 2^{-k})^{n-k}$. Thus

$$\mathbf{Pr}\left[\bigvee_{K \in \binom{V}{k}} A_K\right] \leq \sum_{K \in \binom{V}{k}} \mathbf{Pr}[A_K] \leq \binom{n}{k}(1-2^{-k})^{n-k} < 1.$$

Therefore, with positive probability, no event A_K occurs. This means that there is a tournament on n vertices with the property S_k. □

Let $f(k)$ denote the minimum number of vertices of a tournament that has the property S_k. Using $\binom{n}{k} < (\frac{en}{k})^k$ and $(1-2^{-k})^{n-k} < e^{-(n-k)/2^k}$, we have $f(k) \leq (1+o(1))k^2 2^k \ln(2)$. It is known that $f(k) = \Omega(k \cdot 2^k)$ which is proved by Szekeres.

Let's consider some problems regarding hypergraphs, which is also called set systems.

Definition 1.6. *A hypergraph is a pair $H = (V, E)$ where V is a finite set whose elements are called vertices and E is a family of subsets of V called edges. It is k-uniform if each edge contains k vertices. A k-uniform hypergraph is also called an k-graph.*

An hypergraph H is called intersecting if $A, B \in E(H)$ implies $A \cap B \neq \emptyset$. What would be the maximum number of edges in an n-vertex intersecting k-graph? In other words, we consider $2K_k^{(k)}$ as a hypergraph with two disjoint edges, then how many edges can an n-vertex k-graph with no copies of $2K_k^{(k)}$ have? (Here, $K_n^{(k)}$ denotes the complete k-uniform hypergraph with n vertices) This Turán-type question is completely answered by the following Erdős-Ko-Rado theorem when $n \geq 2k$. Note that a k-uniform hypergraph H on $[n]$ with $E(H) = \{e \in \binom{[n]}{k} : e \ni 1\}$ shows that the bound $\binom{n-1}{k-1}$ is best possible. Here if $n < 2k$, then it is obvious that any n-vertex k-graph is intersecting.

Theorem 1.7 (Erdős-Ko-Rado)**.** *Let $n \geq 2k$. Any n-vertex intersecting k-graph has at most $\binom{n-1}{k-1}$ edges.*

Proof. Let H be an intersecting k-graph on the vertex set $[n]$. Consider the following Claim.

Claim 1. *For $s \in [n]$, let $A_s = \{s, s+1, \ldots, s+k-1\}$ where $n+i = i$ for $i > 0$. Then at most k of the sets $\{A_s : s \in [n]\}$ are edges of H*

Proof. Suppose $A_t \in E(H)$ for some $t \in [n]$. All sets A_s that intersect with A_t can be paired into pairs $\{A_{t-i}, A_{t+k-i}\}$ of two disjoint sets. Hence at most one of such pair are in $E(H)$. This shows that at most k of the sets A_s are edges of H. □

Let σ be a permutation of $[n]$ chosen uniformly at random and let $i \in [n]$ is chosen uniformly at random, where the choices of σ and i are independent. Let $A = \{\sigma(i), \ldots, \sigma(i+k-1)\}$ be the random set obtained from σ and i, where $n+i = i$ for $i > 0$. Conditioning on any choice σ_0, the above lemma gives $\mathbf{Pr}[A \in E(H) : \sigma = \sigma_0] \leq \frac{k}{n}$, hence we have $\mathbf{Pr}[A \in E(H)] \leq \frac{k}{n}$.

But A is chosen uniformly from all k-sets so we have

$$\frac{k}{n} \geq \mathbf{Pr}[A \in E(H)] = \frac{e(H)}{\binom{n}{k}}.$$

This yields $e(H) \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$. □

Definition 1.8. *We say that H has property B or it is 2-colorable if there is a 2-coloring of V such that no edge is monochromatic. Let $m(k)$ be the minimum possible number of edges of an k-uniform hypergraph that is not 2-colorable.*

Proposition 1.9. *Every k-graph with less than 2^{k-1} edges is 2-colorable. Therefore $m(k) \geq 2^{k-1}$.*

Proof. Let $H = (V, E)$ be an k-graph with less than 2^{k-1} edges. Color V randomly by two colors. For each edge $e \in E$, let A_e be the event such that e is monochromatic. Then $\mathbf{Pr}[A_e] = 2^{1-k}$. Therefore

$$\mathbf{Pr}[\bigvee_{e \in E} A_e] \leq \sum_{e \in E} \mathbf{Pr}[A_e] < 1.$$

This shows that there is a 2-coloring without monochromatic edges. □

Later we will improve this lower bound on $m(k)$. For an upper bound, one can simply consider a complete k-graph on $2k-1$ vertices. If we color the vertices of this hypergraph with two colors, then the pigeonhole principle yields a monochromatic edges, so we obtain $m(k) \leq \binom{2k-1}{k} \simeq \frac{4^k}{\sqrt{\pi k}}$.

To get a better upper bound, we change what we randomly choose. Instead of choosing coloring at random, we choose edges at random, and prove that the probability that no coloring works for the resulting hypergraph is positive.

Theorem 1.10. *There exists a k-uniform hypergraph with $(1+o(1))\frac{e \ln 2}{4} k^2 2^k$ edges that is not 2-colorable.*

Proof. Consider k-graphs with m edges and vertex set $[n]$, where n and m will be chosen later to optimize the resulting bound. For a fixed coloring with r points in a color and s in the other color, where $r+s=n$. The probability that a random k-set is monochromatic is $\frac{\binom{r}{k}+\binom{s}{k}}{\binom{n}{k}}$.

This probability is minimized when $r=s$, since $\binom{x}{k}$ is convex function on x. For a coloring $\sigma:[n]\to\{0,1\}$, the probability that a random k-subset of $[n]$ is monochromatic i thus at least $2\binom{n/2}{k}/\binom{n}{k} =: p$. If we select m edges of size k independently and uniformly at random, the probability that none is monochromatic is at most $(1-p)^m$. In other words, the probability that the specified coloring is a proper 2-coloring of the resulting hypergraph is at most $(1-p)^m$.

Since there are 2^n possible colorings and each has probability at most $(1-p)^m$ of being a proper 2-coloring, we have an upper bound of $2^n(1-p)^m$ on the probability that our set of m random edges is 2-colorable. If $2^n(1-p)^m<1$, then some k-uniform hypergraph with n vertices and m edges is not 2-colorable. We seek to minimize m such that $2^n(1-p)^m<1$.

Since $(1-p)\le e^{-p}$, it suffice to have $n\ln 2 - mp<0$. Choose $m=\lceil n\ln 2/p\rceil$.

Now we choose n to minimize n/p. $p=2^{1-k}\prod_{i=0}^{k-1}\frac{n-2i}{n-i}$, then we have

$$\frac{n-2i}{n-i}=1-\frac{i}{n}-O(\frac{i^2}{n^2})=e^{-i/n}+O(\frac{i^2}{n^2}).$$

After summing the exponents, p is asymptotically $2^{1-k}e^{-k^2/(2n)}$ when $k/n\to 0$. So, we now wish to minimize $n2^{k-1}e^{k^2/(2n)}$; calculus tells us to set $n=k^2/2$. With this choice, we obtain $m=(1+o(1))\frac{e\ln 2}{4}k^2 2^k$. □

The probabilistic method can be useful in tackling not only combinatorial problems but also other problems. For example, we can consider the following number theoretic problem.

Definition 1.11. *A subset A of an abelian graph G is called sum-free if $A+A=\{a_1+a_2 : a_1,a_2\in A\}$ does not intersect with A. In other words, there are no triples $a_1,a_2,a_3\in A$ with $a_1+a_2=a_3$.*

Theorem 1.12. *Every set $B=\{b_1,\dots,b_n\}$ of n nonzero integers contain a sum-free subset A with $|A|>\frac{1}{3}n$.*

Proof. Let $p=3k+2$ be a prime such that p is bigger than $|b_i|$ for any $i\in[n]$. We consider the cyclic group $\mathbb{Z}_p$ and let $C=\{k+1,\dots,2k+1\}\subseteq\mathbb{Z}_p$. Note that C is a sum-free subset of $\mathbb{Z}_p$.

We choose an integer $x\in[p-1]$ uniformly at random, and let $d_i\in\mathbb{Z}_p$ such that $d_i\equiv xb_i \pmod p$. As x ranges over all number $1,2,\dots,p-1$, the number d_i also ranges over all nonzero elements of $\mathbb{Z}_p$. Hence

$$\mathbf{Pr}[d_i\in C]=\frac{|C|}{p-1}=\frac{k+1}{3k+1}>1/3.$$

Hence, $C\cap\{d_1,\dots,d_i\}$ has expected size larger than $n/3$. Consequently, there is an $x\in[p-1]$ and a subset A of B of size more than $n/3$ such that $xa(\mathrm{mod}\ p)\in C$ for all $a\in A$. This A is also sum-free as $a_1+a_2=a_3$ implies $xa_1+xa_2\equiv xa_3(\mathrm{mod}\ p)$, contradicting the fact that C is sum-free in $\mathbb{Z}_p$. This provides the desired A. □

One can consider other abelian group rather than integer sets. Alon and Kleitman (1990) proved that every set of n nonzero elements of an arbitrary abelian group contains a sum-free subset with more than $2n/7$ elements and the constant $2/7$ is best possible. The constant $1/3$ in above theorem is also best possible as Eberhard, Green and Manners (2013) proved that it cannot be replaced by $1/3 + \varepsilon$ for any $\varepsilon > 0$.

2. Linearity of Expectation

Definition 2.1. *A function X defined on a (discrete) probability space is called a random variable.*

The range of a random variable is usually $\mathbb{R}$. When the range of the random variable X is $\mathbb{R}$, the expectation $\mathbb{E}[X]$ is $\sum_e X(e)\mathbf{Pr}(e)$ where e runs over all elements of the discrete probability space. Let $X_1, \dots, X_n$ be random variables, then $X = c_1X_1 + \cdots + c_nX_n$ is also a random variable. Linearity of expectation states that

$$\mathbb{E}[X] = c_1\mathbb{E}[X_1] + \cdots + c_n\mathbb{E}[X_n].$$

Also, if $\mathbb{E}[X] = c$ is given, there always exists a choice with $X \geq c$ and also there always exists a choice with $X \leq c$.

2.1. **Max-cut problems.** Consider the following problem: for a given graph G with m edges, what is the maximum number of edges in a bipartite subgraph of G? In other words, what is the maximum number of edges in the induced bipartite graph $G[V_1, V_2]$ where $V_1 \cup V_2$ runs over all partition of $V(G)$? The following theorem shows that we can always find a bipartite subgraph with at least $m/2$ edges.

Theorem 2.2. *Let G be a graph with m edges. Then G contains a bipartite subgraph with at least $m/2$ edges.*

Proof. Let $A \cup B$ be a partition of $V(G)$ obtained as follows: For each $v \in V(G)$, we add v to A or B independently uniformly at random. Call an edge $xy \in E(G)$ crossing if exactly one of x and y belongs to A. Let X be the number of crossing edges, then $X = \sum_{e \in E(G)} X_e$ where X_e is the indicator random variable for xy being crossing. In other words, $X_e = 1$ if e is crossing and $X_e = 0$ otherwise. Then $\mathbb{E}[X_e] = \frac{1}{2}$. Hence the linearity of expectation yields that

$$\mathbb{E}[X] = \sum_{e \in E(G)} \mathbb{E}[X_e] = \frac{m}{2}.$$

Thus there exists a choice of a partition $A \cup B$ of $V(G)$ with at least $m/2$ crossing edges. □

If we consider a complete graph, then we can see that the above theorem is almost best possible, in a sense that we cannot replace $m/2$ with $(\frac{1}{2} + \varepsilon)m$ for any $\varepsilon > 0$. However, can we improve by a sublinear term? We can improve this. Note that if $m = \binom{2n}{2}$, then the K_{2n} shows that the following theorem is best possible.

Theorem 2.3 (Edwards, 1975)**.** *Let G be a graph with m edges. Then G contains a bipartite subgraph with at least $\frac{m}{2} + \frac{-1+\sqrt{8m+1}}{8}$ edges.*

Proof. Let G be a graph with m edges.

Case 1. Let $\chi(G) = 2t$ for some $t \in \mathbb{N}$. Let $V_1 \cup \cdots \cup V_{2t}$ be a partition of $V(G)$ into $2t$ independent sets. We choose $I \in \binom{[2t]}{t}$ uniformly at random. Let $A = \bigcup_{i \in I} V_i$ and $B = V(G) \setminus A$. Let X be the number of crossing edges with respect to the partition

$A \cup B$. Then for each edges e between V_i and V_j, the probability that e is crossing is $\frac{2\binom{2t-2}{t-1}}{\binom{2t}{t}} = \frac{2t^2}{2t(2t-1)} = \frac{t}{2t-1}$. Hence, we have

$$\mathbb{E}[X] = \frac{tm}{2t-1}.$$

Note that an optimal coloring requires at least one edge between every two color classes, hence we have $m \geq \binom{2t}{2}$. This implies that $t \leq \frac{1+\sqrt{8m+1}}{4}$. This shows that

$$\mathbb{E}[X] = \frac{tm}{2t-1} = \frac{m}{2} + \frac{m}{4t-2} \geq \frac{m}{2} + \frac{1+\sqrt{8m+1}}{8}.$$

Case 2. Let $\chi(G) = 2t+1$ for some $t \in \mathbb{N}$. Let $V_1 \cup \cdots \cup V_{2t+1}$ be a partition of $V(G)$ into $2t+1$ independent sets. We choose $I \in \binom{[2t+1]}{t}$ uniformly at random. Let $A = \bigcup_{i \in I} V_i$ and $B = V(G) \setminus A$. Let X be the number of crossing edges with respect to the partition $A \cup B$. Then for each edges e between V_i and V_j, the probability that e is crossing is $\frac{2\binom{2t-1}{t-1}}{\binom{2t+1}{t}} = \frac{t+1}{2t+1}$. Hence, we have

$$\mathbb{E}[X] = \frac{(t+1)m}{2t+1}.$$

Note that an optimal coloring requires at least one edge between every two color classes, hence we have $m \geq \binom{2t+1}{2}$. This implies that $t \leq \frac{-1+\sqrt{8m+1}}{4}$. This shows that

$$\mathbb{E}[X] = \frac{(t+1)m}{2t+1} = \frac{m}{2} + \frac{m}{4t+2} \geq \frac{m}{2} + \frac{-1+\sqrt{8m+1}}{8}.$$

□

Although the above theorem is sharp, it is not easy to find a sharp example when m is some not so nice number. So, Erdős asked whether there are infinite increasing sequence $m_1 < m_2 < \dots$ where any m_i-edge graphs contains max-cut of size more than $\frac{m}{2} + \frac{-1+\sqrt{8m+1}}{8} + f(m_i)$ where $f(x) \to \infty$ as $x \to \infty$. In 1996, Alon proved this with $f(x) = \Omega(x^{1/4})$, and this is sharp as Alon proved that $f(x) = \omega(x^{1/4})$ is impossible. There are some further researches whether one can find a bigger max-cut in graphs with forbidden subgraphs. For example, Alon (1996) proved that if G is a triangle-free graph with m edges, then it has a cut with size at least $\frac{m}{2} + \Omega(m^{4/5})$ and there exists a triangle-free graph with m edges such that every cut of this graph has size at most $\frac{m}{2} + O(m^{4/5})$. Some results are known for H-free graphs for several graphs H.

2.2. **Unbalancing lights.** Let $n \times n$ array of lights be given. Suppose for each row and each column, there is a switch so that if the switch is pulled all of the lights in that line will be switched on to off or off to on. The question is for any initial configuration, how many lights can we turn on? Can we turn on all of them? If all of them is not possible, how much can we do? The following theorem roughly tells how much we can do.

We consider an $n \times n$ matrix with entries $a_{ij} \in \{-1, +1\}$ where 1 indicates on and -1 indicates off. Let $x_i, y_j \in \{-1, +1\}$ be the number indicating whether we activate the switches on i-th row and j-th column,.

Theorem 2.4. *Let $a_{ij} \in \{-1,+1\}$ for $i,j \in [n]$. Then there exists $x_i, y_j \in \{-1,+1\}$ for $i,j \in [n]$ so that*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_iy_j \geq \left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{3/2}.$$

Proof. We first forget $x_1, \ldots, x_n$. We select $y_1, \ldots, y_n$ independently uniformly at random from $\{-1,+1\}$. Let

$$R_i = \sum_{j=1}^{n} a_{ij}y_j, R = \sum_{i=1}^{n} |R_i|.$$

Whatever a_{ij} was, we know that $a_{ij}y_j$ is -1 or $+1$ with probability $1/2$, and they are independent over j. Hence whatever the i-th row initially was, the resulting row $(a_{i1}y_1, a_{i2}y_2, \ldots, a_{in}y_n)$ is uniformly distributed over all 2^n possible rows. So

$$\begin{aligned}\mathbb{E}[|R_i|] &= \sum_{i=0}^{n} |n-2i| \binom{n}{i} 2^{-n} = 2^{1-n} \sum_{i \leq \lfloor (n-1)/2 \rfloor} (n\binom{n-1}{i} - n\binom{n-1}{i-1}) \\ &= n2^{1-n}\binom{n-1}{\lfloor (n-1)/2 \rfloor} = \left(\sqrt{\frac{2}{\pi}} + o(1)\right)\sqrt{n}.\end{aligned}$$

Here, $\binom{n}{i} = 0$ for $i < 0$ or $i > n$. By the linearity of expectation of R, we have

$$\mathbb{E}[R] = \left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{3/2}.$$

This shows that there exists $y_1, \ldots, y_n \in \{-1,+1\}$ with R at least this value. Now we pick x_i with the same sign as R_i, then we have $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_iy_j = \sum_{i=1}^{n} |R_i| \geq \left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{3/2}$. □

However, how tight is this bound? One can show that the above number cannot be bigger than $n^{3/2}$. We know that there exists an $n \times n$ Hadamard matrix A when n is a power of two. This is a matrix where the each column vectors $u_1, \ldots, u_n \in \{1,-1\}^n$ forms an orthogonal basis. Note that if A is Hadamard matrix, then the result of arbitrary row-switchings and column-switchings also yields a Hadamard matrix. On the other hand, we have the following.

Lemma 2.5. *Let A be a Hadamard matrix. Then*

$$\left|\sum_{i,j\in[n]} a_{ij}\right| \leq n^{3/2}.$$

Proof. By Cauchy-Schwarz inequality, we have

$$\begin{aligned}(\sum_{i,j\in[n]} a_{ij})^2 &\leq n\sum_{i\in[n]}(\sum_{j\in[n]} a_{ij})^2 \leq n\sum_{i\in[n]}(n + 2\sum_{j<\ell} a_{ij}a_{i\ell}) \\ &\leq n^3 + 2n\sum_{j<\ell}\sum_{i\in[n]} a_{ij}a_{i\ell} \leq n^3.\end{aligned}$$

Note that the term $\sum_{i\in[n]} a_{ij}a_{i\ell}$ is equal to the dot product of j-th column and ℓ-th column, which is zero as $j < \ell$. This completes the proof. □

Thus, if the initial matrix is Hadamard, no matter how one switches on and off, one cannot do better than $n^{3/2}$.

2.3. **List chromatic number.** Graphs with many edges can have a low chromatic number, e.g a complete bipartite graph. However, this is not true for list chromatic number. To prove lower bound on the list-chromatic number, we need to show that when the lists have a given size there is a list assignment from which no proper coloring can be chosen.

Theorem 2.6 (Alon, 1993)**.** *For some constant $c > 0$, every graph G with average degree d has list chromatic number at least $c\frac{\log d}{\log\log d}$.*

Proof. If suffice to show that $\frac{d}{4} > \binom{s^4}{s}\log(2\binom{s^4}{s})$ implies $\chi_\ell(G) > s$.

Since G has average degree d, it has a subgraph G' with minimum degree at least $d/2$. The subgraph G' in turn has a spanning bipartite subgraph H such that $d_H(v) \geq d_{G'}(v)$ for all $v \in V(G')$, so $\delta(H) \geq d/4$.

We generate a random list assignment L for H with lists of size s. Let A and B be the partite sets of H, with $|A| \geq |B|$. Let $S = [s^4]$ and let $t = \binom{s^4}{s}$. Each vertex receives a random s-subset of S as a list independently, with all t such sets equally likely.

Say that a vertex of A is full if all t possible lists appear on its neighbors. The probability that a particular s-set T fails to appear on the neighbors of x is $(1-1/t)^{d_H(x)}$. Since there are t such sets and $d/4 > t\log(2t)$, we obtain

$$P(x \text{ is not full}) \leq t(1-1/t)^{d/4} < te^{-d/(4t)} < te^{-\log(2t)} = 1/2.$$

Hence for $X =$ the number of full vertices, the expected number $\mathbb{E}[X]$ of full vertices is at least $|A|/2$, and there is some outcome of the random list assignment such that at least $|A|/2$ vertices of A are full. Fix such an assignment.

We now claim that extending this by a random list assignment for A produces with positive probability a list assignment from which no proper coloring can be chosen. Let f be a particular choice of colors from the lists on B. For a full vertex x in A, since all s-sets appear on its neighbors, at most $s-1$ colors fail to be chosen on its neighbors. Hence f can be properly extended to x only if $L(x)$ contains one of these missing colors. There are at most $s-1$ ways to name a usable color, and then $L(x)$ must be filled from the remaining colors, so

$$P(x \text{ can be colored}) \leq \frac{(s-1)\binom{s^4-1}{s-1}}{t} - \frac{s-1}{s^3} < \frac{1}{s^2}.$$

In order to extend f to an L-coloring, all full vertices must be colored, so the probability of extension is bounded by $(1/s^2)^{|A|/2}$, which equals $s^{-|A|}$. Since there are $s^{|B|}$ choices for the coloring f on B from the list assignment on B, the probability that some choice of colors on B extends to an L-coloring is bounded by $s^{|B|}s^{-|A|}$. Since $|A| \geq |B|$, this bound is less than 1. Hence there is some outcome of the random assignment to A such that no proper coloring can be chosen from the lists. □

Alon later improved this to $(\frac{1}{2} - o(1))\ln d$. It is conjectured that for some constant c, every d-regular bipartite graph has choice number at most $c\log d$, but only $O(\frac{d}{\log d})$ is known as an upper bound.

3. Alteration

Often, the expected behavior of random constructions is close to but not quite what we want to prove. For some of such cases, one can make further modifications to improve the obtained random constructions. Consider the following example.

3.1. Ramsey number and hypergraph 2-coloring.

Theorem 3.1. $R(k,k) > (1-o(1))\frac{k2^{k/2}}{e}$.

Proof. Let n be a number which we will determine later. Let X be the random variable counting the number of monochromatic k-cliques in a random 2-coloring of the edges of K_n. Then $X = \sum_C X_C$ where C runs over all k-vertex subset of $V(K_n)$ and X_C is the indicator variable such that $X_C = 1$ if C induces a monochromatic clique and $X_C = 0$ otherwise. Then

$$\mathbb{E}[X] = \sum_C \mathbb{E}[X_C] = \sum_C Pr[X_C = 1] = \binom{n}{k}2^{1-\binom{k}{2}}.$$

Then, there exists a coloring with at most this many monochromatic k-cliques. We delete a vertex of each monochromatic k-clique in such a coloring, we retain a graph with at least $n - \binom{n}{k}2^{1-\binom{k}{2}}$ vertices but no monochromatic k-cliques.

Thus,

$$R(k,k) \geq n - \binom{n}{k}2^{1-\binom{k}{2}} \geq n - (\frac{ne}{k})^k 2^{1-\binom{k}{2}}.$$

We seek n to maximize this bound. Differentiate this to see that choosing n so that $1 = k\frac{e}{k}(\frac{ne}{k})^{k-1}2^{1-k(k-1)/2}$, so we set $n = e^{-1}k2^{k/2}(2e)^{-1/k}$. The factor $(2e)^{-1/k}$ is near 1 when k is large, so we don't gain much from exact maximizing value of n.

So, let n be an integer nearby $e^{-1}k2^{k/2}$ then we have

$$n - (\frac{ne}{k})^k 2^{1-\binom{k}{2}} \geq \frac{1}{e}k2^{k/2}(1 - \frac{2e}{k}).$$

Since $2e/k$ tends to 0 for large k, we obtain the claimed bound. □

Here, we consider a random objects which has small imperfections, and we correct those imperfections. This is called 'alteration method' We consider another example of alteration method.

Definition 3.2. *A set $S \subseteq V(G)$ is dominating if every vertex outside S has a neighbor in S*

Theorem 3.3. *For $k > 1$, every n-vertex graph with minimum degree k has a dominating set of size at most $(\frac{1+\ln(k+1)}{k+1})n$.*

Proof. Form a random vertex subset S in such a graph by including each vertex independently with probability $p = \frac{\ln(k+1)}{k+1}$. Given S, let T be the set of vertices outside S having no neighbors in S. Adding T to S yields a dominating set. We seek the expected number of $|S \cup T|$.

Since each vertex appears in S with probability p, linearity yields $\mathbb{E}[S] = np$. The random variable $|T|$ is the sum of n indicator random variables for whether individual vertices belong to T. We have $v \in T$ if and only if v and its neighbors all fail to be in S. This

has probability at most $(1-p)^{k+1}$ since v has degree at least k. Since $(1-p)^{k+1} < e^{-p(k+1)}$, we have

$$\mathbb{E}[|S| + |T|] \leq np + ne^{-p(k+1)} = (\frac{1+\ln(k+1)}{k+1})n.$$

Hence, there exists a choice of S ensuring $|S| + |T| \leq (\frac{1+\ln(k+1)}{k+1})n$ which completes the proof. □

The following lemma regarding the expectation and the probability will be useful in many instances.

Lemma 3.4 (Markov's inequality)**.** *If X is a discrete random variable, then $Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$. Thus $\mathbb{E}[X] \to 0$ implies $P[X = 0] \to 1$.*

Proof. $\mathbb{E}[X] = \sum_{k\geq 0} kPr[X = k] \geq t\sum_{k\geq t} Pr[X = k] = tP[X \geq t]$. □

Now we use the alteration method to prove the following theorem.

Theorem 3.5. *Given $k \geq 3$ and $g \geq 3$, there exists a graph with girth at least g and chromatic number at least k.*

Proof. We generate graphs with vertex set $[n]$ by letting each pair be an edge with the probability p, independently. We show that for large p it has no large independent set, and use $\chi(G) \geq n/\alpha(G)$ to prove that $\chi(G)$ is large. We also show that for small p, the expected number of short cycles are small. If we choose right p, then we obtain a graph with large chromatic number, and we delete some vertices to delete all short cycles in there.

Let $p = n^{t-1}$, where $t = 1/g$. Since there are at most $n^j/(2j)$ possible j-cycles and each cycle appear with probability p^j, the total number X of cycles of length less than g has expectation

$$\mathbb{E}[X] = \sum_{i=3}^{g-1} \frac{n^j}{2_i} p^i \leq \sum_{i=3}^{g-1} \frac{n^{ti}}{2i}.$$

Since $t = 1/g$ and g is fixed, we have $\mathbb{E}[X] < gn^{1-1/g}$. By Markov's inequality, we can now conclude that $Pr[X \geq n/2] \to 0$ as $n \to \infty$. For n large enough, we have $Pr[X \geq n/2] < 1/2$.

Since $\alpha(G)$ cannot grow when we delete vertices, at least $(n - X)/\alpha(G)$ independent sets are needed to color the vertices remaining when we delete a vertex of each cycle. If $X < n/2$ and $\alpha(G) \leq n/(2k)$, then at least k colors are needed for the graph remaining. With $r = \lceil \frac{3\ln n}{p} \rceil$, we have

$$Pr[\alpha(G) \geq r] \leq \binom{n}{r}(1-p)^{\binom{r}{2}} < (ne^{-p(r-1)/2})^r.$$

This tends to 0 as $n \to \infty$.

Since $r = \lceil 3n^{1-t} \ln n \rceil$ and k is fixed, we can choose n large enough to obtain $r < n/(2k)$. For large enough n, we have

$$Pr[X \geq n/2] < 1/2 \text{ and } Pr[\alpha(G) \geq r) < 1/2.$$

Then there exists an n-vertex graph G with $\alpha(G) \leq n/(2k)$ and it has fewer than $n/2$ cycles of length less than g. We delete a vertex from each short cycle and obtain a graph with girth at least g and chromatic number at least k. □

Let's consider another example of 2-coloring of hypergraphs. We have proved that $2^{k-1} \leq m(k) \leq O(k^2 2^{k-1})$. Beck 1978 improved the lower bound to $\Omega(k^{1/3}2^k)$ and Radhakrishnan and Srinivasan 2000 further improved this to $\Omega(2^k(k/\log k)^{1/2})$.

Again to show the lower bound, we need to find a way to color edges of a hypergraph. However, if we just randomly color, then each edge become monochromatic with probability $\frac{1}{2^{k-1}}$. So simple random coloring would not work. We need to alter the random coloring to obtain a desired 2-coloring.

Theorem 3.6 (Radhakrishnan and Srinivasan, 2000)**.** *If there exists $p \in [0,1]$ with $s(1-p)^k + s^2 p < 1$, then $m(k) > s2^{k-1}$.*

Proof. This proof is by Cherkashin and Kozik 2015.

For our convenience, we consider the following rather continuous time framework. For each vertex $v \in V$, let x_v be a real number in $[0,1]$ chosen uniformly at random, we call this the label of vertex v. Note that all labels x_v are distinct for distinct vertices with probability 1. This gives an ordering of the vertices according to the values of x_v. For an edge e, we say that $v \in e$ is the last vertex of e if x_v is bigger than all x_u with $u \in e - \{v\}$. Let

$$L = [0, \frac{1-p}{2}), M = [\frac{1-p}{2}, \frac{1+p}{2}), R = [\frac{1+p}{2}, 1]$$

be three subintervals of $[0,1]$.

If x_v is in $L \cup M$, then we color v blue and if x_v is in R, then we color v red. As p is positive, it is likely that we can avoid red monochromatic edges, while obtaining many blue monochromatic edges. Now we recolor some vertices to destroy all blue edges as follows:

for any vertex v with $x_v \in M$, if v is the last vertex of a blue monochromatic edge e, then we recolor v to red. (3.1)

Now, we compute the expected number of monochromatic edges. There are two types of monochromatic edges, the edges e whose labels all lie in L or R, or the edges e which becomes red monochromatic only after recoloring.

For the first type, the expected number of such edges is $2e(H)(\frac{1-p}{2})^k \leq s(1-p)^k$.

For the second type, the first vertex of e must be blue before recoloring but becomes red after recoloring. Hence, such an edge e belongs to a pair (f, e) of edges of H where $e \cap f$ is a single vertex v and $x_v \in M$ and v is the last vertex of f and first vertex of e. For such a pair (e, f), assuming $x_v \in M$ is chosen, the probability of the above event happening is that $x_v^{k-1}(1-x_v)^{k-1} \leq (\frac{1}{4})^{k-1}$. As there are at most $s^2 4^{k-1}$ such pairs (e, f) and $x_v \in M$ happens with the probability p, the expected number of monochromatic edges of second type is at most

$$s^2 4^{k-1} \cdot p(\frac{1}{4})^{k-1} \leq s^2 p.$$

Hence, the expected number of monochromatic edges is at most

$$s(1-p)^k + s^2 p < 1.$$

This shows that the above algorithm produces a coloring with no monochromatic edges with positive probability. □

Using this, one can show the following corollary as desired.

Corollary 3.7. $m(k) = \Omega(2^k(k/\log k)^{1/2})$.

Proof. As $(1-p) \le e^{-p}$, we have $s(1-p)^k + s^2p \le se^{-pk} + s^2p$. The right hand side is minimized at $p = \frac{\ln(k/s)}{k}$. We substitute this back in then we have $\frac{s^2}{k}(1+\ln(k/s))$. To make this smaller than 1, let $s = (k/\ln k)^{1/2}$ with sufficiently large k, then we have $s(1-p)^k + s^2p < 1$. With this, we can apply the previous theorem to finish the proof. □

3.2. **dependent random choice.** In undergraduate graph theory, we learned about Turán's theorem and Kövári-Sós-Turán theorem. These theorems estimate the maximum number of edges in a graph without certain subgraphs. In other words, this theorem concerns about finding certain subgraphs in a graph with many edges. This problem becomes difficult especially when the graph we wish to find is bipartite. In this subsection, we learn some tools that we can use. We want to extend the Kövári-Sós-Turán Theorem into the following.

Theorem 3.8. *Let $s \in \mathbb{N}$ and let H be a bipartite graph with vertex partition A, B such that all vertices in B has degree at most s. Then there exists $c = c(H)$ such that $ex(n, H) \le cn^{2-1/s}$.*

A very natural way to find a graph H into G is the following. We order the vertices in A into $(x_1, \dots, x_h)$ and we map x_i into a vertex $\phi(x_i) \in V(G)$ one by one. While doing this, we make sure that for all $y \in B$, the common neighborhood of vertices in $\phi(N_H(y))$ is large. Once we map all vertices of x in this way, there are many choices of vertices in G for y to embed. Hence, we can embed vertices in B one by one into different vertices. Hence, the ideal situation is when we have a set $U \subseteq V(G)$ of vertices which satisfies the following property for some large r, m. Then we can freely embed vertices in A into U.

Every r vertices of U have at least m common neighbors.

If $r, m > h$, then we can arbitrarily embed each x_i into U to obtain an embedding of H into G. How can we obtain such a set U? We can find such a set using alteration method.

Lemma 3.9. *Let $a, d, m, n, r \in \mathbb{N}$. Let G be an n-vertex graph with $d(G) = d$. If there exists $t \in \mathbb{N}$ satisfying*

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

then there exists a subset $U \subseteq V(G)$ with $|U| \ge a$ such that every r vertices in U have at least m common neighbors in G.

Proof. We will randomly choose vertices, and prove that chosen vertices have the desired property with positive probability. Let $N_G^*(W) := \bigcap_{x \in W} N_G(x)$.

Assume that we have for two sets W_1 and W_2 of size r with $|N^*(W_1)| > |N^*(W_2)|$. In order to increase the probability of obtaining a desired set, we want W_1 to be more likely

to be included in U than W_2 (as our goal is to make all r-subsets of U to have many common neighbors). For this, we can randomly choose some vertex, say v, and we let U be the neighborhood $N_G(v)$. In this way, we can ensure that W_1 is more likely to be included in U than W_2 as

$$\mathbf{Pr}[W_1 \subseteq U] = \mathbf{Pr}[v \in N^*(W_1)] = \frac{|N^*(W_1)|}{n} > \frac{|N^*(W_2)|}{n} = \mathbf{Pr}[v \in N^*(W_2)] = \mathbf{Pr}[W_2 \subseteq U].$$

Now we start the proof. For each $i \in [t]$, we choose a random vertex $v_i \in V(G)$ independently uniformly at random. Note that two vertex v_i and v_j may be the same as we choose independently. Let $A = N^*(\{v_1, \dots, v_t\})$ and $X = |A|$ be the random variable denoting the size of A. Linearity of expectation implies

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

Here, we obtain the penultimate inequality by the convexity of the function $z \to z^t$.

Let

$$\mathcal{R} := \{R \in \binom{V(G)}{r} : |N^*(R)| \leq m\}.$$

Let Y be the random variable counting the number of subsets $R \subseteq A$ of size r such that $|N^*(R)| \leq m$. For given $R \in \mathcal{R}$, we have $\mathbf{Pr}[R \subseteq A] = (\frac{|N^*(R)|}{n})^t$. Thus

$$\mathbb{E}[Y] = \sum_{R \in \mathcal{R}} (\frac{|N^*(R)|}{n})^t \leq \binom{n}{r} (\frac{m}{n})^t.$$

By linearity of expectation,

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

This implies that there exists a choice of $v_1, \dots, v_t$ which yields a set A satisfying $X - Y \geq a$. Consider such a set A and delete one vertex from each subset $R \in \mathcal{R}$ lying inside A. Let U be the set of remaining vertices, then U has size at least $X - Y \geq a$, and U is our desired subset. □

Observe that we choose U in a dependent way. By choosing something else, and the choice of U depends from the earlier choice. Hence we call it *dependent random choice*. By this, we achieve that certain sets will be included in U more likely that some other sets. By using this lemma, we can prove Theorem 3.8.

Proof of Theorem 3.8. Let $a := |A|, b := |B|, m := a + b, d := 2cn^{1-1/s}$ and $c \geq 3ma$. Suppose that G is an n-vertex graph with $e(G) \geq cn^{2-1/s}$, hence $d(G) \geq d$. We have

$$\frac{d^s}{n^{s-1}} - \binom{n}{s} (\frac{m}{n})^s \geq 2c^s - \frac{n^s}{s!} (\frac{m}{n})^s \geq c^s \geq a.$$

Thus Lemma 3.9 implies that there exists a set U with $|U| = a$ such that any s vertices in U has at least m common neighbors.

We take an arbitrary injective map $\phi : A \to U$. Let $B = \{y_1, \dots, y_b\}$. We embed $y_1, \dots, y_b$ one by one in order. Right before we embed y_i, we have $|N^*_G(\phi(N_H(y_i)))| \geq m = a + b$, thus there exists a vertex $v_i \in N^*_G(\phi(N_H(y_i))$ which is not an image of any

vertices in $A \cup \{y_1, \ldots, y_{i-1}\}$. We embed y_i to v_i. By repeating this, we obtain a copy of H in G. This prove the theorem. $\square$

4. The second moment

So far, many of our proof concerns about expectation $\mathbb{E}[X]$ of certain random variable X and show that it is possible to achieve as good as its expectation. Often times, this is not good enough and we want to prove that X is actually very close to $\mathbb{E}[X]$ with high probability (which is called 'concentration'). Indeed, we know one concept which measures how spread out X is from its expectation.

Definition 4.1. $\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$ *is the variance of* X.

As $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, the study of $\mathrm{Var}[X]$ is essentially same as the study about $\mathbb{E}[X^2]$, which is called the second moment of the random variable X. For given random variable we often write $\mu = \mathbb{E}[X], \sigma^2 = \mathrm{Var}[X]$.

The following is a basic concentration theorem regarding the second moment. Note that for a nonnegative random variable X and $a > 0$, $\mathbf{Pr}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$ holds and this is called Markov's inequality.

Theorem 4.2 (Chebyshev's inequality)**.** *For any* $\lambda > 0$*, we have*

$$\mathbf{Pr}[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}.$$

Proof. By the definition of expectation, we have

$$\sigma^2 = \mathrm{Var}[X] = \mathbb{E}[(X - \mu)^2] \geq \lambda^2\sigma^2\mathbf{Pr}[|X - \mu| \geq \lambda\sigma].$$

This proves the theorem. □

Using Chebyshev's inequality is called the second moment method. Chebyshev's inequality is best possible by considering the case where X is μ or $\mu + \lambda\sigma$ or $\mu - \lambda\sigma$ with probability $1 - 1/\lambda^2, 1/(2\lambda^2), 1/(2\lambda^2)$. However, in certain cases, we can obtain a better bound, which we will study later.

As before, we will frequently consider some random variable X which is sum of several simpler random variables $X_1, \ldots, X_m$. Then we have

$$\mathrm{Var}[X] = \sum_{i \in [m]} \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j],$$

where we define $\mathrm{Cov}[X_i, X_j]$ as follows.

Definition 4.3. *The covariance* $\mathrm{Cov}[Y, Z]$ *is* $\mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]$.

4.1. An application in Number theory.

Let's consider some applications of the second moment method. Let $\nu(n)$ be the number of prime numbers p dividing n. Hardy and Ramanujan in 1920 showed that $\nu(n)$ is close to $\ln\ln n$ for almost all n. Later Turán gave the following much simpler proof using the probabilistic method.

Theorem 4.4. *Let* $\omega(n) \to \infty$ *arbitrarily slowly. Then the number of* x *in* $[n]$ *such that*

$$|\nu(x) - \ln\ln n| > \omega(n)\sqrt{\ln\ln n}$$

is $o(n)$.

Proof. Let x be a number in $[n]$ chosen uniformly at random. For a prime p, let

$$X_p = \begin{cases} 1 & \text{if } p \mid x \\ 0 & \text{otherwise.} \end{cases}$$

Let

$$M = n^{1/10} \quad \text{and} \quad X = \sum_{p \le M, p \text{ prime}} X_p.$$

Then every $x \in [n]$ has at most 10 prime factors larger than M, we have $\nu(x) - 10 \le X(x) \le \nu(x)$, so we can focus on estimating the value of $X(x)$ rather than ν. First, we estimate the expectation of X. By the linearity of expectation,

$$\mathbb{E}[X] = \sum_p \mathbb{E}[X_p] = \sum_{p \le M} \frac{\lfloor n/p \rfloor}{n} = \sum_{p \le M} (\frac{1}{p} + O(\frac{1}{n})) = \ln\ln n + O(1).$$

Here, we used the fact that $\sum_{p \le x} 1/p = \ln\ln x + O(1)$.

Now we want to estimate the variance to show the concentration. As $\mathrm{Var}[X_p] = (1/p)(1-1/p) + O(1/n)$ and $\sum_{x \in \mathbb{N}} 1/x^2 = O(1)$ we have

$$\sum_{p \le M} \mathrm{Var}[X_p] = \sum_{p \le M} \frac{1}{p} + O(1) = \ln\ln n + O(1).$$

Also, for two distinct primes p, q, we have $X_p X_q = 1$ if and only if $pq \mid x$, hence

$$\begin{aligned} \mathrm{Cov}[X_p, X_q] = \mathbb{E}[X_p X_q] - \mathbb{E}[X_p]\mathbb{E}[X_q] &= \frac{\lfloor n/(pq)) \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n}\frac{\lfloor n/q \rfloor}{n} \\ &= (\frac{1}{pq} \pm \frac{1}{n}) - (\frac{1}{p} \pm \frac{1}{n})(\frac{1}{q} \pm \frac{1}{n}) = 0 \pm \frac{3}{n}(\frac{1}{p} + \frac{1}{q}). \end{aligned}$$

Thus, we have

$$|\sum_{p \ne q \le M} \mathrm{Cov}[X_p, X_q]| \le \frac{3}{n} \sum_{p \ne q} (\frac{1}{p} + \frac{1}{q}) \le \frac{6M}{n} \sum_{p \le M} \frac{1}{p} \le O(n^{-9/10} \ln\ln n) = o(1).$$

Hence, $\mathrm{Var}[X] = \ln\ln n + O(1)$, and Chebyshev's inequality gives that for any $\lambda > 0$, we have

$$\mathbf{Pr}\left[|X - \ln\ln n| > \lambda\sqrt{\ln\ln n}\right] < \lambda^{-2} + o(1).$$

Taking $\lambda = \omega(n)$ and using the fact that X and ν has difference at most 10, we conclude the theorem. □

Let's consider one more result.

Definition 4.5. *A set $\{x_1, \dots, x_k\}$ of positive integers have distinct sums if all sums $\sum_{i \in S} x_i$ are distinct over all subsets $S \subseteq [k]$. Let $f(n)$ be the largest k such that there exists a set $\{x_1, \dots, x_k\} \subseteq [n]$ with distinct sums.*

One example of a set with distinct sums is $\{1, 2, 2^2, \dots, 2^{\lfloor \log_2 n \rfloor}\}$ which shows $f(n) \ge 1 + \lfloor \log_2 n \rfloor$. Erdős offered \$300 for a proof or disproof that $f(n) \le \log_2 n + C$ for some constant C. As all $2^{f(n)}$ sums must distinct and less than $nf(n)$, we have $2^{f(n)} \le nf(n)$, so we have $f(n) < \log_2 n + \log_2\log_2 n + O(1)$. We can use the second moment method to prove the following.

Theorem 4.6. $f(n) \le \log_2 n + \frac{1}{2}\log_2\log_2 n + O(1)$.

Proof. Fix a set $\{x_1, \ldots, x_k\}$ having distinct sums. Let $\varepsilon_1, \ldots, \varepsilon_k$ be independent random variables having values 0 or 1 with probability 1/2. Let $X = \sum_{i\in[k]} \varepsilon_i x_i$ be our random variable. Then

$$\mu = \mathbb{E}[X] = \frac{1}{2}\sum_{i\in[k]} x_i \text{ and } \sigma^2 = \mathrm{Var}[X] = \frac{1}{4}\sum_{i\in[k]} x_i^2 \leq \frac{n^2 k}{4}.$$

By Chebyshev's inequality, for any $\lambda > 1$, we have

$$\mathbf{Pr}[|X - \mu| < \lambda n\sqrt{k}/2] \geq 1 - \frac{1}{\lambda^2}.$$

However, as $\{x_1, \ldots, x_k\}$ have distinct sums, for any value t, $\mathbf{Pr}[X = t] \in \{0, 2^{-k}\}$. Thus

$$1 - \frac{1}{\lambda^2} \leq \mathbf{Pr}[|X - \mu| < \lambda n\sqrt{k}/2] \leq 2^{-k}(\lambda n\sqrt{k} + 1)$$

This implies $n \geq \frac{2^k(1-\lambda^{-2})-1}{\sqrt{k}\lambda}$. Putting $\lambda > 1$ implies that $f(n) \leq \log_2 n + \frac{1}{2}\log_2\log_2 n + O(1)$. □

4.2. **Random graphs.**

Definition 4.7. *The binomial random graph model $G(n,p)$ (it is also called the Erdős-Rényi random graph model) is a probability space over the set of graphs on the vertex set $[n]$ determined by $\mathbf{Pr}[ij \in G] = p$ with these events mutually independent.*

Definition 4.8. *If a property P holds for $G(n,p)$ with probability $1 - o(1)$ where $o(1)$ term tends to zero as n tends to infinity, then we say that $G(n,p)$ satisfies P with high probability (or asymptotically almost surely). We sometimes write whp or a.a.s instead.*

Definition 4.9. *A function $r(n)$ is a threshold function for some property P, if whenever $p = p(n)$ satisfies $p(n)/r(n) \to 0$, then the probability that $G(n,p)$ satisfies P tends to 0 and whenever $p(n)/r(n) \to \infty$ then the probability that $G(n,p)$ satisfies P tends to 1.*

When p is small, say $o(1/n)$, $G(n,p)$ is likely to be a forest, lacking of any subgraphs. But when p is a positive constant, then $G(n,p)$ is likely to be almost complete, containing all small subgraphs. For a specific small graph H, one can ask at which point of p, does $G(n,p)$ starts to have H as a subgraph. To investigate this, we define the following concept first.

Definition 4.10. *Let H be a graph with n vertices and e edges. We call $\rho(H) = e/v$ the density of H. We call H balanced if every subgraph H' has $\rho(H') \leq \rho(H)$. We call H strictly balanced if every proper subgraph H' satisfies $\rho(H') < \rho(H)$.*

We wish to show that if H is balanced, then $n^{-v/e}$ is the threshold function for having H as a subgraph. For this we prove the theorem below.

We will use the Second moment method to prove the concentration of X which is a random variable counting the subgraphs isomorphic to H. Note that Chebyshev's inequality says that for any $\varepsilon > 0$,

$$\mathbf{Pr}[|X - \mathbb{E}[X]| \geq \varepsilon\mathbb{E}[X]] \leq \frac{\mathrm{Var}[X]}{\varepsilon^2\mathbb{E}[X]^2}.$$

Hence, often times, in order to prove concentration, it suffices to show that $\mathrm{Var}[X] = o(\mathbb{E}[X]^2)$ holds.

Theorem 4.11. *Let H be balanced with v vertices, e edges and a automorphisms. Let X be the number of copies of H in $G(n,p)$. If $pn^{v/e} = o(1)$, then $X = 0$ whp, and if $pn^{v/e} = \omega(1)$ then we have $X = (1+o(1))\frac{n^v p^e}{a}$ whp.*

Proof. Let $V(H) = [v]$ and for each ordered tuples $(x_1, \dots, x_v)$ of vertices in $G = G(n,p)$, let $A_{x_1,\dots,x_v}$ be the event that $x_1, \dots, x_v$ provides a copy of H in that order. Let $I_{x_1,\dots,x_v}$ be the indicator variable for $A_{x_1,\dots,x_v}$. We say that $\mathbf{x} = (x_1, \dots, x_v)$ and $\mathbf{y} = (y_1, \dots, y_v)$ are equivalent if $y_{\sigma(i)} = x_i$ for some automorphism σ of H. Then $X = \sum I_{x_1,\dots,x_v}$ counts the number of copies of H in G where the sum is taken for all equivalence class defined above. Then it is easy to see

$$\mathbb{E}[X] = \frac{n(n-1)\dots(n-v+1)p^e}{a} = (1+o(1))\frac{n^v p^e}{a}.$$

If $pn^{v/e} = o(1)$, then this is $o(1)$ implying that $X = 0$ with probability $1 - o(1)$. Assume $\mathbb{E}[X] = \omega(1)$.

Claim 2. *If $\mathbb{E}[X] = \omega(1)$, then $\mathrm{Var}[X] = o(\mathbb{E}[X])^2$.*

Proof. For any indicator variables Y, Z, we have

$$\mathrm{Var}(Y) = \mathbf{Pr}[Y=1]\mathbf{Pr}[Y=0] \le \mathbf{Pr}[Y=1] = \mathbb{E}[Y], \text{ and}$$

$$\mathrm{Cov}(Y,Z) \le \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z] \le \mathbb{E}[YZ] = \mathbf{Pr}[Y \wedge Z].$$

Also, if $|\{x_1, \dots, x_v\} \cap \{y_1, \dots, y_v\}| \le 1$, then $A_{\mathbf{x}}$ and $A_{\mathbf{y}}$ are independent, so their covariance is zero. We write $\mathbf{x} \sim \mathbf{y}$ if their intersection has size at least two. Hence,

$$\begin{aligned}\mathrm{Var}[X] &= \sum \mathrm{Var}[I_{\mathbf{x}}] + \sum_{\mathbf{x}\sim\mathbf{y}} \mathrm{Cov}[I_{\mathbf{x}}, I_{\mathbf{y}}] \\ &\le \mathbb{E}[X] + \sum_{\mathbf{x}} \sum_{\mathbf{y}\sim\mathbf{x}} \mathbf{Pr}[A_{\mathbf{x}} \wedge A_{\mathbf{y}}] = \mathbb{E}[X] + \sum_{\mathbf{x}} \mathbf{Pr}[A_{\mathbf{x}}] \sum_{\mathbf{y}\sim\mathbf{x}} \mathbf{Pr}[A_{\mathbf{y}} \mid A_{\mathbf{x}}].\end{aligned}$$

Note that the events $A_{\mathbf{x}}$ are symmetric in the following sense: for two $\mathbf{x}, \mathbf{y}$ there exists a measure-preserving map of the underlying probability space that permute the events and send $A_{\mathbf{x}}$ to $A_{\mathbf{y}}$. Hence, to estimate the last term, we only have to compute

$$\Delta^* = \sum_{\mathbf{y}\sim\mathbf{x}} \mathbf{Pr}[A_{\mathbf{y}} \mid A_{\mathbf{x}}].$$

There are $v!/a$ terms with $\{y_1, \dots, y_v\} = \{x_1, \dots, x_v\}$ and each of them contributes at most 1 to Δ^*. Assume $\{y_1, \dots, y_v\} \cap \{x_1, \dots, x_v\} = S$ has i elements with $2 \le i \le v-1$. Then as H is balanced, at most ie/v edges of H corresponding to $A_{\mathbf{y}}$ lies inside S, and $e - (ie/v)$ of them does not lie in $\{x_1, \dots, x_v\}$. Hence $\mathbf{Pr}[A_{\mathbf{y}} \mid A_{\mathbf{x}}] = O(p^{e-(ie/v)})$. As there are $O(n^{v-i})$ tuples $\mathbf{y}$ intersecting with $\mathbf{x}$ at i vertices, so

$$\Delta^* = \sum_{i=2}^{v-1} O(n^{v-i}p^{e-(ie/v)}) = o(n^v p^e) = o(\mathbb{E}[X]).$$

Thus

$$\mathrm{Var}[X] \le \mathbb{E}[X] + \sum_{\mathbf{x}} \mathbf{Pr}[A_{\mathbf{x}}] o(\mathbb{E}[X]) \le o(\mathbb{E}[X]^2).$$

We obtain the final inequality as $\mathbb{E}[X] = (\frac{1}{a} + o(1))n^v p^e$ tends to infinity as n grows. □

By applying Chebyshev's inequality, for any $\varepsilon > 0$, we have

$$\mathbf{Pr}[|X - \mathbb{E}[X]| \geq \varepsilon \mathbb{E}[X]] \leq \frac{\mathrm{Var}[X]}{\varepsilon^2 \mathbb{E}[X]^2}.$$

As $\mathrm{Var}[X] = o(\mathbb{E}[X]^2)$, this probability can be arbitrary small when n is sufficiently large. This proves the theorem. □

4.3. **The Rödl nibble.** For a given hypergraph H and a set $S \subseteq V(H)$, we write $d_H(S)$ to denote the number of edges of H containing S. Let $\Delta_i(H) = \max_{S \in \binom{V(H)}{i}} d_H(S)$. We say $\Delta(H) = \Delta_1(H)$ be the maximum degree of H and $\Delta_2(H)$ be the maximum co-degree of H.

Definition 4.12. *A matching in a hypergraph H is a collection of vertex-disjoint edges.*

In this subsection, we want to prove the theorem below, which is a refinement of Rödl nibble. Here, when we write "a statement holds if $0 < x \ll y, z < 1$" this means that there exists some function f such that "a statement holds if $x < f(y, z)$". We will not specify this function f explicitly. Note that with this definition, "if $0 < x \ll y, z$, then a statement holds" is equivalent to the sentence "for given y, z there exists x_0 such that the statement holds for all $0 < x < x_0$".

Theorem 4.13 (Pippenger)**.** *Let k, D be integers and $\varepsilon, \delta \in \mathbb{R}$ such that $0 < 1/D, \delta \ll \varepsilon, 1/k \leq 1$. Let H be an n-vertex k-graph satisfying the following.*

(1) (Almost regular) All vertices of H has degree $(1 \pm \delta)D$.
(2) (Small codegree) $\Delta_2(H) < \delta D$

Then H contains a matching of size at least $(1 - \varepsilon)n/k$.

This theorem has surprisingly many applications. One of the application is the following.

Definition 4.14. *(n, ℓ, k)-block design is a collection $\mathcal{L}$ of ℓ-sets in $[n]$ satisfying the following: For any $K \in \binom{[n]}{k}$, there is exactly one ℓ-set $L \in \mathcal{L}$ containing K.*

For example, $(n, 3, 2)$-block design is called Steiner triple system. This is equivalent to finding an edge-decomposition of K_n into triangles, which exists if and only if n is 1 or 3 modulo 6. It is easy to see that this is necessary by counting the edges and considering the degree of the complete graph.

In general, (n, ℓ, k)-block design is an edge-decomposition of the complete k-graph $K_n^{(k)}$ into copies of complete k-graph $K_\ell^{(k)}$. One big conjecture was whether (n, ℓ, k)-block design exists when the necessary divisibility condition holds. Recently Keevash proved that when the divisibility condition holds and n is sufficiently larger than k, ℓ, then design exists.

However, before Keevash proved this, this problem was open for more than 100 years. In 1963, Erdős and Hanani conjectured an approximate version of this, stating that for any $\varepsilon > 0$, if n is sufficiently large then an ε-approximate decomposition exists. Here ε-approximate decomposition of $K_n^{(k)}$ into $K_\ell^{(k)}$ is a collection of edge-disjoint copies of $K_\ell^{(k)}$ which covers at least $(1 - \varepsilon)$-fraction of the host hypergraph.

Theorem 4.15 (Rödl)**.** *For $\ell \geq k$ and $\varepsilon > 0$, if n is sufficiently large, then ε-approximate decomposition of $K_n^{(k)}$ into $K_\ell^{(k)}$ exists.*

Proof. Let $r = \binom{\ell}{k}$ and H be the r-graph whose vertices are all k-sets in $[n]$ and whose edges are all $\binom{\ell}{k}$ k-sets that lie in an ℓ-set. Then each vertex of H has degree $D = \binom{n-k}{\ell-k}$ and every two distinct vertices has codegree at most $\binom{n-k-1}{\ell-k-1} = o(D)$. As n is sufficiently large, Theorem 4.13 yieds that H has a matching of size at least $(1-\varepsilon)\binom{n}{k}/\binom{\ell}{k}$, which yields an ε-approximate decomposition of $K_n^{(k)}$ into copies of $K_\ell^{(k)}$. □

In order to prove that $H = H_0$ has a large matching, we use the following strategy.

(1) We choose edges of H_0 independently at random with probability $p = \alpha/D$.
(2) We throw away all vertices inside a chosen edge, and let the remaining vertices induces a hypergraph H_1.
(3) If some of the chosen edges has a common vertex, we discard them so that remaining edges form a matching E_0 disjoint from $V(H_1)$.
(4) Prove that H_1 is also almost regular and having small codegree. We repeat this for $H_1, H_2, \dots$.

For this purpose, we prove the following lemma which allow us to proceed one step of the above iteration. Note that 'almost regularity' splits into two weaker conditions, which makes our technical computations easier.

Lemma 4.16. *Assume $0 < 1/D, \delta_0 \ll \delta_1, \alpha, 1/K, 1/k < 1$. Let H be an n-vertex k-graph satisfying the following.*

(H1) For all $x \in V(H)$ but at most $\delta_0 n$ of them has degree $(1 \pm \delta_0)D$.
(H2) $\Delta(H) < KD$.
(H3) $\Delta_2(H) < \delta_0 D$.

Then H has a set M' of edges with the following properties where $V' = V - \bigcup_{e\in M'} e$ and $H' = H[V']$.

(M1) $|M'| = (1 \pm \delta_1)\frac{\alpha n}{k}$.
(M2) $|V'| = (1 \pm \delta_1)ne^{-\alpha}$.
(M3) All vertices $x \in V'$ but at most $\delta_1|V'|$ of them has degree $d_{H'}(x) = (1\pm\delta_1)De^{-\alpha(k-1)}$.

Let's first see how this lemma proves what we want.

Proof of Theorem 4.13. Choose α with $\alpha \ll \varepsilon, 1/k, 1/K < 1$ and let t be an integer with $e^{-\alpha t} < \alpha$ and choose $\delta_0, \dots, \delta_t$ and D be a large enough number so that the following holds

$$1/D, \delta_0 \ll \delta_1 \ll \cdots \ll \delta_{t+3} \ll \alpha \ll \varepsilon, 1/k, 1/K < 1.$$

Let $H_0 = H$ and let $D_i = De^{-\alpha(k-1)i}$ and $n_i = ne^{-\alpha i}$ and $K_i = Ke^{\alpha(k-1)i}$. For given H_i satisfying

(Hi-1) All $x \in V(H_i)$ but at most $\delta_i n_i$ of them has degree $(1 \pm \delta_i)D_i$.
(Hi-2) $\Delta(H_i) < K_i D_i$.
(Hi-3) $\Delta_2(H_i) < \delta_i D_i$.
(Hi-4) $|V(H_i)| = (1 \pm \delta_i)n_i$.

we apply Lemma 4.16 with $\delta_i, \delta_{i+1}/10$ playing the roles of δ_0, δ_1 there to obtain a graph H_{i+1} satisfying (H$(i+1)$-1)–(H$(i+1)$-4) and a set M'_{i+1} of size $|M'_i| = (1 \pm \delta_{i+1})\alpha n_i/k$ covering $(1 \pm \delta_i)n_i - (1 \pm \delta_{i+1})n_i e^{-\alpha} \geq (1 - \delta_{i+2})(\alpha - \alpha^2)n_i$ vertices.

Let M_i be the collection of edges in M_i' which does not intersect with any other edges in M_i'. As M_i' has $(1 \pm \delta_{i+1})\alpha n_i/k$ edges covering at least $(1-\delta_{i+2})(\alpha - \alpha^2)n_i$ vertices, we have $|M_i| \geq (\alpha - 2k\alpha^2)n_i/k$.

Consider a matching $M = M_0 \cup \cdots \cup M_t$, the number of edges in this is at least

$$\sum_{i=0}^{t} \frac{(\alpha - 2k\alpha^2)n_i}{k} \geq (\alpha - 2k\alpha^2)\frac{n}{k}\sum_{i=0}^{t} e^{-\alpha i} \geq (\alpha - 2k\alpha^2)\frac{n}{k}(\frac{1-e^{-t\alpha}}{1-e^{-\alpha}})$$
$$\geq (\alpha - 2k\alpha^2)\frac{n}{k}(\frac{1-\alpha}{\alpha}) \geq (1-\varepsilon)\frac{n}{k}.$$

Note that these inequality holds as $\alpha \ll 1/k, \varepsilon$ and $e^{-\alpha t} < \alpha$. Hence, this forms the desired matching. □

Proof of Lemma 4.16. Choose $\delta_{0.1}, \ldots, \delta_{0.6}$ such that

$$1/D, \delta_0 \ll \delta_{0.1} \ll \delta_{0.2} \ll \cdots \ll \delta_{0.5} \ll \delta_1, \alpha, 1/K, 1/k.$$

This hierarchy of numbers will allow us to omit many computations. Of course, we have $D \leq \binom{n-1}{k-1}$, hence n is also large as D is large.

Let M' be a random subset of $E(H)$ obtained by picking each edge of H independently at random with probability $p = \alpha/D$. We will show that (M1)–(M3) holds with positive probability.

First we show (M1). (H1) together with (H2) ensures that $e(H) = (1 \pm \delta_{0.1})Dn/k$. So, $\mathbb{E}[|M'|] = (1 \pm \delta_{0.1})\alpha n/k$ and $\mathrm{Var}[|M'|] = e(H)p(1-p) \leq (1+\delta_{0.1})(\alpha n/k)$. Using Chebyshev's inequality, we have

$$\mathbf{Pr}[|M'| = (1 \pm \delta_{0.2})\frac{\alpha n}{k}] \geq 1 - \frac{4}{\delta_{0.2}^2 \alpha n/k} \geq 0.99.$$

Hence $(M1)$ holds with probability at least 0.99.

Second, we show (M2). For each $x \in V(H)$, let I_x be the indicator variable of x belonging to V'. So, $I_x = 1$ if $x \in V'$ and 0 if $x \in \bigcup_{e \in M'} e$. Then $|V'| = \sum_{x \in V} I_x$. Let $x \in V(H)$ be good if it has the correct degree $d_H(x) = (1 \pm \delta_0)D$ and bad otherwise. If x is good, then

$$\mathbb{E}[I_x] = \mathbf{Pr}[I_x = 1] = (1-p)^{d(x)} = (1 - \frac{\alpha}{D})^{(1 \pm \delta_0)D} = e^{-\alpha}(1 \pm \delta_{0.1}).$$

If x is bad, then $0 \leq \mathbb{E}[I_x] \leq 1$, but there are only $\delta_0 n$ bad vertices, so

$$\mathbb{E}[|V'|] = e^{-\alpha}(1 \pm \delta_{0.1})(1 \pm \delta_0)n \pm \delta_0 n = ne^{-\alpha}(1 \pm \delta_{0.2}).$$

Now we compute the variance of $|V'|$. Note that

$$\mathrm{Cov}[I_x, I_y] = \mathbb{E}[I_x I_y] - \mathbb{E}[I_x]\mathbb{E}[I_y] = (1-p)^{d(x)+d(y)-d(\{x,y\})} - (1-p)^{d(x)+d(y)}$$
$$\leq (1-p)^{-d(\{x,y\})} - 1 \leq (1 - \frac{\alpha}{D})^{-\delta_0 D} - 1 \leq \delta_{0.1}.$$

Hence,

$$\begin{aligned}\mathrm{Var}[|V'|] &= \sum_{x\in V}\mathrm{Var}[I_x] + \sum_{x\neq y}\mathrm{Cov}[I_x, I_y] \le \mathbb{E}[|V'|] + \sum_{x\neq y}\delta_{0.1} \\ &\le (1\pm\delta_{0.2})ne^{-\alpha} + \delta_{0.1}n^2 \le \delta_{0.2}\mathbb{E}[|V'|]^2.\end{aligned}$$

Hence, we can apply Chebyshev's inequality to conclude that

$$\mathbf{Pr}[|V'| = (1\pm\delta_{0.4})ne^{-\alpha}] \ge \mathbf{Pr}[|V'| = (1\pm\delta_{0.3})\mathbb{E}[|V'|]] \ge 0.99.$$

So, (M2) holds with probability at least 0.99.

Now we prove (M3). For this, we first prove the following claim.

Claim 3. *All but at most $\delta_{0.1}n$ vertices x satisfy the following two.*

(A) $d(x) = (1\pm\delta_0)D$.

(B) all but at most $\delta_{0.1}D$ edges $e \in E(H)$ with $x \in e$ satisfy

$$|\{f \in E(H) : x \notin f, e\cap f \neq \emptyset\}| = (1\pm\delta_{0.1})(k-1)D. \tag{4.1}$$

Proof. By (H1), all but at most $\delta_0 n \le \delta_{0.1}n/2$ vertices satisfy (A).

Recall that the bad vertices are the vertices with degree not $(1\pm\delta_0)D$ but at most KD. As there are at most $\delta_0 n$ bad vertices, there are at most $\delta_0 nKD$ edges containing a bad vertex. Hence, the number of vertices contained in more than $\delta_{0.1}D$ such edges is at most $\delta_0 nKDk/(\delta_{0.1}D) \le \delta_{0.1}n/2$.

If $x \in e$ and e does not contain any bad vertex, then as $\Delta_2(H) \le \delta_0 D$, the number of edges f not containing x that intersect with e is $(k-1)(1\pm\delta_0)D \pm \binom{k-1}{2}\delta_0 D \pm k\delta_0 D = (1\pm\delta_{0.1})(k-1)D$, satisfying (4.1).

So, in total, there are at most $\delta_{0.1}n$ vertices violating (A) or (B). □

Now we want to show that most of the vertices satisfying (A) and (B) satisfies (M3). Let x be a vertex satisfying (A) and (B). We call an edge e good if it satisfies (4.1). Conditioning on $x \in V'$, the probability that a good edge containing x stays in H' is $(1-p)^{(1\pm\delta_{0.1})(k-1)D}$. As at most $\delta_{0.1}n$ edges which is not good can stay with probability between 0 and 1, we have

$$\mathbb{E}[d_{H'}(x)] = (1\pm\delta_0\pm\delta_{0.1})D(1-p)^{(1\pm\delta_{0.1})(k-1)D} \pm \delta_{0.1}D = (1\pm\delta_{0.2})e^{\alpha(k-1)}D.$$

Now, we want to estimate the variance. For each edges e containing x, let I_e be the indicator random variable which is 1 if $e \subseteq V'$ and 0 otherwise. Then $d_{H'}(x) = \sum_{e\ni x} I_e$. So,

$$\begin{aligned}\mathrm{Var}[d_{H'}(x)] &\le \mathbb{E}[d_{H'}(x)] + \sum_{e\ni x, f\ni x}\mathrm{Cov}[I_e, I_f] \\ &\le \mathbb{E}[d_{H'}(x)] + 2\delta_{0.1}D(1\pm\delta_0)D + \sum_{e,f\text{good}, e\ni x, f\ni x}\mathrm{Cov}[I_e, I_f].\end{aligned}$$

We are left to bound the last term. For this, we fix a good e and bound $\sum_{f\ni x, f\text{good}}\mathrm{Cov}[I_e, I_f]$. There are at most $(k-1)\delta_0 D$ edges f with $|e\cap f| > 1$, and their contribution to the sum is at most $(k-1)\delta_0 D$. If $e\cap f = \{x\}$, then let $t(e,f)$ be the number of edges of H that intersect both e, f while not containing x. Then we have $t(e,f) \le (k-1)^2\delta_0 D$. Let a be

the number of edges not containing x but intersecting e, and b be the number of edges not containing x but intersecting with f. For such e and f, we have

$$\begin{aligned}\mathrm{Cov}[I_e, I_f] &\leq \mathbb{E}[I_e I_f] - \mathbb{E}[I_e]\mathbb{E}[I_f] \leq (1-p)^{a+b-t(e,f)} - (1-p)^{a+b} \\ &\leq (1-p)^{-t(e,f)} - 1 \leq \delta_{0.1}.\end{aligned}$$

Hence, we have

$$\sum_{e,f\text{good},e\ni x,f\ni x} \mathrm{Cov}[I_e, I_f] \leq \sum_e ((k-1)\delta_0 D + (1+\delta_0)D\delta_{0.1}) \leq \sum_e \delta_{0.2} D \leq \delta_{0.3} D^2/2.$$

Hence, conditioning on $x \in V'$, we have

$$\mathrm{Var}[d_{H'}(x)] \leq \mathbb{E}[d_{H'}(x)] + \delta_{0.3} D^2 \leq \delta_{0.4}(\mathbb{E}[d_{H'}(x)])^2.$$

By Chebyshev's inequality, $d_{H'}(x) = (1 \pm \delta_{0.5})De^{-\alpha(k-1)}$ does not hold with probability at most $\delta_{0.5}$. By using Markov's inequality, with probability at least 0.99, for all but at most $100\delta_{0.5}n$ vertices x satisfying (A) and (B) satisfy (M3). As at most $\delta_{0.1}n$ vertices not satisfy (A) and (B), and $\delta_{0.1}n + 100\delta_{0.5}n \leq \delta_1 n$, this proves that $(M3)$ holds with probability at least 0.99. This finishes the proof of the lemma. □

5. The local lemma

In many cases, we actually showed that such events holds with high probability while we only wanted to show that such events holds with a positive probability.

On the other hand, there are some cases where the probability of the event is actually a small positive number. For example, considering n independent events of probability $p > 0$, the probability that all of them happens is p^n an exponentially small number. Here the independency trivially provides such an answer. What about the case where there are many events loosely independent of each other? Can we still somehow use 'weak dependency' to prove what we want? We first have to quantify how one can measure 'dependency'.

Definition 5.1. *Let $A_1, \dots, A_n$ be events. A compound event specifies the occurrence of A_i for $i \in S$ and the non-occurrence of A_j for $j \in T$ where S and T are disjoint subsets of $[n]$. An event B is mutually independent of $A_1, \dots, A_n$ if B is independent of each compound event specified by disjoint subsets of $[n]$.*

The following is a symmetric version of local lemma.

Lemma 5.2 (Symmetric local lemma; Erdős-Lovasz, 1975)**.** *Let $A_1, \dots, A_n$ be events such that each is mutually independent of some set of all but at most d events. Suppose that $Pr(A_i) \le p < 1$ for all $i \in [n]$. If $ep(d+1) < 1$, then $Pr(\bigcap \overline{A_i}) > 0$.*

The following is general version of the local lemma.

Definition 5.3. *Let $A_1, \dots, A_n$ be events in a probability space. A directed graph $D = ([n], E)$ on the vertex set $[n]$ is called a dependency digraph for the events $A_1, \dots, A_n$ if for each $i \in [n]$, the event A_i is mutually independent of all the events $\{A_j : (i,j) \notin E\}$*

Lemma 5.4 (The local lemma, general version)**.** *Suppose that D is a dependency digraph for the events $A_1, \dots, A_n$ and suppose there are real numbers $x_1, \dots, x_n \in [0,1)$ such that $\mathbf{Pr}[A_i] \le x_i \prod_{j \in N_D^+(i)} (1 - x_j)$ for all $i \in [n]$. Then*

$$\mathbf{Pr}[\bigwedge_{i=1}^{n} \overline{A_i}] \ge \prod_{i=1}^{n} (1 - x_i).$$

In particular, with positive probability, no events A_i holds.

Proof. We use induction on s to show that for any $S \subseteq [n]$ of size $s < n$ and $i \notin S$,

$$\mathbf{Pr}[A_i \mid \bigwedge_{j \in S} \overline{A_j}] \le x_i.$$

This is true if $s = 0$. Assume that is holds for all $s' < s$. Let

$$S_1 = S \cap N_D^+(i) \text{ and } S_2 = S \setminus S_1.$$

Then we have

$$\mathbf{Pr}[A_i \mid \bigwedge_{j \in S} \overline{A_j}] = \frac{\mathbf{Pr}[A_i \wedge \bigwedge_{j \in S_1} \overline{A_j} \mid \bigwedge_{\ell \in S_2} \overline{A_\ell}]}{\mathbf{Pr}[\bigwedge_{j \in S_1} \overline{A_j} \mid \bigwedge_{\ell \in S_2} \overline{A_\ell}]}.$$

Now we bound the numerator and denominator. As A_i is mutually independent to all events corresponding to S_2, we have

$$\mathbf{Pr}[A_i \wedge \bigwedge_{j\in S_1} \overline{A_j} \mid \bigwedge_{\ell\in S_2} \overline{A_\ell}] \le \mathbf{Pr}[A_i \mid \bigwedge_{\ell\in S_2} \overline{A_\ell}] = \mathbf{Pr}[A_i] \le x_i \prod_{j\in N_D^+(i)} (1-x_j). \tag{5.1}$$

We now bound the denominator using induction hypothesis. If $|S_1| = 0$, then this denominator is 1. Otherwise, we have $S_1 = \{j_1, \dots, j_r\}$. Then

$$\mathbf{Pr}[\bigwedge_{j\in S_1} \overline{A_j} \mid \bigwedge_{\ell\in S_2} \overline{A_\ell}] = (1-\mathbf{Pr}[A_{j_1} \mid \bigwedge_{\ell\in S_2} \overline{A_\ell}])(1-\mathbf{Pr}[A_{j_2} \mid \bigwedge_{\ell\in S_2\cup\{j_1\}} \overline{A_\ell}]) \cdots (1-\mathbf{Pr}[A_{j_r} \mid \bigwedge_{\ell\in S_2\cup\{j_1,\dots,j_{r-1}\}} \overline{A_\ell}])$$
$$\ge \prod_{\ell=1}^{r}(1-x_{j_\ell}) \ge \prod_{j\in N_D^+(i)} (1-x_j).$$

By using these bound, we conclude that $\mathbf{Pr}[A_i \mid \bigwedge_{j\in S} \overline{A_j}] \le x_i$.

Now we use this to prove the lemma.

$$\mathbf{Pr}[\bigwedge_{i=1}^{n} \overline{A_i}] = (1-\mathbf{Pr}[A_1])(1-\mathbf{Pr}[A_2 \mid \overline{A_1}]) \cdots (1-\mathbf{Pr}[A_n \mid \bigwedge_{i=1}^{n-1} \overline{A_i}]) \ge \prod_{i=1}^{n}(1-x_i).$$

□

Proof of the symmetric case. Assume $d > 0$ as it is trivial otherwise. Note that the dependency digraph has maximum outdegree at most d. Take $x_i = \frac{1}{d+1} < 1$, then we have $(1-\frac{1}{d+1})^d > 1/e$ for all d. This guarantees that we can apply the general version of the local lemma. □

5.1. **Several applications.** In many of our applications, the low dependency among the events actually comes from the fact that the dependency of the events are 'local'. This motivates the name 'local lemma'. The following proposition describes this localness.

Proposition 5.5 (Mutual independence principle)**.** *Let $Z_1, \dots, Z_m$ be independent experiments and $A_1, \dots, A_n$ be events such that each A_i is determined by a subset S_i of $Z_1, \dots, Z_m$. If S_i is disjoint from $S_{j_1} \cup \cdots \cup S_{j_k}$ then A_i is mutually independent of $\{A_{j_1}, \dots, A_{j_k}\}$.*

Theorem 5.6. $R(k,k) > (1+o(1))\frac{\sqrt{2}}{e} k 2^{k/2}$.

Proof. We color $E(K_n)$ giving each edge red or blue independently at random. For each vertex set S of size k, let A_S be the event that the subgraph induced by S is monochromatic. Knowing the color of all edges outside $\binom{S}{2}$ has no effect on the probability of A_S. Hence we can let d in the symmetric local lemma be the number of k-sets in $[n]$ that share at least two elements with S, the event A_S is mutually independent of the sets of all other events.

We have

$$d < \binom{k}{2}\binom{n-2}{k-2} < \frac{k^2}{2}(\frac{ne}{k-2})^{k-2}.$$

We also have $Pr(A_S) = 2^{1-\binom{k}{2}} =: p$ for all S.

To apply local lemma, it suffices to make n small enough so that

$$\frac{k^2}{2}(\frac{ne}{k-2})^{k-2} < \frac{1}{ep} = \frac{1}{2e}2^{k/2}(2^{k/2})^{k-2}.$$

Since $2^{k/2} = 2\sqrt{2}^{k-2}$, it suffices to have $n \le c\frac{\sqrt{2}}{e}k2^{k/2}$, where $c = (\frac{2}{ek^2})^{1/(k-2)}\frac{k-2}{k}$. Since $c \to 1$ as $k \to \infty$, the claimed bound holds. □

Although we only gained factor $\sqrt{2}$, it is best we have so far. By using general version of the local lemma, we can actually prove $R(k,3) = \Omega(k^2/\log^2 k)$ and $R(k,4) > k^{5/2+o(1)}$.

By considering red edges as edges and blue edges as non-edges, this implies that there exists a triangle-free graph with independence number at most k with $ck^2/\log^2 k$ vertices. Hence, a graph with chromatic number at least $s = k/\log^2 k$ and the number of vertices at most $s^2 \log^2 s$.

Theorem 5.7 (Erdos, 1961)**.** $R(3,k) \ge \Omega(\frac{k^2}{\log^2 k})$.

Proof. Generate a random graph with vertex set $[n]$ by letting edge edge occur with probability p which we determine later. We must avoid $\binom{n}{3}$ possible triangles, each with probability p^3 and $\binom{n}{k}$ independent k-sets, each with probability $(1-p)^{\binom{k}{2}}$, occurences of these are our events A_i.

We can let $N_D^+(i)$ in the local lemma correspond to those events determined by edge sets intersecting the edge set for A_i.

	# tirangles in $N_D^+(i)$	# k-sets in $N_D^+(i)x$
A_i is triangle	$< 3n$	$< \binom{n}{k}$
A_i is k-sets	$< \binom{k}{2}n$	$< \binom{n}{k}$

We want to find weights y for triangle and z for k-sets such that

$$p^3 < y(1-y)^{3n}(1-z)^{\binom{n}{k}} \text{ and } (1-p)^{\binom{k}{2}} < z(1-y)^{k^2n/2}(1-z)^{\binom{n}{k}}.$$

First, consider k, p, y, z as all functions of n, and later we will consider n as a function of k. We choose $y = p^3(1+\delta)$ for some $\delta > 0$ and y, z small enough so that $(1-y)^{3n}, (1-z)^{\binom{n}{k}}$ approach to 1. For this, we want

$$ny \to 0 \text{ and } \binom{n}{k}z \to 0.$$

To guarantee this, let $z = \binom{n}{k}^{-(1+\delta)}$ and let $p = c_1 n^{-1/2}$ for some constant c_1. This choices guaranatees $ny \to 0$ and $\binom{n}{k}z \to 0$, so we have $(1-y)^{3n} \sim e^{-3ny} \to 1$ and $(1-z)^{\binom{n}{k}} \sim e^{-z\binom{n}{k}} \to 1$. So, the first inequality holds with this choice for large enough n when $\delta > 0$ is a fixed small number.

Consider the second inequality. As our choice of z ensures that the last term in the second inequality is asymptotically 1, we only need to ensure $e^{-pk^2/2} < z(1-y)^{k^2n/2}$. If we can choose k so that $\log\binom{n}{k} \sim c_2 k^2 n^{-1/2}/2$ for some constant c_2, then we only need

$$pk^2/2 > (1+\delta)c_2 n^{-1/2}k^2/2 + yk^2n/2.$$

For this, it suffice to have

$$c_1 > (c_2 + c_1^3)(1+\delta).$$

This holds for some appropriate choices of $0 < c_1, c_2, \delta < 0$. When $k = o(n)$, we have $\log \binom{n}{k} \sim k \log n/k$, so we want to have $k \log n/k \sim c_2 n^{-1/2} k^2/2$. Choosing $k = c_2^{-1} n^{1/2} \log n$ accomplishes this. This gives $n \sim \frac{1}{4} c_2^2 k^2 / \log^2 k$. □

5.2. **Linear arboricity of graphs.** Let's collect the following Chernoff bound which will be useful later. Note the denominator $4pn$ on the exponent.

Lemma 5.8 (Chernoff's bound)**.** *Suppose that $X_1, \ldots, X_n$ are independent random variables such that $\mathbf{Pr}[X_i = 1] = p_i$ and $\mathbf{Pr}[X_i = 0] = 1 - p_i$ for all $i \in [n]$. Let $X := X_1 + \cdots + X_n$ and $p = \frac{1}{n} \sum_{i \in [n]} p_i$. Then for all $0 < t < pn/4$, we have*

$$\mathbf{Pr}[|X - \mathbb{E}[X]| \geq t] \leq 2e^{-t^2/(4pn)}.$$

Proof. Let $Y_i = X_i - p_i$ and $Y = \sum_{i \in [n]} Y_i$, then $\mathbb{E}[Y] = 0$ and it is enough to show that $\mathbf{Pr}[|Y| > t] \leq 2e^{-t^2/(4pn)}$.

Let $\lambda > 0$, and consider the expectation of $e^{\lambda Y_i}$, then

$$\mathbb{E}[e^{\lambda Y_i}] = p_i e^{\lambda(1-p_i)} + (1 - p_i) e^{-\lambda p_i} = e^{-\lambda p_i}(p_i e^{\lambda} + (1 - p_i)).$$

Thus

$$\mathbb{E}[e^{\lambda Y}] = \prod_{i \in [n]} \mathbb{E}[e^{\lambda Y_i}] = \prod_{i \in [n]} (p_i e^{\lambda(1-p_i)} + (1 - p_i)) \leq e^{-\lambda pn}(pe^{\lambda} + (1 - p))^n.$$

The final inequality comes from the concavity of $f(x) = \ln(xe^{\lambda} + 1 - x)$ which yields $\sum f(p_i) \leq nf(p)$. Hence we have

$$\mathbf{Pr}[Y \geq t] = \mathbf{Pr}[e^{\lambda Y} > e^{\lambda t}] < \mathbb{E}[e^{\lambda Y}]/e^{\lambda t} \leq e^{-\lambda pn}(pe^{\lambda} + (1 - p))^n e^{-\lambda t}.$$

Let $\lambda = \ln[(\frac{1-p}{p})(\frac{t+np}{n-(t+np)})]$ and use $\ln(1 + x) \geq x - x^2/2$ then we have

$$\mathbf{Pr}[Y \geq t] \leq e^{-t^2/(2pn) + t^3/(4p^2n^2)} \leq e^{-t^2/(4pn)}.$$

Now we need to estimate $\mathbf{Pr}[Y \leq -t]$. Note that a symmetric argument only gives $e^{-t^2/(2(1-p))n + t^3/(4(1-p)^2n^2)}$ bound which is not what we want.

Again, we have $\mathbb{E}[e^{-\lambda Y}] \leq e^{\lambda pn}(pe^{-\lambda} + (1 - p))^n$. Thus

$$\mathbf{Pr}[Y \leq -t] = \mathbf{Pr}[e^{-\lambda Y} \leq e^{\lambda t}] \leq e^{\lambda pn}(pe^{-\lambda} + (1 - p))^n e^{-\lambda t}.$$

Use standard calculus to simplify the above expression and put $\lambda = t/(np)$, then we have

$$\mathbf{Pr}[Y \leq -t] \leq e^{-t^2/(4pn)}.$$

This proves the lemma. □

Definition 5.9. *A linear forest is a disjoint union of paths. The linear arboricity* $\mathrm{la}(G)$ *of G is the minimum number of linear forests in G whose union is the set of all edges of G.*

The following is a well-known conjecture raised by Akiyama, Exoo and Harary in 1981.

Conjecture 5.10 (The linear arboricity conjecture)**.** *The linear arboricity of a d-regular graph is $\lceil \frac{d+1}{2} \rceil$.*

As every linear forest has at most $n-1$ edges, we have $\mathrm{la}(G) \geq \frac{e(G)}{n-1} = \frac{dn}{2(n-1)} > d/2$. This yields the lower bound for the conjecture. However, the upper bound is difficult. An asymptotic upper bound was proved by Alon in 1988. We prove the following theorem.

Theorem 5.11. *For every d-regular graph G, we have $la(G) \leq \frac{d}{2} + O(d^{3/4}\log^{1/2} d)$.*

To make the analysis easier, we consider a directed version of the conjecture.

Definition 5.12. *A d-regular digraph is a digraph where every vertex has out-degree d and in-degree d. A linear directed forest is a disjoint union of directed paths. The dilinear arboricity* $\mathrm{dla}(G)$ *of a digraph G is the minimum number of linear directed forests in G whose union covers all edges of G.*

Conjecture 5.13. *For every d-regular digraph D,* $\mathrm{dla}(D) = d+1$.

Note that every $2d$-regular graph can be oriented into a d-regular digraph. So this conjecture implies the linear arboricity conjecture for even regular graphs.

We collect the following proposition to show what we want.

Proposition 5.14. *Let H be a graph with maximum degree d and $V(H) = V_1 \cup \dots V_r$ be a partition of V into r sets. Suppose that for each $i \in [r]$ we have $|V_i| \geq 2ed$. Then there is an independent set W that contains a vertex from each V_i.*

Proof. Let $g = \lceil 2ed \rceil$ and delete some vertices if necessary to assume that $|V_i| = g$ for all $i \in [r]$. For each $i \in [r]$, we pick a vertex in V_i independently uniformly at random, and let W be the set of chosen vertices. We show that with positive probability, W is an independent set. As we only choose one vertex from each V_i, we may assume that H has no edges inside V_i for any $i \in [r]$.

For each edge $f \in E(H)$, let A_f be the event that W contains both endpoints of f. Then $\mathbf{Pr}[A_f] \leq \frac{1}{g^2}$. Moreover, if $f \subseteq V_i \cup V_j$, then A_f is mutually independent of all events $\{A_{f'} : f' \cap (V_i \cup V_j) = \emptyset\}$. Thus the dependency digraph for the events $A_f : f \in E(H)$ has maximum degree less than $2g\Delta(H) = 2gd$. As $e \cdot 2gd \cdot (1/g^2) < 1$, Lovasz Local Lemma implies that with positive probability none of the events A_f happen. This means that W is an independent set containing a vertex from each V_i, with positive probability. This concludes the proof. □

Now we use this to prove the above conjecture for graphs with large girth. Again, girth of digraph is the length of the shortest (directed) cycle.

Theorem 5.15. *Let D be a d-regular digraph with girth $g \geq 8ed$. Then* $\mathrm{dla}(G) = d+1$.

Proof. By using Hall's theorem, it is easy to prove that D can be decomposed into d distinct 1-regular spanning digraphs $F_1, \dots, F_d$, and each 1-regular spanning digraph is a union of vertex disjoint cycles. Hence $E(D)$ can be decomposed into $E_1, \dots, E_r$ where each E_i forms a cycle and $|E_i| \geq g \geq 8ed$.

We consider the line graph of the underlying graph of D, which is $4d-2$-regular. As each E_i has size at least $8ed \geq 2e(4d-2)$, the previous proposition implies that there exists a set M of edges of D which contains an edge of E_i and it forms a matching of the underlying graph of D. Therefore $M, F_1 \setminus M, \dots, F_d \setminus M$ are $d+1$ directed linear forests

covering all edges of D. Hence $\mathrm{dla}(D) \le d+1$. As D has $d|D|$ edges and each directed forest has $|D|-1$ edges, so $\mathrm{dla}(D) \ge \frac{|D|d}{|D|-1} > d$, so this shows the equality. □

Now our strategy of proving Theorem 5.11 is to first decompose the digraph into several digraphs with high girth. For this, we prove the following lemma first.

Lemma 5.16. *Let d be sufficiently large and let D be a d-regular digraph and let $p \in [10\sqrt{d}, 20\sqrt{d}]$ be an integer. Then there exists a p-coloring of the vertices of D by the colors $[p]$ with the following: for each $i \in [p]$ and $v \in V(D)$, the number of out/in-neighbors of v with color i is $d/p \pm 5\sqrt{\frac{d\log d}{p}}$.*

Proof. Let G be the underlying graph of D. Let $f : V(D) \to [p]$ be a random vertex coloring of V obtained by chosing the color of each vertex independently uniformly at random. For v, i, let $A^+_{v,i}$ be the event that $X =$ the number of out neighbors of v with color i is not $d/p \pm 5\sqrt{d/p}\sqrt{\log d}$. As X is a binomial random vairiable, using Lemma 5.8 we have

$$\mathbf{Pr}[A^+_{v,i}] \le \frac{1}{d^4}.$$

Similarly, we define $A^-_{v,i}$ which also happens with probability at most $1/d^4$. As the events $A^+_{v,i}$ or $A^-_{v,i}$ are mutually independent of all events $A^+_{u,j}, A^-_{u,j}$ for all u that do not have a common neighbor with v in G. Thus, the dependency digraph has maximum degree at most $(2d)^2 p$. As $e \cdot (1/d^4)((2d)^2 p + 1) < 1$, Lovasz local lemma implies that with positive probability, no events $A^+_{v,i}$ or $A^-_{v,i}$ holds. This provides the desired coloring f. □

Proof of Theorem 5.11. Let p be a prime satisfying $10d^{1/2} \le p \le 20d^{1/2}$. By using the previous lemma, find a vertex-coloring $f : V \to [p]$ satisfying the conclusion of the lemma. For each $i \in [p]$, let D_i be the spanning subgraph of D with edges $\{(u,v) \in E(D) : f(v) \equiv f(u) + i (\mathrm{mod}\ p)\}$. Then each D_i forms a digraph such that all vertices have in/out-degree $d/p \pm 5\sqrt{\frac{d\log d}{p}}$. For each $i \in [p-1]$, each D_i has girth at least p as all cycles in D_i has length divisible by p. One can add vertices and edges to each D_i to convert it into a d_i-regular digraph D'_i of girth at least p where $d_i = d/p + 5\sqrt{\frac{d\log d}{p}}$. By the previous theorem, each D'_i decomposes into at most $d_i + 1$ directed linear forest, and D_0 decomposes into $d/p + 5\sqrt{\frac{d\log d}{p}}$ digraphs with maximum out/in-degree one, and each such graph decomposes into $2d/p + 10\sqrt{\frac{d\log d}{p}}$ directed linear forests. In sum, we have

$$\mathrm{dla}(G) \le (p-1)(d/p + 5\sqrt{\frac{d\log d}{p}}) + \frac{2d}{p} + 10\sqrt{\frac{d\log d}{p}} + p - 1 \le d + O(d^{3/4}(\log d)^{1/2}).$$

□

5.3. **Lopsided Local Lemma.** Note that in the proof of Local lemma, the equality in (5.1) does not have to be equality. As long as we have $\mathbf{Pr}[A_i \mid \bigwedge_{\ell \in S_2} \overline{A_\ell}] \le \mathbf{Pr}[A_i]$, the proof works. This motivates the following definition.

Definition 5.17. *D is a negative dependency digraph for the events $A_1, \dots, A_n$ if for every event A_i and subset $S \subseteq [n] \setminus N^+_D(i)$, we have*

$$\mathbf{Pr}[A_i \mid \bigwedge_{\ell \in S} \overline{A_\ell}] \le \mathbf{Pr}[A_i].$$

Note that what we have above is equivalent to the following property: $\mathbf{Pr}[A_i \mid \bigvee_{\ell\in S} A_\ell] \geq \mathbf{Pr}[A_i]$ for all $S \subseteq [n] \setminus N_D^+(i)$. This says that the out-neighborhood $N_D^+(i)$ contains all events A_j satisfying $\mathbf{Pr}[A_i \mid A_j] < \mathbf{Pr}[A_i]$. In other words, all events A_j to which A_i is negatively related to are in $N_D^+(i)$, hence it is called a negative dependency digraph.

This definition with the proof of Local lemma yields the following theorem.

Theorem 5.18 (Lopsided Local Lemma)**.** *Let $A_1, \dots, A_n$ be events and suppose that D is a negative dependency digraph on the vertex set $[n]$ and suppose that there exists $x_1, \dots, x_n \in (0,1)$ such that for every $i \in [n]$ we have $\mathbf{Pr}[A_i] \leq x_i \prod_{j\in N_D^+(i)}(1-x_j)$. Then*

$$\mathbf{Pr}[\bigwedge_{i\in[n]} \overline{A_i}] \geq \prod_{i\in[n]}(1-x_i).$$

Or, we can even directly assume the last inequality of (5.1), the proof of Local lemma still works.

Theorem 5.19 (Extended Lopsided Local Lemma)**.** *Suppose that D is a digraph on vertex set $[n]$ and there exists $x_1, \dots, x_n \in (0,1)$ such that for every $i \in [n]$ and $J \subseteq [n] \setminus N_D^+(i)$, we have $\mathbf{Pr}[A_i \mid \bigwedge_{j\in J} \overline{A_j}] \leq x_i \prod_{j\in N_D^+(i)}(1-x_j)$. Then*

$$\mathbf{Pr}[\bigwedge_{i\in[n]} \overline{A_i}] \geq \prod_{i\in[n]}(1-x_i).$$

These versions of local lemma is applicable to problems dealing with probability spaces with more structures than independent random variables. We consider an application of this to Latin transversals.

Definition 5.20. *Let A be an $n\times n$ matrix with integer entries. A set $\{(i_1,j_1),\dots,(i_s,j_s)\}$ of pairs is called a partial (Latin) transversal of size s if $i_1,\dots,i_s$ are all distinct and $j_1,\dots,j_s$ are all distinct and the entries $A_{i_1,j_1},\dots,A_{i_s,j_s}$ are all distinct. A partial Latin transversal is called a (Latin) transversal if $s=n$.*

For given $n \times n$ matrix filled with numbers in $[n]$, it is called a Latin square if every row and every column contains each element of $[n]$ exactly once. Ryser-Brualdi-Stein conjecture states that every odd Latin square has a Latin transversal and every even Latin square has a partial Latin transversal of size $n-1$. Brouwer, De vries and Wieringa proved that it has a partial Latin transversal of size $n-\sqrt{n}$ and this was improved in 1982 by Shor to $n - O((\log n)^2)$ and recently by Keevash-Pokrovskiy-Sudakov-Yepremen to $n - O(\frac{\log n}{\log\log n})$.

On the other hand, Stein conjectured that any $n \times n$ matrix where each element of appears in at most n entries must have a partial Latin transversal of size $n-1$. Recently Pokrovskiy and Sudakov disproved this by showing that there is a such matrix with no partial Latin transversal of size $n - \frac{1}{42}\ln n$. On the other hand Aharoni, Berger, Kotlar and Ziv in 2017 proved that any such a matrix has a partial Latin transversal of size at least $2n/3$.

In the course of attacking Stein's conjecture, Erdős-Spencer in 1991 proved the following theorem. This states that Stein's conjecture is true if we strengthen the condition on the matrix.

Theorem 5.21. *Let $k \leq \frac{n-1}{4e}$ and no integer appears in more than k entries of $n \times n$ matrix A. Then A has a Latin transversal.*

Proof. Note that once we choose a permutation π on $[n]$, then we obtain a collection $\{(1, \pi(1)), \ldots, (n, \pi(n))\}$ of pairs. Using this correspondence, we say that π is a Latin transversal if the corresponding set of pairs is a Latin transversal. We choose a permutation π uniformly at random among all permutations on $[n]$. Let

$$T := \{(i, j, i', j') : i < i', j \neq j', A_{i,j} = A_{i',j'}\}.$$

For each $\mathbf{z} \in T$, let $A_{\mathbf{z}}$ denote the event that $\pi(i) = j, \pi(i') = j'$. Then, the event of π being a Latin transversal is that none of these bad events $A_{\mathbf{z}}$ happens. Then we have $\mathbf{P}[A_{\mathbf{z}}] = \frac{1}{n(n-1)}$.

Now we define a graph G on the vertex set T by making (i, j, i', j') adjacent to (p, q, p', q') if and only if $\{i, i'\} \cap \{p, p'\} \neq \emptyset$ or $\{j, j'\} \cap \{q, q'\} \neq \emptyset$. We replace each edge of G into two oppositely directed edges to obtain a digraph D.

Claim 4. *D is negative dependency graph for the events $\{A_{\mathbf{z}} : \mathbf{z} \in T\}$.*

Proof. To prove this, fix $\mathbf{z}$ and $S \subseteq T \setminus N_D^+(\mathbf{z})$. We want to prove

$$\mathbf{Pr}[A_{\mathbf{z}} \mid \bigwedge_{\mathbf{y} \in S} \overline{A_{\mathbf{y}}}] \leq \frac{1}{n(n-1)} = \mathbf{Pr}[A_{\mathbf{z}}].$$

By symmetry, it suffices to consider for the case when $\mathbf{z} = (1, 1, 2, 2)$. For each $i \neq j \in [n]$, let

$$P_{i,j} = \{\pi : \pi(1) = i, \pi(2) = j, \pi \in \bigwedge_{\mathbf{y} \in S} \overline{A_{\mathbf{y}}}\}.$$

Then, this partitions the events $\bigwedge_{\mathbf{y} \in S} \overline{A_{\mathbf{y}}}$ into $n(n-1)$ blocks. Now we will show that $|P_{1,2}| \leq |P_{i,j}|$. To see this, first consider the case when $\{i, j\} \cap \{1, 2\} = \emptyset$ for each $\pi \in P_{1,2}$, let π^* be a permutation in $P_{i,j}$ where

$$\pi^*(\ell) = \begin{cases} \pi(\ell) & \text{if } \pi(\ell) \notin \{1, 2, i, j\} \\ i & \text{if } \pi(\ell) = 1 \\ 1 & \text{if } \pi(\ell) = i \\ j & \text{if } \pi(\ell) = 2 \\ 2 & \text{if } \pi(\ell) = j \end{cases}$$

Indeed, π^* is in $P_{i,j}$ as $\pi^*(1) = i, \pi^*(2) = j$ and if $\pi \in \bigwedge_{\mathbf{y} \in S} \overline{A_{\mathbf{y}}}$ then $\pi^* \in \bigwedge_{\mathbf{y} \in S} \overline{A_{\mathbf{y}}}$ because the events $A_{\mathbf{y}}$ do not involve 1 or 2. This correspondence $\pi \to \pi^*$ is clearly injective, so we have $|P_{1,2}| \leq |P_{i,j}|$.

Note that the above π^* is the composition $(1, i) \cdot (2, j) \cdot \pi$ of transpositions $(1, i)$ and $(2, j)$ with π.

Similarly, we can take π^* as follows for the rest of the cases.

$$\pi^* = \begin{cases} (1,2)\cdot\pi & \text{if } (i,j)=(1,2) \\ (1,i)\cdot\pi & \text{if } 2=j, i\neq 1 \\ (2,j)\cdot\pi & \text{if } 1=i, j\neq 2 \\ (2,j)\cdot\pi & \text{if } 1=i, j\neq 2 \\ (21i)\cdot\pi & \text{if } j=1, i\neq 2 \\ (12j)\cdot\pi & \text{if } i=2, j\neq 1 \end{cases}$$

Again, this defines an injective map as before. Hence this shows that

$$\mathbf{P}[A_{1,1,2,2} \mid \bigwedge_{\mathbf{y}\in S} \overline{A_{\mathbf{y}}}] = \frac{|P_{1,2}|}{\sum_{i\neq j}|P_{i,j}|} \leq \frac{1}{n(n-1)}.$$

□

Now we estimate the maximum out-degree of D. For given (i,j,i',j'), there are at most $(4n-4)$ choices for pairs (p,q) with $p\in\{i,i'\}$ or $q\in\{j,j'\}$. For this fixed (p,q) there are at most $k-1$ pairs (p',q') with $A_{p,q}=A_{p',q'}$ and $(p,q)\neq(p',q')$. So, the maximum out-degree d of D satisfies

$$(d+1) < (4n-4)k \leq \frac{n(n-1)}{e}.$$

Hence, if we let $x_{\mathbf{z}} = \frac{1}{d+1}$ then we have

$$\mathbf{Pr}[A_i] = \frac{1}{n(n-1)} \leq \frac{1}{d+1}(1-\frac{1}{d+1})^d \leq x_i \prod_{j\in N_D^+(i)}(1-x_j).$$

Hence the Lopsided Local Lemma implies that we can avoid all such events with positive probability. This finishes the proof. □

5.4. **Algorithmic aspects of Local Lemma.** We often seek for an efficient algorithms for the problems, either deterministic or randomized.

For example, in an undergraduate graph theory, we learned a simple example of 'derandomization' which converts a probabilistic argument into a deterministic polynomial-time algorithm. A simplest way of derandomization is the method of conditional probabilities which is illustrated in the following example.

Proposition 5.22. *For fixed k, there exists a polynomial-time algorithm to color the edges of K_n so that the number of monochromatic K_k is at most $\binom{n}{k}2^{1-\binom{k}{2}}$.*

Proof. Consider a complete graph K_n on vertex set $[n]$ and consider a partial 2-coloring f of $E(K_n)$ which colors edges of a subgraph H of $E(K_n)$. For each set $S\in\binom{[n]}{k}$, we let

$$w(S) = \begin{cases} 0 & \text{if } f \text{ colors two edges in } \binom{S}{2} \text{ by different colors} \\ 2^{1-\binom{k}{2}} & \text{if no edge in } \binom{S}{2} \text{ is colored} \\ 2^{-r} & \text{if } f \text{ colors all but } r \text{ edges in } \binom{S}{2} \text{ using one color.} \end{cases}$$

Let $w(f)=\sum_{S\in\binom{[n]}{k}} w(S)$. Although this weight is a deterministic weight, it has the following probabilistic interpretation. For given partial coloring f, we randomly color remaining uncolored edges. The probability $Pr[S \text{ is monochromatic}]$ that S induces a monochromatic k-clique in this experiment equals the weight $w(S)$. Hence, $w(f)=\mathbb{E}[X:$

$\phi\,|_{E(H)}= f]$ where X is the number of monochromatic k-cliques in a random coloring ϕ, where we condition on the event that the resulting coloring is f when it's restricted to H. So it's a conditional expectation.

Now we choose an edge e not colored by f. Let f_1, f_2 be two extension of f by coloring e with color $1, 2$, respectively. As the weight is conditional expectation, we have $w(f) = \frac{1}{2}w(f_1) + \frac{1}{2}w(f_2)$. Hence, there exists a choice $i \in \{1, 2\}$ with $w(f) \geq w(f_i)$. We color e by the color i and we proceed in this manner. Since any edge we consider belongs to n^{k-2} cliques whose weight will be updated once we color the edges, the algorithm runs in time $O(n^k)$. □

However, when we solve a problem by applying Local Lemma, it is difficult to imagine how one can turn this proof into even an efficient randomized algorithm. Local lemma shows that a given desired event holds with positive probability, but the probability can be exponentially small in terms of the size of the problem. Consequently, one might have to repeat the random process exponentially many times until one obtain a solution. However, the folloiwng simple randomized algorithm by Moser 2009 and Moser-Tardos 2010 enable us to turn most of the applications of Local Lemma into a randomized algorithm. We consider the following set-up.

Set-up 5.23. *Let Ω be a finite set and for each $v \in \Omega$, let $C[v]$ denote a random variable and the random variables $C[v]$ are mutually independent. Let I be an index set and for each $\alpha \in I$, there is a set $A[\alpha] \subseteq \Omega$ and an event $BAD[\alpha]$ depending on $(C[v] : v \in A[\alpha])$ with $p[\alpha] = \mathbf{Pr}[BAD[\alpha]]$. Let $\alpha \sim \beta$ if $A[\alpha] \cap A[\beta] \neq \emptyset$ for each $\alpha, \beta \in I$.*

For example, consider the problem of avoiding monochromatic K_ks in a random coloring of K_n. We consider a random coloring of edges of K_n, $\Omega = E(K_n)$ and $C[e]$ is a random variable denoting the colors of edges $e \in E(K_n)$. And $I = \binom{V(K_n)}{k}$ and $A[\alpha] = E(K_n[\alpha])$ for $\alpha \in I$. Then $BAD[\alpha]$ is the event that α induces a monochromatic clique. In the above set up, the relation $\sim$ yields a dependency graph on the events $BAD[\alpha]$.

Algorithm 5.24 (Moser's fix-it algorithm)**.** *For each $v \in \Omega$, choose $C[v]$ at random. Pick one arbitrary bad event $BAD[\alpha]$ occurred and select all the $C[v]$ at random again for all $v \in A[\alpha]$. Repeat this while there's bad events occured.*

Note that resampling all $C[v]$s might induces some other bad events. But we just keep repeating this, while hoping all bad events will disappear at some point.

Definition 5.25. *Let LOG be the sequence $\alpha_1\alpha_2 \ldots \alpha_u$ where α_t is the α selected at t-th time in the above algorithm and let $TLOG = u$.*

In order to analyze the LOG above, we introduce the following concept of Moser tree.

Definition 5.26. *The depth of a vertex in a rooted tree is its distance from the root, and the depth of a rooted tree is the maximum depth of its vertices. A Moser Tree is a finite rooted tree T with labeling $f : V(T) \to I$ satisfying the following:*

(1) If u with label α is a child of v with label β, then $\alpha \sim \beta$.

(2) If u has label α and v has label β with $\alpha \sim \beta$, then u and v has different depths.

For a labelled rooted tree (T, f) with lables, let $p(T) = \prod_{v \in V(T)} p(f(v))$.

The reason why we define LOG and Moser trees are as follows. In the Moser's fix-it algorithm, we will obtain a sequence of labels as its LOG. We want to bound the probability of getting a specific sequence as its LOG, so that we will be able to prove that the expected length of the LOG is finite. For this, we will correspond a Moser tree T to a given possible LOG, and show that $p(T)$ is an upper bound on the probability of having the sequence as our LOG.

Definition 5.27. *Let $\alpha_1 \dots \alpha_u$ be a prefix of LOG. We associate to it a labeled rooted tree T_u as follows. The root has label α_u, and let t run from $u-1$ down to 1. If we do not have $\alpha_t \sim \alpha_{t'}$ for any $t < t' \leq u$ for which $\alpha_{t'}$ has been already placed in T_u, then we ignore α_t. Otherwise, among all such t', select one such that the vertex labeled $\alpha_{t'}$ is at the greatest depth. (Ties can be broken arbitrarily) Add a vertex with label α_t and make it the child of the vertex labeled $\alpha_{t'}$.*

Claim 5. *For a prefix $\alpha_1 \dots \alpha_u$ of LOG, its corresponding labeled rooted tree T_u is a Moser tree.*

Proof. When a vertex with label β is created as a child of a vertex with label α, we must have $\alpha \sim \beta$. Moreover, if two vertices at the same depth D has labels $\alpha \sim \beta$, the one which was added later in the process must have been placed so that it has depth at least $D+1$, a contradiction. This proves the claim. □

Claim 6. *For a prefix $\alpha_1 \dots \alpha_v$ of LOG and $u < v$, the Moser trees T_u and T_v are not equal.*

Proof. Suppose $T_u = T_v$. Then we have $\alpha_u = \alpha_v$ as they are the labels of two same Moser trees. For all $j \in [u]$, all α_j in T_u must be also in T_v while T_v have an additional vertex α_v, hence $|T_v| > |T_u|$ a contradiction. □

This claim shows that a LOG of length u will generate u distinct Moser trees. Hence,

$$\mathbb{E}[TLOG] = \sum_{n \geq 1} \mathbf{Pr}[TLOG \geq n] = \sum_{T\text{: a Moser tree}} \mathbf{P}[T_n = T \text{ for some } n].$$

To show the following lemma, we define a weak Moser tree as follows.

Definition 5.28. *A rooted tree with vertex labels in I is a weak Moser tree if the following holds.*

(1) If u with label α is a child of v with label β, then $\alpha \sim \beta$.
(2) Labels of the children of a vertex are all distinct.

Lemma 5.29. *For any Moser tree T, we have*

$$\mathbf{Pr}[T = T_n \text{ for some } n] \leq p(T).$$

Proof. Note that we will resample $C[v]$ several times, and each resampling result is independent with the previous ones. Hence, we may as well assume there are $C[v,0], C[v,1], \dots$ mutually independent identically distributed random variables.

At the beginning of Moser's fix-it algorithm, we use $C[v,0]$ and later if we resample $C[v]$ after using $C[v,0],\dots,C[v,t]$, we sample from $C[v,t+1]$.

Let $\alpha_1\alpha_2\dots\alpha_u$ be the LOG which has Moser tree T. By deleting some unused labels from the LOG, we assume that T has exactly u vertices, and $w_1,w_2,\dots,w_u$ are the corresponded vertices in T with w_i being labeled with α_i. For each $i\in[n]$ and $v\in A[\alpha_i]$, let $f_i(v)=|\{j\le i : v\in A[\alpha_j]\}|-1$.

For $T_n=T$ to happen, it is necessary that $BAD[\alpha_t]$ holds for all t. Then $\mathbf{Pr}[BAD[\alpha_t]\mid \bigwedge_{0\le t'<t} BAD[\alpha_{t'}]]$ depends on $(C[v,f_i(v)] : v\in A[\alpha_t])$. As all $C[v,i]$ are mutually independent, this happens with probability $\mathbf{P}[BAD[\alpha_t]]=p(\alpha_t)$. This proves that

$$\mathbf{Pr}\left[\bigwedge_{0\le t'<u} BAD[\alpha_{t'}]\right] \le \prod_{0\le t\le u} \mathbf{Pr}\left[BAD[\alpha_t]\mid \bigwedge_{0\le t'<t} BAD[\alpha_{t'}]\right] = p(T).$$

□

Now we estimate the following.

Lemma 5.30. *Suppose that there exists $x(\alpha)\ge p(\alpha)$ for each $\alpha\in I$ such that*

$$x(\alpha)\ge p(\alpha)\prod_{\beta\sim\alpha}(1+x(\beta)).$$

Then $\mathbb{E}[TLOG]\le\sum_{\alpha\in I}x(\alpha)$.

Proof. In order to upper bound $\mathbb{E}[TLOG]$, by the previous claim, we only have to bound

$$\sum_{T\text{: a Moser tree}} p(T) \le \sum_{T\text{: a weak Moser tree}} p(T).$$

Let $w(D,\alpha)=\sum_{\substack{T\text{: weak Moser tree with root}\\ \text{label }\alpha\text{ with depth }\le D}} p(T)$. Then, every weak Moser tree with root labeled α decomposes into the root and some weak Moser tree with roots $\beta\sim\alpha$. Hence, we have the following recursion.

$$w(D,\alpha)=p(\alpha)\prod_{\beta\sim\alpha}(1+w(D-1,\beta)).$$

We show $w(D,\alpha)\le x(\alpha)$ for all $\alpha\in I$ by induction on D. For $D=0$, we have $w(0,\alpha)=p(\alpha)\le x(\alpha)$. Suppose that $w(D-1,\beta)\le x(\beta)$ holds for all $\beta\in I$. Then the above recursion implies

$$w(D,\alpha)=p(\alpha)\prod_{\beta\sim\alpha}(1+w(D-1,\beta))\le p(\alpha)\prod_{\beta\sim\alpha}(1+x(\beta))\le x(\alpha).$$

Thus

$$\sum_{T\text{: a weak Moser tree}} p(T)=\sum_{\alpha\in I}\lim_{D\to\infty} w(D,\alpha)\le\sum_{\alpha\in I}x(\alpha).$$

This proves the lemma. □

Note that the condition $\mathbf{Pr}[A_i]\le x_i\prod_{j\in N_D^+(i)}(1-x_j)$ holds, then we let $x(i)=x_i$ for all $i\in I$. As $\mathbf{Pr}[A_i]=p(i)$ and $j\in N_D^+(i)$ is same as $i\sim j$ in this set-up, for all $\alpha=i\in I$, we have

$$x(\alpha)\ge p(\alpha)\prod_{\beta\sim\alpha}(1-x(\beta))^{-1}\ge p(\alpha)\prod_{\beta\sim\alpha}(1+x(\beta)).$$

Thus, as long as the Local Lemma works, then the above lemma ensures that the Moser's fix-it algorithm has expected TLOG at most $\sum_{\alpha\in I} x_i$. In most of the applications, there are polynomially many bad events, and x_i are polynomially small, this expectation is at most a polynomial, say $O(n^s)$. Using Markov's inequality, one can conclude that Moser's fix-it algorithm ends within a polynomial times $O(n^{s+s'})$ with probability at least $1 - O(n^{-s'})$.

In the symmetric case, with dependency digraph having degree at most d, taking $x = d^{-1}$ yields that $ep(d+1) \leq 1$ implies $\mathbb{E}[TLOG] \leq \frac{|I|}{d}$.

6. Correlation inequality

Consider a random graph $G(n,p)$ and let H be the event that the random graph is Hamiltonian and P be the event that the random graph is Planar. Intuitively, being Hamiltonian suggests that it has many edges, and being planar suggests that it has a few edges. Therefore, we might suspect $\mathbf{Pr}[P \mid H] \leq \mathbf{Pr}[P]$, implying $\mathbf{Pr}[P \wedge H] \leq \mathbf{Pr}[H] \cdot \mathbf{Pr}[P]$.

Definition 6.1. *A graph property is a collection of graphs. A graph property $\mathcal{P}$ is monotonically increasing if for each $G \in \mathcal{P}$ and a supergraph G' of G with $V(G) = V(G')$, we have $G' \in \mathcal{P}$. A graph property $\mathcal{P}$ is monotonically decreasing if for each $G \in \mathcal{P}$ and a subgraph G' of G with $V(G) = V(G')$, we have $G' \in \mathcal{P}$.*

Note that Hamiltonicity is a monotonically increasing property and Planarity is a monotonically decreasing property. We want to consider some graph properties $\mathcal{P}, \mathcal{Q} \subseteq 2^{\binom{[n]}{2}}$ where each of $\mathcal{P}, \mathcal{Q}$ is either monotonically increasing or monotonically decreasing, we want to compare two probabilities

$$\mathbf{Pr}[G(n,p) \in \mathcal{P} \cap \mathcal{Q}] \quad \text{and} \quad \mathbf{Pr}[G(n,p) \in \mathcal{P}] \cdot \mathbf{Pr}[G(n,p) \in \mathcal{Q}].$$

In order to deduce the desired inequality, we will first prove a more general "the four functions theorem" and deduce less general theorems "the FKG inequality" from it. Eventually the less general theorem will imply what we want.

Note that $\mathbf{Pr}[G(n,p) \in \mathcal{P}]$ can be written as $\sum_{G \in \mathcal{P}} \mu(G)$ where $\mu(G) = \mathbf{Pr}[G(n,p) = G]$ is the measure of a specific labeled graph. Moreover,

$$\sum_{G \in \mathcal{P}} \mu(G) = \sum_{G\text{: any labeled graphs}} \mathbb{1}_{\mathcal{P}}(G)\mu(G).$$

By replacing the indicator function to more general functions, this motivates the following theorem.

Theorem 6.2 (The FKG inequality)**.** *Let L be a finite distributive lattice, and let $\mu : L \to \mathbb{R}^+$ be a log-supermodular function. Then, for any two increasing functions $f, g : L \to \mathbb{R}^+$, we have*

$$\left(\sum_{x \in L} \mu(x) f(x)\right)\left(\sum_{x \in L} \mu(x) g(x)\right) \leq \left(\sum_{x \in L} \mu(x) f(x) g(x)\right)\left(\sum_{x \in L} \mu(x)\right).$$

We will later define the terms used here, like 'distributive lattice' and 'log-supdermodular functions' and etc. However, notice that letting $f(x) = \mathbb{1}_{\mathcal{P}}, g(x) = \mathbb{1}_{\mathcal{Q}}$ and $\mu(G) = \mathbf{Pr}[G(n,p) = G]$ and applying The FKG inequality will yields what we want, as $\mathbb{1}_{\mathcal{P}} \cdot \mathbb{1}_{\mathcal{Q}} = \mathbb{1}_{\mathcal{P} \cap \mathcal{Q}}$. Moreover, the above inequality is regarding four functions $\mu \cdot f, \mu \cdot g, \mu \cdot f \cdot g, \mu$ with certain good property. By analyzing the essential properties of these functions making the inequality true, we can make the statement more abstract to obtain the following more general theorem, called the four function theorem.

Note that a graph G can be identified with a subset A of $[n]$ if we identify each pair in $\binom{V}{2}$ with a number in $[n]$ when $n = \binom{|V|}{2}$. So, the set-up in the following subsection is more general than what we want.

6.1. **The four function theorem.** Suppose $n \geq 1$ and let $2^{[n]}$ denote the set of all subsets of $[n]$. For a function $\varphi : 2^{[n]} \to \mathbb{R}^+$ and a family $\mathcal{A} \subseteq 2^{[n]}$, we write $\varphi(\mathcal{A}) = \sum_{A\in\mathcal{A}} \varphi(A)$. For $\mathcal{A}, \mathcal{B} \subseteq 2^{[n]}$, let

$$\mathcal{A} \cup \mathcal{B} = \{A \cup B : A \in \mathcal{A}, B \in \mathcal{B}\} \text{ and } \mathcal{A} \cap \mathcal{B} = \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}.$$

Theorem 6.3 (The four function theorem)**.** *Let $\alpha, \beta, \gamma, \delta : 2^{[n]} \to \mathbb{R}^+$ be four functions from the set of all subsets of $[n]$ to the nonnegative reals. Suppose that for every two subsets $A, B \subseteq [n]$, we have*

$$\alpha(A)\beta(B) \leq \gamma(A \cup B)\delta(A \cap B). \tag{6.1}$$

Then for any two families of subsets $\mathcal{A}, \mathcal{B} \subseteq 2^{[n]}$, we have

$$\alpha(\mathcal{A})\beta(\mathcal{B}) \leq \gamma(\mathcal{A} \cup \mathcal{B})\delta(\mathcal{A} \cap \mathcal{B}). \tag{6.2}$$

Proof. It is easy to check that if we replace $\alpha, \beta, \gamma, \delta$ with $\alpha \cdot \mathbb{1}_{\mathcal{A}}, \beta \cdot \mathbb{1}_{\mathcal{B}}, \gamma \cdot \mathbb{1}_{\mathcal{A}\cup\mathcal{B}}, \delta \cdot \mathbb{1}_{\mathcal{A}\cap\mathcal{B}}$, (6.1) still holds and (6.2) stays the same. After this replacement, we can assume that $\mathcal{A} = \mathcal{B} = \mathcal{A} \cap \mathcal{B} = \mathcal{A} \cap \mathcal{B} = 2^{[n]}$.

With this assumption, we prove the inequality using induction on n. Consider the base case when $n = 1$. For each $\varphi \in \{\alpha, \beta, \gamma, \delta\}$, let $\varphi_0 = \varphi(\emptyset)$ and $\varphi_1 = \varphi([1])$. Then (6.1) implies that we have

$$\alpha_0\beta_0 \leq \gamma_0\delta_0, \quad \alpha_0\beta_1 \leq \gamma_1\delta_0, \quad \alpha_1\beta_0 \leq \gamma_1\delta_0, \quad \alpha_1\beta_1 \leq \gamma_1\delta_1. \tag{6.3}$$

To show (6.2), we only have to prove the following.

$$(\alpha_0 + \alpha_1)(\beta_0 + \beta_1) \leq (\gamma_0 + \gamma_1)(\delta_0 + \delta_1). \tag{6.4}$$

If $\gamma_1 = 0$ or $\delta_0 = 0$, then this is obvious from (6.3). Otherwise, (6.3) implies $\gamma_0 \geq \alpha_0\beta_0/\delta_0$ and $\delta_1 \geq \alpha_1\beta_1/\gamma_1$. So, (6.4) can be derived from

$$(\frac{\alpha_0\beta_0}{\delta_0} + \gamma_1)(\delta_0 + \frac{\alpha_1\beta_1}{\gamma_1}) \geq (\alpha_0 + \alpha_1)(\beta_0 + \beta_1)$$

which is equivalent to

$$(\alpha_0\beta_0 + \gamma_1\delta_0)(\gamma_1\delta_0 + \alpha_1\beta_1) \geq (\alpha_0 + \alpha_1)(\beta_0 + \beta_1)\gamma_1\delta_0$$

also equivalent to

$$(\gamma_1\delta_0 - \alpha_0\beta_1)(\gamma_1\delta_0 - \alpha_1\beta_0) \geq 0.$$

This follows from (6.3). This complete the proof for $n = 1$.

Suppose that the theorem holds for $n - 1$. For each $\varphi \in \{\alpha, \beta, \gamma, \delta\}$ and $A \subseteq [n - 1]$, let $\varphi'(A) = \varphi(A) + \varphi(A \cup \{n\})$. As $\varphi'(2^{[n-1]}) = \varphi(2^{[n]})$, we only have to prove $\alpha'(2^{[n-1]})\beta'(2^{[n-1]}) \leq \gamma'(2^{[n-1]})\delta'(2^{[n-1]})$. For this we want to use the induction hypothesis, so we only have to check the assumption (6.1) for $\alpha', \beta', \gamma', \delta'$ on $[n - 1]$.

Fix $A', B' \subseteq [n - 1]$, and let $\overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\delta}$ be functions on $2^{[1]}$ with

$$\begin{aligned}
\overline{\alpha}(\emptyset) = \alpha(A'), &\quad \overline{\alpha}([1]) = \alpha(A' \cup \{n\}) \\
\overline{\beta}(\emptyset) = \beta(B'), &\quad \overline{\beta}([1]) = \beta(B' \cup \{n\}) \\
\overline{\gamma}(\emptyset) = \gamma(A' \cup B'), &\quad \overline{\gamma}([1]) = \gamma(A' \cup B' \cup \{n\}) \\
\overline{\delta}(\emptyset) = \delta(A' \cap B'), &\quad \overline{\delta}([1]) = \delta((A' \cap B') \cup \{n\}).
\end{aligned}$$

Then (6.1) implies that we have $\overline{\alpha}(S)\overline{\beta}(R) \le \overline{\gamma}(S\cup R)\overline{\delta}(S\cap R)$ for all $S, R \subseteq [1]$, hence by the case $n = 1$ of the statement we already proved, we have

$$\alpha'(A')\beta'(B') = \overline{\alpha}(2^{[1]})\overline{\beta}(2^{[1]}) \le \overline{\gamma}(2^{[1]})\overline{\delta}(2^{[1]}) = \gamma'(A'\cup B')\delta'(A'\cap B').$$

This is the desired inequality so that (6.1) holds for $\alpha', \beta', \gamma', \delta'$ on $[n-1]$. Hence we have $\alpha'(A')\beta'(B') \le \gamma'(A'\cup B')\delta'(A'\cap B')$. Induction hypothesis yields that we have $\alpha'(2^{[n-1]})\beta'(2^{[n-1]}) \le \gamma'(2^{[n-1]})\delta'(2^{[n-1]})$ this implies that $\alpha(2^{[n]})\beta(2^{[n]}) \le \gamma(2^{[n]})\delta(2^{[n]})$. Again, as we have replaced $\alpha, \beta, \gamma, \delta$ with $\alpha\cdot\mathbb{1}_{\mathcal{A}}, \beta\cdot\mathbb{1}_{\mathcal{B}}, \gamma\cdot\mathbb{1}_{\mathcal{A}\cup\mathcal{B}}, \delta\cdot\mathbb{1}_{\mathcal{A}\cap\mathcal{B}}$ at the beginning, this implies (6.2), concluding the proof of the theorem. □

This theorem actually is about more general structures than just the subsets of $[n]$.

Definition 6.4. *A lattice is a partially ordered set in which every elements x and y have a unique minimal upper bound $x\vee y$ called the join of x and y, and a unique maximal lower bound $x\wedge y$ called the meet of x and y. A lattice L is distributed if for all $x, y, z \in L$,*

$$x\wedge(y\vee z) = (x\wedge y)\vee(x\wedge z) \Leftrightarrow x\vee(y\wedge z) = (x\vee y)\wedge(x\vee z).$$

Definition 6.5. *For two sets $X, Y \subseteq L$, define $X\vee Y = \{x\vee y : x\in X, y\in Y\}$ and $X\wedge Y = \{x\wedge y : x\in X, y\in Y\}$.*

For example, considering partially ordered set on $L\subseteq 2^{[n]}$ ordered by inclusion, which is closed under the union and intersection is a distributive lattice. Here, the join is same as the union and meet is same as the intersection. It is standard to check that every finite distributive lattice L is isomorphic to a sublattice of $2^{[n]}$ for some n. This fact generalize the previous theorem to the following.

Theorem 6.6. *Let L be a finite distributive lattice and let α, β, γ and δ be four functions from L to $\mathbb{R}^+$ such that $\alpha(x)\beta(y) \le \gamma(x\vee y)\delta(x\wedge y)$ holds for all $x, y \in L$. Then for every $X, Y \subseteq L$, we have $\alpha(X)\beta(Y) \le \gamma(X\vee Y)\delta(X\wedge Y)$.*

6.2. **The FKG inequality.** Using the four function theorem, we can prove the FKG inequality.

Definition 6.7. *A function $\mu : L\to\mathbb{R}^+$ where L is a finite distributive lattice is called log-supermodular if $\mu(x)\mu(y) \le \mu(x\vee y)\mu(x\wedge y)$ for all $x, y\in L$. A function $f : L\to\mathbb{R}^+$ is increasing if $f(x)\le f(y)$ whenever $x\le y$ and decreasing if $f(x)\ge f(y)$ whenever $x\le y$.*

Now we prove the folowing theorem.

Theorem 6.8 (The FKG inequality)**.** *Let L be a finite distributed lattice, and let $\mu : L\to\mathbb{R}^+$ be a log-supermodular function. Then, for any two increasing functions $f, g : L\to\mathbb{R}^+$, we have*

$$\left(\sum_{x\in L}\mu(x)f(x)\right)\left(\sum_{x\in L}\mu(x)g(x)\right) \le \left(\sum_{x\in L}\mu(x)f(x)g(x)\right)\left(\sum_{x\in L}\mu(x)\right).$$

Proof. Define four functions $\alpha = \mu\cdot f$, $\beta = \mu\cdot g$, $\gamma = \mu\cdot f\cdot g$ and $\delta = \mu$. We claim that these functions satisfy the hypothesis of Theorem 6.6. Indeed, as f, g are increasing and

μ is log-supermodular, for $x, y \in L$ then we have

$$\alpha(x)\beta(y) = \mu(x)f(x)\mu(y)g(y) \leq \mu(x \vee y)f(x)g(y)\mu(x \wedge y) \leq \mu(x \vee y)f(x \vee y)g(x \vee y)\mu(x \wedge y)$$
$$= \gamma(x \vee y)\delta(x \wedge y).$$

Thus, by Theorem 6.6, we have $\alpha(L)\beta(L) \leq \gamma(L)\delta(L)$, proving the inequality. □

Note that if f, g are decreasing, then the proof still works by switching the roles of γ, δ. If f is increasing and g is decreasing, the opposite inequality holds. This can be proved by applying the above theorem to two increasing functions $f(x), k - g(x)$ where k is $\max_{x \in L} g(x)$.

Again, one can consider μ as a measure on L.

Definition 6.9. *Consider a finite set N. Let $\mathcal{A}$ be a family of subsets of N is monotonically decreasing if $A \in \mathcal{A}$ and $A' \subseteq A$ implies $A' \in \mathcal{A}$. Similarly, it is monotonically increasing if $A \in \mathcal{A}$ and $A' \supseteq A$ implies $A' \in \mathcal{A}$.*

We consider $N = \binom{[n]}{2}$, then this coincides with the definition on the set of graphs. Let $|N| = m$, and consider a real vector $(p_1, \ldots, p_m)$ where $0 \leq p_i \leq 1$. Consider the probability space whose elements are all members of 2^N where for each $A \subseteq N$, $\mathbf{Pr}_p[A] = \prod_{i \in A} p_i \prod_{j \notin A}(1 - p_j)$. Define $\mu = \mu_p : 2^N \to \mathbb{R}^+$ by $\mu(A) = \mathbf{Pr}[A]$. With this, it is easy to check that μ is log-supermodular. Moreover, we have $\mu(A)\mu(B) = \mu(A \cup B)\mu(A \cap B)$. Applying FKG inequality, we obtain the following.

Theorem 6.10. *Let $\mathcal{A}, \mathcal{B}$ be two monotonically increasing families of subsets of N and let $\mathcal{C}, \mathcal{D}$ be two monotonically decreasing families of subsets of N. Then for any real vector $p = (p_1, \ldots, p_m)$, $0 \leq p_i \leq 1$, we have*

$$\mathbf{Pr}_p[\mathcal{A} \cap \mathcal{B}] \geq \mathbf{Pr}_p[\mathcal{A}] \cdot \mathbf{Pr}_p[\mathcal{B}],$$
$$\mathbf{Pr}_p[\mathcal{C} \cap \mathcal{D}] \geq \mathbf{Pr}_p[\mathcal{C}] \cdot \mathbf{Pr}_p[\mathcal{D}],$$
$$\mathbf{Pr}_p[\mathcal{A} \cap \mathcal{C}] \leq \mathbf{Pr}_p[\mathcal{A}] \cdot \mathbf{Pr}_p[\mathcal{C}].$$

Note that with this set up and letting $N = \binom{[n]}{2}$, we obtain the probability space of $G(n, p)$, concluding

Theorem 6.11. *Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_4$ be graph properties where $\mathcal{Q}_1, \mathcal{Q}_2$ are monotonically increasing and $\mathcal{Q}_3, \mathcal{Q}_4$ are monotonically decreasing. Let $G = G(n, p)$ be a graph obtained from the binomial random graph model. Then*

$$\mathbf{Pr}[G \in \mathcal{Q}_1 \cap \mathcal{Q}_2] \geq \mathbf{Pr}[G \in \mathcal{Q}_1] \cdot \mathbf{Pr}[G \in \mathcal{Q}_2],$$
$$\mathbf{Pr}[G \in \mathcal{Q}_3 \cap \mathcal{Q}_4] \geq \mathbf{Pr}[G \in \mathcal{Q}_3] \cdot \mathbf{Pr}[G \in \mathcal{Q}_4],$$
$$\mathbf{Pr}[G \in \mathcal{Q}_1 \cap \mathcal{Q}_3] \leq \mathbf{Pr}[G \in \mathcal{Q}_1] \cdot \mathbf{Pr}[G \in \mathcal{Q}_3].$$

7. Martingales and tight concentrations

Assume that we are interested in a random variable X. As an example, you might consider the number of heads you get after flipping coins n times. The value of X can be computed thought a process, which yields intermediate values $X_0, X_1, \ldots, X_n = X$. In our example, X_i is the number of heads in the first i coin tosses. We know $\mathbb{E}[X] = n/2$, but can we show that $|X - \mathbb{E}[X]|$ is small with a very high probability (tight concentration)? More generally, we expect X_i to be close to $f(i) = i/2$, so $f(i)$ is 'the expected behavior' of the random variable X_i, and we wish to show that X_i is not too far from $f(i)$ with high probability. This can be shown if the expected value of $X_{i+1} - X_i$ is same as $f(i+1) - f(i)$ and $|X_{i+1} - X_i|$ is never too big. These two conditions are very natural condition to obtain tight concentrations. Let's first focus on the first condition. As this function $f(i)$ can vary over various random variables, and it's not convenient to deal with, let $X_i' = X_i - f(i)$. Then it suffices to show that X_i' is close to zero. Conventionally, we instead consider $X_i' = X_i - f(i) + X_0$ so that it is enough to show that X_i' is close to X_0. With this set up, the first condition become that "the expected value of $X_i' - X_{i+1}'$ is zero. We give the following name to random variables satisfying this condition.

Definition 7.1. *A martingale is a sequence of random variables $X_0, X_1, \ldots, X_m$ such that for every i and $x_0, \ldots, x_i \in \mathbb{R}$, we have*

$$\mathbb{E}[X_{i+1} \mid X_0 = x_0, \ldots X_i = x_i] = x_i.$$

In other words, we might just write $\mathbb{E}[X_{i+1} \mid X_0, \ldots, X_i] = X_i$ instead of the above equation.

As an example, consider a gambler with initial amount of money X_0. Each time the gambler flips a fair coin and if it's head the gambler earns 1 otherwise the gambler loses 1. As the game is fair, it is easy to check $\mathbb{E}[X_{i+1}|X_0, \ldots, X_i] = X_i$.

We might change the game a bit. Assume that a gambler can decide the amount of money the gambler can bet. If the coin is head, the gambler earns the money as much as the gambler and otherwise loses the money the gambler bet. The gambler might decide the amount of money the gambler bet based on the previous history $X_0, \ldots, X_i$. For example, the gambler can double the bet until the gambler win. However, no matter what strategy the gambler uses, we still have the equality $\mathbb{E}[X_{i+1}|X_0, \ldots, X_i] = X_i$. Hence $X_0, \ldots, X_m$ is still a martingale.

In many of our applications, the martingale we deal with is of the following form.

Definition 7.2. *Suppose that we have random variables $X_1, \ldots, X_n$ which are not necessarily independent, and each X_i has a value in V_i and let $f : V_1 \times \cdots \times V_n \to \mathbb{R}$. Let*

$$Y_i = \mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i].$$

Then $Y_0, Y_1, \ldots, Y_n$ forms a martingale and this is called the Doob martingale or the exposure martingale.

Note that Y_i is also a random variable as the values of $X_1, \ldots, X_i$ are randomly determined and these values determine the value of Y_i. Moreover, we can easily check that $\mathbb{E}[Y_{i+1} : Y_0, \ldots, Y_i] = Y_i$. Let's see some examples of the Exposure martingale.

Definition 7.3 (The edge exposure martingale)**.** *Consider the random graph $G(n,p)$. Let $e_1,\dots,e_m$ be an enumeration of pairs in $\binom{[n]}{2}$ and let f be any graph theoretic function, and let G be the random graph drawn from the distribution $G(n,p)$. Let I_i be the indicator random variable for the event $e_i \in E(G)$. Let $X_i = \mathbb{E}[f(G) : I_1,\dots,I_i]$, then $X_0,\dots,X_m$ is an martingale called the edge exposure martingale.*

Definition 7.4 (The vertex exposure martingale)**.** *Consider the random graph $G(n,p)$. Let f be any graph theoretic function, and $v_1,\dots,v_n$ be an enumeration of $[n]$ and let G be the random graph drawn from the distribution $G(n,p)$. Let $I^i \in \{0,1\}^{i-1}$ be a vector where I^i_j is the indicator random variable for the event $v_jv_i \in E(G)$ for each $j \le i-1$. Let $X_i = \mathbb{E}[f(G) : I_1,\dots,I_i]$, then $X_0,\dots,X_m$ is an martingale called the vertex exposure martingale.*

7.1. **Concentration inequality.** Where does the name 'martingale' comes from? In a martingale $X_0,\dots,X_m$, when X_i is fixed, X_{i+1} is somewhat tied to X_i. So X_{i+1} tends to not to be deviated too much from X_i, so hopefully we might be able to anticipate the value of X_m with a good probability. However this is not true in general. It could be that X_{i+1} is $X_i + M$ or $X_i - M$ with probability $1/2$ where M is very large, then our anticipation would be always wrong with probability more than $1/2$. However, as long as we can limit the deviation of X_{i+1} from X_i, we can prove the following concentration inequality.

Theorem 7.5 (Azuma's inequality)**.** *Let $X_0,\dots,X_m$ be a martingale with $|X_{i+1}-X_i| \le c_i$ for all $i < m$. For $\lambda > 0$, we have*

$$\mathbf{Pr}[|X_m - X_0| > \lambda] \le 2\exp(-\frac{\lambda^2}{2\sum_i c_i^2}).$$

Proof.

Claim 7. *For a random variable X with $\mathbb{E}[X] = 0$ and $|x| \le c$, we have*

$$\mathbb{E}[e^X] \le \frac{e^c + e^{-c}}{2} \le e^{c^2/2}.$$

Proof. Note that the function e^x is convex function on $[-c,c]$, hence we have $e^x \le \frac{1}{2c}((c-x)e^c + (c+x)e^{-c}) = \frac{e^c+e^{-c}}{2} + \frac{e^c-e^c}{2c}x$. This implies that

$$\mathbb{E}[e^X] \le \frac{e^c + e^{-c}}{2} \le e^{c^2/2}.$$

The last inequality can be checked using Taylor expansion. □

Replace X_i with $X_i - X_0$ if neccesary to assume $X_0 = 0$. Let $Y_i = X_i - X_{i-1}$, then we have $|Y_i| \le c_i$ and $\mathbb{E}[Y_i : X_0,\dots,X_{i-1}] = 0$, hence we have

$$\mathbb{E}[e^{tY_i} \mid X_0,\dots,X_{i-1}] \le e^{t^2c_i^2/2}.$$

Moreover, we have

$$\mathbb{E}[e^{tX_i}] = \mathbb{E}[e^{tX_{i-1}+tY_i}] = \mathbb{E}[\mathbb{E}[e^{tY_i} \mid X_{i-1}]e^{tX_{i-1}}] \le e^{t^2c_i^2/2}\mathbb{E}[e^{tX_{i-1}}].$$

We repeat this for each $i = m, m-1,\dots,1$, then we obtain

$$\mathbb{E}[e^{tX_m}] \le \exp(\frac{t^2\sum_i c_i^2}{2}).$$

Hence, by Markov's inequality, and applying $t = \lambda(\sum c_i^2)^{-1}$, we have

$$\mathbf{Pr}[X_n \geq \lambda] \leq e^{-t\lambda}\mathbb{E}[e^{tX_n}] \leq \exp(-t\lambda + \frac{t^2\sum_i c_i^2}{2}) = \exp(-\frac{\lambda^2}{2\sum c_i^2}).$$

By a symmetric argument, we also obtain

$$\mathbf{Pr}[X_n \leq \lambda] \leq \exp(-\frac{\lambda^2}{2\sum c_i^2}).$$

which yields the desired inequality □

Let $\Omega = A^B$ be the set of functions $g : B \to A$ and we define a measure by fixing values p_{ab} and setting $\mathbf{Pr}[g(b) = a] = p_{ab}$ where the values $g(b)$ are mutually independent and $\sum_{a\in A} p_{ab} = 1$. We fix a sequence of sets (called a gradation)

$$\emptyset = B_0 \subseteq B_1 \subseteq \cdots \subseteq B_m = B.$$

Let $L : A^B \to \mathbb{R}$ be a function and for each $h \in A^B$, we define a martingale $X_0, X_1, \ldots, X_m$ with

$$X_i(h) = \mathbb{E}[L(g) : g(b) = h(b) \text{ for all } b \in B_i].$$

Definition 7.6. *For $\mathbf{c} = (c_1, \ldots, c_m)$, a function L satisfies the $\mathbf{c}$-Lipschitz condition relative to the gradation if for each i, we have*

$$h, h' \text{ differs only on } B_{i+1} - B_i \Rightarrow |L(h) - L(h')| \leq c_i.$$

The c-Lipschitz condition indicates $(c, \ldots, c)$-Lipschitz condition and the Lipschitz condition indicates 1-Lipschitz condition.

Theorem 7.7. *Let L satisfy the $\mathbf{c}$-Lipschitz condition relative to the gradation $\emptyset = B_0 \subseteq \cdots \subseteq B_m = B$, then the corresponding martingale $X_i(h) = \mathbb{E}[L(g) : g(b) = h(b)$ for all $b \in B_i]$ satisfy $|X_{i+1}(h) - X_i(h)| \leq c_i$ for all $i < m$ and $h \in A^B$.*

Proof. Let $g \in A^B$ be a random function drawn from A^B according to the measure defined before. $H_j := \{h' \in A^B : h(b) = h'(b) \text{ for all } b \in B_j\}$, then we have

$$X_{i+1}(h) = \sum_{h'\in H_{i+1}} L(h')\mathbf{Pr}[g = h' \mid g(b) = h(b) \text{ for all } b \in B_{i+1}] = \sum_{h'\in H_{i+1}} L(h')w_{h'},$$

where $w_{h'} = \mathbf{Pr}[g = h' \mid g(b) = h(b) \text{ for all } b \in B_{i+1}]$. For each $h' \in H_{i+1}$, let $H[h'] = \{h^* \in A^B : h^*(b) = h'(b) \text{ for all } b \notin B_{i+1} - B_i\}$. Then, we have

$$\begin{aligned} X_i(h) &= \sum_{h''\in H_i} L(h'')\mathbf{Pr}[g = h'' \mid g(b) = h(b) \text{ for all } b \in B_i] \\ &= \sum_{h'\in H_{i+1}} \sum_{h^*\in H[h']} L(h^*)\mathbf{Pr}[\,{}^{g(b)=h^*(b)}_{\text{for all } b\in B_{i+1}} \mid {}^{g(b)=h^*(b)}_{\text{for all } b\in B_i}] \cdot \mathbf{Pr}[g = h^* \mid {}^{g(b)=h^*(b)}_{\text{for all } b\in B_{i+1}}] \\ &= \sum_{h'\in H_{i+1}} \sum_{h^*\in H[h']} L(h^*)q_{h^*}w_{h'}, \end{aligned}$$

where $q_{h^*} = \mathbf{Pr}[\,{}^{g(b)=h^*(b)}_{\text{for all } b\in B_{i+1}} \mid {}^{g(b)=h^*(b)}_{\text{for all } b\in B_i}]$. Here, we obtain the last inequality because $h^* \in H[h']$, which implies $h'(b) = h^*(b)$ for all $b \in B \setminus B_{i+1}$ and the values $g(b)$ are

mutually independent. Thus

$$\begin{aligned}|X_{i+1}(h) - X_i(h)| &= \left|\sum_{h'\in H_{i+1}} w_{h'}\left(L(h') - \sum_{h^*\in H[h']} L(h^*)q_{h^*}\right)\right| \\ &\le \sum_{h'\in H_{i+1}} w_{h'} \sum_{h^*\in H[h']} \left|(L(h') - L(h^*))q_{h^*}\right| \\ &\le c_i \sum_{h'\in H_{i+1}} w_{h'} \sum_{h^*\in H[h']} q_{h^*} \le c_i.\end{aligned}$$

□

This together with Azuma's inequality yields the following.

Theorem 7.8. *Let L satisfy the* **c***-Lipschitz condition relative to a gradation $\emptyset = B_0 \subseteq \cdots \subseteq B_m = B$. Then for all λ,*

$$\mathbf{Pr}[|L(g) - \mu| \ge \lambda] \le 2\exp(-\frac{\lambda^2}{2\sum_i c_i^2}).$$

Let's consider some easy applications of Martingale. Let g be the random function from $[n]$ to $[n]$ chosen from all n^n possible functions uniformly at random. Let $L(g)$ be the number of values not in the range of g, meaning $L(g) = |[n]\setminus g([n])|$. Then the expectation can be computed by using the linearity of expectation that $\mathbb{E}[L(g)] = n(1-\frac{1}{n})^n = \frac{n\pm 1}{e}$. Let $B_0 = \emptyset \subseteq \cdots \subseteq B_n$ with $B_i = [i]$ be a gradation. It is easy to see that L satisfies the Lipschitz condition relative to this gradation as changing the value of $g(i)$ can change $L(g)$ by at most 1. Thus, Azuma's inequality shows the following concentration.

Theorem 7.9. $\mathbf{Pr}[|L(g) - \frac{n}{e}| > \lambda\sqrt{n} + 1] \le 2e^{-\lambda^2/2}$.

7.2. **Clique number and chromatic number.** Now we consider the concentration of a graph parameter of random graphs. Of course, it varies over the resulting graph $G \sim G(n, 1/2)$, but we can show that the $\chi(G)$ is very close to some number with high probability.

Let $f(x) = \binom{n}{x}2^{-\binom{x}{2}}$ and let k_0 be the number satisfying $f(k_0 - 1) > 1 \ge f(k_0)$. Let $k = k_0 - 4$, then we have $k = (2 + o(1))\log_2 n$ and $f(k) > n^{3+o(1)}$.

Theorem 7.10. *There exists a constant $c > 0$ with* $\mathbf{Pr}[\omega(G) < k] \le e^{-\frac{cn^2}{\log^8 n}}$.

Proof. To prove this, we want to show a concentration of $\omega(G)$. However, this is difficult to control. Because, if we want to prove that it is concentrated near $\mathbb{E}[\omega(G)] \simeq \Theta(\log n)$, then the Azuma's inequality give an upper bound of the probability $O(e^{-\varepsilon^2\mathbb{E}[\alpha(G)]^2})$ which is not small enough for our purpose. Hence, we consider the following random variable. Let $Y = Y(G)$ be the maximum size of a family of of edge-disjoint cliques of size k. In this way, the $\mathbb{E}[Y]$ becomes much larger, so we can show better concentration.

Claim 8. $\mathbb{E}[Y] \ge (1 + o(1))\frac{n^2}{2k^2}$.

Proof. Let $\mathcal{K}$ denote the family of k-cliques of G, then $\mu := \mathbb{E}[|\mathcal{K}|] = \binom{n}{k}2^{-\binom{k}{2}} = f(k) > n^{3+o(1)}$. Let W be the number of unordered pair $\{C_1, C_2\}$ of k-cliques of G with $|C_1 \cap C_2| >$

1. Then

$$\begin{aligned}\mathbb{E}[W] &= \sum_{A\neq B\in\binom{n}{k},|A\cap B|>1} \mathbf{Pr}[G[A]=K_k \wedge G[B]=K_k] \\ &= \frac{1}{2}\sum_{A}\mathbf{Pr}[G[A]=K_k]\sum_{B:|A\cap B|>1}\mathbf{Pr}[G[B]=K_k \mid G[A]=K_k] = \frac{1}{2}\mu\Delta^*.\end{aligned}$$

where $\Delta^* = \sum_{B:|A\cap B|>1}\mathbf{Pr}[G[B]=K_k \mid G[A]=K_k]$, then we have

$$\Delta^* = \sum_{i=2}^{k-1}\binom{k}{i}\binom{n-k}{k-i}2^{\binom{i}{2}-\binom{k}{2}} = \mu\sum_{i=2}^{k-1} g(i),$$

where $g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}}2^{\binom{i}{2}}$. Here $g(2) = (1+o(1))k^4/n^2$ and $g(k-1) = (2+o(1))kn2^{-k}/\mu < o(g(2))$ as $\mu > n^{3+o(1)}$ and $2^{-k} = n^{-2+o(1)}$. By some calculations, one can show that $\sum_{3\le i\le k-1} g(i) = o(g(2))$. Hence, we have

$$\mathbb{E}[W] = (1+o(1))\frac{\mu^2k^4}{2n^2}.$$

Now we use alteration method. For each $C\in\mathcal{K}$, we add it to $\mathcal{C}$ independently at random with probability q, which we determine later. Let W' be the number of unordered pair $\{C_1,C_2\}$ with $C_1,C_2\in\mathcal{C}$ with $|C_1\cap C_2|>1$. Then

$$\mathbb{E}[W'] = q^2\mathbb{E}[W] = (1+o(1))\frac{\mu^2k^4q^2}{2n^2}.$$

We delete one set of each such pair $\{C_1,C_2\}$ from $\mathcal{C}$ to obtain a set $\mathcal{C}^*$ of edge-disjoint k-cliques of G. Then

$$\mathbb{E}[Y] \ge \mathbb{E}[|\mathcal{C}^*|] \ge \mathbb{E}[|\mathcal{C}|] - \mathbb{E}[W'] = \mu q - (1+o(1))\frac{\mu^2k^4q^2}{2n^2} = (1+o(1))\frac{n^2}{2k^4},$$

if we let $q = n^2/(\mu k^4)$. This proves the claim. □

Now, enumerate the pairs in $\binom{[n]}{2}$ into $e_1,\dots,e_m$ with $m=\binom{n}{2}$ and let $B_i=\{e_1,\dots,e_i\}$. Then the gradation $B_0=\emptyset\subseteq\cdots\subseteq B_m$ and $Y(G)$ satisfies the Lipschitz condition relative to this gradation as changing whether e_i lies in $G=G(n,1/2)$ only changes Y by 1. G has no k-clique if and only if $Y=0$, hence Azuma's inequality gives that

$$\begin{aligned}\mathbf{Pr}[\omega(G)<k] = \mathbf{Pr}[Y=0] &\le \mathbf{Pr}[|Y-\mathbb{E}[Y]|\le\mathbb{E}[Y]] \\ &\le 2e^{-\frac{\mathbb{E}[Y]^2}{2\binom{n}{2}}} \le e^{-(c'+o(1))n^2/k^8} \le e^{-(c+o(1))n^2/\log^8 n}.\end{aligned}$$

□

Now we use this to prove the following theorem.

Theorem 7.11 (Bollobás, 1988)**.** *With high probability, we have* $\chi(G) = (1+o(1))\frac{n}{2\log_2 n}$.

Proof. As $\alpha(G)=\omega(\overline{G})$ has the same distribution as $\omega(G)$ for $G=G(n,1/2)$, we have $\alpha(G)\le(2+o(1))\log_2 n$ with high probability. Hence with high probability we have

$$\chi(G) \ge \frac{n}{\alpha(G)} \ge (1+o(1))\frac{n}{2\log_2 n}.$$

To show upper bound, let $m = \lfloor \frac{n}{\log^2 n} \rfloor$. For any set S of m vertices the induced subgraph $G[S]$ has the distribution of $G(m, 1/2)$. Let $k = k_0(m) - 4$ as before, then $k = (2 + o(1)) \log_2 n$. Then $\mathbf{Pr}[\alpha(G[S]) < k] < e^{-\frac{cm^2}{\log^8 m}}$. As the number of such sets S is $\binom{n}{m} < n^m \leq 2^{\frac{n}{\log n}}$, we have

$$\mathbf{Pr}[\alpha(G[S]) < k \text{ for some } S \in \binom{[n]}{m}] < 2^{\frac{n}{\log_2 n}} e^{-\frac{cm^2}{\log^8 m}} = o(1).$$

Hence, with high probability, every m-sets contain an independent set of size k.

Suppose G has this property, then we take out an independent set of size k and take this as a color class. We repeat this until there are less than m vertices left. then we give each vertex a distinct color. Then we obtain that

$$\chi(G) \leq \lceil \frac{n-m}{k} \rceil + m \leq \frac{n}{k} + m \leq (1 + o(1)) \frac{n}{2 \log_2 n}.$$

This happens with high probability, so we have $\chi(G) = (1 + o(1)) \frac{n}{2 \log_2 n}$ with high probability. □

Moreover, if p is much smaller, then we can get a much better concentration of the chromatic number of $G(n, p)$. The following theorem says that $\chi(G)$ is one of four values with high probability.

Theorem 7.12. *Let $p = n^{-\alpha}$ with $\alpha > 5/6$ fixed. Let $G = G(n, p)$, then there exists $u = u(n, p)$ so that with high probability we have $u \leq \chi(G) \leq u + 3$.*

Proof.

Claim 9. *Fix $c > 0$. Then with high probability, every $c\sqrt{n}$ vertices of $G = G(n, p)$ can be three-colored.*

Proof. Let T be a minimal set of size t that is not three-colored. Then the minimality ensures that any vertex $x \in T$ has at least three other neighbors inside T. This implies that $G[T]$ must has at least $3t/2$ edges.

The probability that there exists a set T of size t with at most $c\sqrt{n}$ vertices having at least $3t/2$ edges is at most

$$\sum_{t=4}^{c\sqrt{n}} \binom{n}{t} \binom{\binom{t}{2}}{3t/2} p^{3t/2} \leq \sum_{t=4}^{c\sqrt{n}} (\frac{ne}{t} (\frac{te}{3})^{3/2} n^{-3\alpha/2})^t \leq \sum_{t=4}^{c\sqrt{n}} (cn^{\frac{3}{2}(5/6-\alpha)})^t = o(1).$$

This proves the claim. □

Let $\varepsilon > 0$ be an arbitrary positive number. We prove that if n is large enough, then we have $u \leq \chi(G) \leq u + 3$ for some u with probability at least $1 - 3\varepsilon$.

Let u be the least integer so that $\mathbf{Pr}[\chi(G) \leq u] > \varepsilon$.

Let $Y(G)$ be the minimal size of a set of vertices S for which $G - S$ can be u-colored. This $Y(G)$ satisfies vertex Lipschitz condition as changes on the edges incident to one vertex may change $Y(G)$ by at most one. We consider the vertex exposure martingale $Y_i = \mathbb{E}[Y : G[\{1, \ldots, i\}]]$. Let $\mu = \mathbb{E}[Y] = Y_0$, we have

$$\mathbf{Pr}[Y \leq \mu - \lambda\sqrt{n}] < e^{-\lambda^2/2} \quad \text{and} \quad \mathbf{Pr}[Y \geq \mu + \lambda\sqrt{n}] < e^{-\lambda^2/2}.$$

Let λ be the number such that $e^{-\lambda^2/2} = \varepsilon$.

By the definition of u, we have $Y = 0$ with probability at least ε. Hence, the first inequality ensures that $\mu \leq \lambda\sqrt{n}$ and the second inequality gives

$$\mathbf{Pr}[Y \geq 2\lambda\sqrt{n}] \leq \mathbf{Pr}[Y \geq \mu + \lambda\sqrt{n}] \leq \varepsilon.$$

Hence, with probability at least $1-\varepsilon$, there is a u-coloring of all but at most $c'\sqrt{n}$ vertices for some constant $c' > 0$ depending on ε. By the previous claim, with probability $1-o(1)$, every $c'\sqrt{n}$ vertices of G can be colored with 3 colors. Hence we have $\chi(G) \leq u+3$ with probability at least $1-\varepsilon-o(1) \geq 1-2\varepsilon$ for large n.

On the other hand, by the minimality of u ensures that $\mathbf{Pr}[\chi(G) \geq u] \geq 1-\varepsilon$. Hence, we have

$$\mathbf{Pr}[u \leq \chi(G) \leq u+3] \geq 1-3\varepsilon.$$

As $\varepsilon > 0$ can be arbitrary, we have $\chi(G) \in \{u, u+1, u+2, u+3\}$ with high probability. □

It is known that for $\alpha > 1/2$, $\chi(G)$ is concentrated on at most two values. For $G(n, 1/2)$ on the other hand, it is recently proved by Heckel that $\chi(G(n, 1/2))$ is not concentrated on fewer than $n^{1/4-o(1)}$ consecutive integers. Hence, such a strong concentration for $G(n, 1/2)$ does not hold.

7.3. **Talagrand's inequality and the concentration around the median.** So far, we have learned about concentrations around the expectation(=mean). One can also consider a concentration around the median.

For example, consider a Lipschitz graph theoretic function f where $f(G(n,1/2))$ has the median m. Then $\mathcal{A} = \{G : f(G) \leq m\}$ contains $\frac{1}{2} \cdot 2^{\binom{n}{2}}$ graphs. If f is a Lipschitz function, then the following would be a desired inequality leading towards a 'one-sided' concentration: at least $(1-\varepsilon)2^{\binom{n}{2}}$ graphs can be obtained from a graph in $\mathcal{A}$ by changing at most t adjacencies. As f is Lipschitz, this shows a one-sided concentration $\mathbf{Pr}[f(G(n,1/2)) \leq m+t] \geq 1-\varepsilon$ as we wished.

As before, the set of graphs can be generalized into a collection of subsets of $[n]$ for appropriate $n = \binom{|V|}{2}$. Thus, ths above discussion motivates to study the Hamming distance over $\{0,1\}^n$. Let ρ be the Hamming distance over $\{0,1\}^n$ and for $A \subseteq \{0,1\}^n$, let $B(A,s) = \{y \in \{0,1\}^n : \rho(x,y) \leq s \text{ for some } x \in A\}$. A simple application of Azuma's inequality yields the following.

Theorem 7.13. *Let $\lambda > 0$ and $\varepsilon = e^{-\lambda^2/2}$ and let $A \subseteq \{0,1\}^n$ with $|A| \geq \varepsilon 2^n$. Then we have $|B(A, 2\lambda\sqrt{n})| \geq (1-\varepsilon)2^n$.*

Proof. Choose a point z in $\{0,1\}^n$ uniformly at random. For each $y \in \{0,1\}^n$, let $X(y) = \min_{x \in A} \rho(x,y)$. Let $X_0, X_1, \ldots, X_n$ be the martingale by exposing one coordinate of z at a time. Then the Lipschitz condition holds for X as $|X(y) - X(y')| \leq 1$ if y, y' differ in only one coordinate. Thus, we have the following where $\mu = \mathbb{E}[X]$.

$$\mathbf{Pr}[X < \mu - \lambda\sqrt{n}] < e^{-\lambda^2/2} = \varepsilon \text{ and } \mathbf{Pr}[X > \mu + \lambda\sqrt{n}] < e^{-\lambda^2/2} = \varepsilon.$$

However, we have $\mathbf{Pr}[X = 0] = |A|2^{-n} \geq \varepsilon$, this implies $\mu \leq \lambda\sqrt{n}$. Thus we have

$$\mathbf{Pr}[X > 2\lambda\sqrt{n}] < \varepsilon.$$

As our choice of z is uniformly at random, we have

$$|B(A, 2\lambda\sqrt{n})| = 2^n \mathbf{Pr}[X \leq 2\lambda\sqrt{n}] \geq (1-\varepsilon)2^n.$$

This proves the theorem. □

Let $B(s)$ denote a ball of radius s about $(0,0,\ldots,0)$. The above inequality can be rephrased as $|A| \geq |B(r)| \Rightarrow |B(A, 2\lambda\sqrt{n})| \geq |B(r + 2\lambda\sqrt{n})|$ where r is not so much smaller than $n/2$. Actually, this statement also holds when the condition on r is removed. Harper 1966 proved that $|A| \geq |B(r)| \to |B(A,s)| \geq |B(r+s)|$ holds for all r, s.

In application, the Hamming distance is not exactly what we want. Imagine that we are interested in the variable X which is the largest number of edge-disjoint k-cliques in a graph G. Once we delete an edge to a graph G, then the value of X might or might not change depending on which edge we delete. In other words, when we want to estimate $f(x)$ by using the Hamming distance between a point and x and a set A, some coordinates can be more important than the other coordinate. In order to better cope with this situation, we introduce the following definitions.

Let $\Omega = \prod_{i \in [n]} \Omega_i$ where each Ω_i is a probability space and Ω has the product measure. Let $A \subseteq \Omega$ and let $\vec{x} = (x_1, \ldots, x_n) \in \Omega$.

Definition 7.14. *Let $\rho(A, \vec{x})$ be the least value such that for any $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ with $|\vec{\alpha}| = 1$, there exists $\vec{y} = (y_1, \ldots, y_n) \in A$ with*

$$\sum_{x_i \neq y_i} \alpha_i \leq \rho(A, \vec{x}).$$

For any real $t \geq 0$, let $A_t = \{\vec{x} \in \Omega : \rho(A, \vec{x}) \leq t\}$.

Note that the above definition can deal with any choice $\vec{\alpha}$ of weightings on coordinates on Ω.

Theorem 7.15 (Talagrand's inequality)**.** *For any $A \subseteq \Omega$, we have*

$$\mathbf{Pr}[A](1 - \mathbf{Pr}[A_t]) \leq e^{-t^2/4}.$$

In particular, we have a function f on Ω with the median m, then $A = \{\vec{x} : f(\vec{x}) \leq m\}$ satisfies $\mathbf{Pr}[A] \geq 1/2$, and we obtain that $\mathbf{Pr}[A_t]$ is close to 1 when t is large. Hence providing a concentration from above around the median. Similarly, we can take $A = \{\vec{x} : f(\vec{x}) \geq m\}$ to obtain a concentration from below.

Definition 7.16. *Let $U(A, \vec{x}) = \{\vec{s} \in \{0, 1\}^n : \exists \vec{y} \in A$ such that $x_i \neq y_i \Rightarrow s_i = 1\}$. In other words, $\{i \in [n] : s_i = 1\}$ contains a set of coordinates we need to change to obtain a point in A from $\vec{x}$.*

Note that the Hamming distance between $\vec{x}$ and $\vec{y} \in A$ is same with $|\vec{s}|$ for some $\vec{s} \in U(A, \vec{x})$. As $\rho(A, \vec{x})$ is somehow 'convex combination' of normal Hamming distance, the following theorem shows that $\rho(A, \vec{x})$ is the norm of a 'convex combination' of the points $\vec{y}$ in $U(A, \vec{x})$. This provides an alternative characterization of ρ. Here, $\mathrm{conv}(X)$ denotes the convex hull of X which is the smallest convex set containing X. We write $V(A, \vec{x}) = \mathrm{conv}(U(A, \vec{x})$.

Theorem 7.17.

$$\rho(A, \vec{x}) = \min_{\vec{v} \in V(A, \vec{x})} |\vec{v}|.$$

Proof. Note that $\rho(A, \vec{x})$ is the least real number so that for all $\vec{\alpha}$ with $|\vec{\alpha}| = 1$ there exists $\vec{s} \in U(A, \vec{x})$ with $\vec{\alpha} \cdot \vec{s} \leq \rho(A, \vec{x})$.

Let $\vec{v} \in V(A, \vec{x})$ be the point achieving the minimum. Consider the line from the origin to $\vec{v}$ and the hyperplane P through $\vec{v}$ orthogonal to the line. Then P separates $V(A, \vec{x})$ from the origin so that all $\vec{s} \in V(A, \vec{x})$ have $\vec{s} \cdot \vec{v} \geq \vec{v} \cdot \vec{v}$.

Let $\vec{\alpha} = \frac{1}{|\vec{v}|}\vec{v}$, then all $\vec{s} \in U(A, \vec{x})$ satisfies $\vec{s} \cdot \vec{\alpha} \geq \frac{\vec{v} \cdot \vec{v}}{|\vec{v}|} = |\vec{v}|$. This proves $\rho(A, \vec{x}) \geq \min_{\vec{v} \in V(A, \vec{x})} |\vec{v}|$.

Conversely, take any $\vec{\alpha}$ with $|\vec{\alpha}| = 1$, then we have $\vec{\alpha} \cdot \vec{v} \leq |\vec{v}|$. As we have $\vec{v} \in V(A, \vec{x})$, we may write $\vec{v} = \sum \lambda_i \vec{t_i}$ for some $t_i \in U(A, \vec{x})$ with all $\lambda_i \geq 0$ and $\sum \lambda_i = 1$. Then we have

$$|\vec{v}| \geq \sum \lambda_i (\vec{\alpha} \cdot \vec{t_i})$$

and hence there exists i such that $\vec{\alpha} \cdot \vec{t_i} \leq |\vec{v}|$. This proves that for any $\vec{\alpha}$ with $|\vec{\alpha}| = 1$, there exists $\vec{s} \in U(A, \vec{x})$ with $\vec{\alpha} \cdot \vec{s} \leq |\vec{v}|$. Hence, we have $\rho(A, \vec{x}) \leq \min_{\vec{v} \in V(A, \vec{x})} |\vec{v}|$. □

In case when $\Omega = \{0,1\}^n$, then $U(A,\vec{0})$ is same as A and $U(A,\vec{x})$ is the set A' obtained from A by flipping coordinates on $\{i \in [n] : x_i = 1\}$. Hence $\rho(A,\vec{x})$ is the Euclidean distance from $\vec{x}$ to the convex hull of A'.

Theorem 7.18.

$$\int_\Omega \exp\left[\frac{1}{4}\rho^2(A,\vec{x})\right] d\vec{x} \leq \frac{1}{\mathbf{Pr}[A]}.$$

Proof. We use induction on n. For $n = 1$, $\rho(A,\vec{x})$ is 1 if $\vec{x} \notin A$ and zero otherwise. So we have

$$\int_\Omega \exp\left[\frac{1}{4}\rho^2(A,\vec{x})\right] d\vec{x} = \mathbf{Pr}[A] + (1-\mathbf{Pr}[A])e^{1/4} \leq \frac{1}{\mathbf{Pr}[A]}$$

as the inequality $u + (1-u)e^{1/4} \leq u^{-1}$ for $0 < x \leq 1$ holds.

Assume we have the result for n and consider $\Omega = \prod_{i\in[n+1]} \Omega_i$. Write $\Omega^n = \prod_{i\in[n]} \Omega_i$. Then any $z \in \Omega$ can be uniquely written $z = (x,w)$ with $x \in \Omega^n$ and $w \in \Omega_{n+1}$. Let

$$B = \{x \in \Omega^n : (x,w) \in A \text{ for some } w \in \Omega_{n+1}\}$$

be the projection of A into Ω^n. For each $w \in \Omega_{n+1}$, let

$$A_w = \{x \in \Omega^n : (x,w) \in A\}.$$

From this definition, we can observe the followings.

If $\vec{s} \in U(B,x)$, then $(\vec{s},1) \in U(A,(x,w))$.

If $\vec{t} \in U(A_w,x)$, then $(\vec{t},0) \in U(A,(x,w))$.

Hence, if $\vec{s} \in V(B,x)$ and $\vec{t} \in V(A_w,x)$, then $(\vec{s},1)$ and $(\vec{t},0)$ are in $V(A,(x,w))$. Thus, for given $\lambda \in [0,1]$, we have

$$((1-\lambda)\vec{s} + \lambda\vec{t}, 1-\lambda) \in V(A,(x,w)).$$

The previous theorem implies that

$$\rho^2(A,(x,w)) \leq |(1-\lambda)\vec{s} + \lambda\vec{t}|^2 + (1-\lambda)^2 \leq (1-\lambda)|\vec{s}|^2 + \lambda|\vec{t}|^2 + (1-\lambda)^2.$$

By choosing $\vec{s}$ and $\vec{t}$ satisfying $|\vec{s}| = \rho(B,x)$ and $|\vec{t}| = \rho(A_w,x)$, we obtain

$$\rho^2(A,(x,w)) \leq (1-\lambda)^2 + \lambda\rho^2(A_w,x) + (1-\lambda)\rho^2(B,x). \tag{7.1}$$

Now we fix w and use the above to compute the integral as follows.

$$\int_x \exp[\frac{1}{4}\rho^2(A,(x,w))] \leq e^{(1-\lambda)^2/4} \int_x (\exp[\frac{1}{4}\rho^2(A_w,x)])^\lambda (\exp[\frac{1}{4}\rho^2(B,x)])^{1-\lambda}.$$

Using Hölder's inequality, this is at most

$$e^{(1-\lambda)^2/4} \left[\int_x \exp[\frac{1}{4}\rho^2(A_w,x)]\right]^\lambda \left[\int_x \exp[\frac{1}{4}\rho^2(B,x)]\right]^{1-\lambda}.$$

Using induction hypothesis, this is at most

$$e^{(1-\lambda)^2/4} (\frac{1}{\mathbf{Pr}[A_w]})^\lambda (\frac{1}{\mathbf{Pr}[B]})^{1-\lambda} \leq \frac{1}{\mathbf{Pr}[B]} e^{(1-\lambda)^2/4} r^{-\lambda},$$

where $r = \mathbf{Pr}[A_w]/\mathbf{Pr}[B] \leq 1$. Now we wish to choose $\lambda = 1 + 2\ln r$ for $e^{-1/2} \leq r \leq 1$ and $\lambda = 0$ otherwise to optimize the above expression. By tedious calculations, we can check $e^{(1-\lambda)^2/4} r^{-\lambda} \leq 2 - r$ for the above choice of λ. Thus we have

$$\int_x \exp[\frac{1}{4}\rho^2(A,(x,w))] \leq \frac{1}{\mathbf{Pr}[B]}(2 - \frac{\mathbf{Pr}[A_w]}{\mathbf{Pr}[B]}).$$

Now, we integrate this over w, then we have

$$\int_{(x,w)\in\Omega} \exp[\frac{1}{4}\rho^2(A,(x,w))] \leq \frac{1}{\mathbf{Pr}[B]}(2 - \frac{\mathbf{Pr}[A]}{\mathbf{Pr}[B]}) \leq \frac{1}{\mathbf{Pr}[A]} x(2-x),$$

where $x = \mathbf{Pr}[A]/\mathbf{Pr}[B] \in [0,1]$. As $x(2-x) \leq 1$, this completes the induction step and hence proves the theorem. □

Using the above theorem, we can prove Talagrand's inequality as follows. Fix A and consider a random variable $X = \rho(A, \vec{x})$ where $\vec{x}$ is chosen at random according to the given probability distribution over Ω. Then Markov's inequality yields

$$\mathbf{Pr}[\Omega \setminus A_t] = \mathbf{Pr}[X \geq t] = \mathbf{Pr}[e^{X^2/4} \geq e^{t^2/4}] \leq \mathbb{E}[e^{X^2/4}] e^{-t^2/4}.$$

The above theorem states that $\mathbb{E}[e^{X^2/4}] \leq 1/\mathbf{Pr}[A]$ and this finishes the proof.

Now we consider applications of Talagrand's inequality. In $\Omega = \prod_{i\in[n]} \Omega_i$, we call $h : \Omega \to \mathbb{R}$ a *K-Lipschitz* function if $|h(x) - h(y)| \leq K$ whenever x, y differ in at most one coordinate.

Definition 7.19. *Let $f : \mathbb{N} \to \mathbb{N}$ and $h : \Omega \to \mathbb{R}$. We say that h is f-certifiable if whenever $h(x) \geq s$, there exists $I \subseteq [n]$ with $|I| \leq f(s)$ so that all $y \in \Omega$ that agree with x on the coordinates I have $h(y) \geq s$.*

As our choice of weighted Hamming distance was for arbitrary $\vec{\alpha}$, we can derive the following corollary.

Theorem 7.20. *Let $f : \mathbb{N} \to \mathbb{N}$ and $h : \Omega \to \mathbb{R}$ be a K-Lipschitz function which is f-certifiable. Let x be a point randomly chosen from Ω according to the product measure and $X = h(x)$ be a random variable. Then, for any b, t, we have*

$$\mathbf{Pr}\left[X \leq b - tK\sqrt{f(b)}\right] \mathbf{Pr}\left[X \geq b\right] \leq e^{-t^2/4}.$$

Proof. Let $A = \{x : h(x) < b - tK\sqrt{f(b)}\}$. Suppose that $h(y) \geq b$, we claim $y \notin A_t$. As h is f-certifiable, there exists a set I of indices of size at most $f(b)$ that certifies $h(y) \geq b$. Let $\alpha_i = 0$ for $i \notin I$ and $\alpha_i = |I|^{-1/2}$ when $i \in I$.

Suppose $y \in A_t$. As $\rho(A, y) \leq t$, by the definition of ρ, there exists $z \in A$ such that $\sum_{i\in I: y_i \neq z_i} |I|^{-1/2} \leq t$. Thus, z differs from y in at most $t\sqrt{I} \leq t\sqrt{f(b)}$ coordinates of I. Let y' agree with y on I and agree with z outside I. Then, by the certification, $h(y') \geq b$. Now, y' and z differ in at most $t\sqrt{f(b)}$ coordinates and so by the Lipschitz condition, we have

$$h(z) \geq h(y') - tK\sqrt{f(b)} \geq b - tK\sqrt{f(b)},$$

but then $z \notin A$, which is a contradiction. So, $\mathbf{Pr}[X \geq b] \leq 1 - \mathbf{Pr}[A_t]$. By Talagrand's inequality, we have

$$\mathbf{Pr}[X < b - tK\sqrt{f(b)}]\mathbf{Pr}[X \geq b] \leq e^{-t^2/4}.$$

As this holds for arbitrary t and the right hand side is continuous, we can replace $<$ with $\leq$ as well. This proves the theorem. □

Again, this theorem yields a concentration of a random variable around its median. The median is much more difficult to compute than the mean in general, but tight concentration shows that the mean and median are not far away. Let's consider some applications of this theorem.

Theorem 7.21. *Let $x = (x_1, \dots, x_n)$ where each x_i is independently and uniformly chosen from $[0, 1]$. Let $X = h(x)$ be the length of the longest increasing subsequence of x. Then there exists $m = \Theta(n^{1/2})$ such that for any $t \geq 0$, we have*

$$\mathbf{Pr}[|X - m| > tn^{1/4}] \leq 4e^{-\Omega(t^2)}.$$

Proof. We will take m as the median of X. We first want to show that $m = \Theta(n^{1/2})$. For a given set $I \subseteq [n]$ with $|I| = k$, the probability that $(x_i : i \in I)$ induces an increasing subsequence is $\frac{1}{k!} \leq (\frac{e}{k})^k$. Also there are $\binom{n}{k} \leq (\frac{en}{k})^k$ choices of $I \subseteq [n]$ with size k. Hence, for $k = Cn^{1/2}$ for some large constant C, the probability that there are no increasing subsequence of size k is at least

$$1 - \frac{1}{k!}\binom{n}{k} \geq 1 - o(1).$$

Similarly, we can prove that with probability $1 - o(1)$ it does not contain any decreasing sequence of length at least $Cn^{1/2}$. Erdős–Szekeres theorem states that such a sequence of length n contains an increasing subsequence of length at least $\frac{1}{C}n^{1/2}$. This shows that $\mathbf{Pr}[\frac{1}{C}n^{1/2} \leq |X| \leq Cn^{1/2}] \geq 1 - o(1)$, hence we have $m = \Theta(n^{1/2})$.

Note that if we use Azuma's inequality, then the concentration we can get is a concentration with the window size $O(n^{1/2})$, but it is not so good as m itself is $\Theta(n^{1/2})$. So we use Talagrand's inequality.

X is f-certifiable with $f(s) = s$ as the s coordinates of the increasing subsequence of length s certify that $X \geq s$. Also f is a Lipschitz function. Let t' be such that $t'm^{1/2} = tn^{1/4}$. Then we have

$$\mathbf{Pr}[X < m - t'm^{1/2}] \leq e^{-t'^2/4}\mathbf{Pr}[X \geq m]^{-1} \leq 2e^{-t'^2/4}.$$

Let b be such that $b - t'b^{1/2} = m$, hence $b = m + (1 + o(1))t'm^{1/2}$. Then we have

$$\mathbf{Pr}[X > b] \leq e^{-t'^2/4}\mathbf{Pr}[X \leq m]^{-1} \leq 2e^{-t'^2/4}.$$

Combining this, we prove the theorem. □

Now we consider the clique number of the random graph $G(n, 1/2)$ again. Again, recall that we can consider this as a collection of $\binom{n}{2}$ independent random variables, where each variable is the random variable indicating each pair being an edge. Note that the following theorem yields a better concentration than Theorem 7.10.

Theorem 7.22. *Let G, k be as in Theorem 7.10. then* $\mathbf{Pr}[\omega(G) < k] \leq \exp[-\Omega(\frac{n^2}{\ln^6 n})]$.

Proof. Let Y be the maximal number of edge-disjoint k-cliques. By Theorem 7.10, we have $\mathbb{E}[Y] = \Omega(n^2k^{-4})$ and Y is tightly concentrated around its mean. Hence the median m of Y also satisfies $m = \Omega(n^2k^{-4})$ with $k = \Theta(\ln n)$ and Y is 1-Lipschitz. Also, Y is

f-certifiable with $f(x) = \binom{k}{2}s$ as the edges of the s-cliques certify that $Y \geq s$. Hence we have

$$\mathbf{Pr}[Y \leq m - tm^{1/2}\binom{k}{2}^{1/2}] < e^{-t^2/4}\mathbf{Pr}[Y \geq m]^{-1}.$$

Setting $t = \Theta(m^{1/2}/k)$ so that $m = tm^{1/2}\binom{k}{2}^{1/2}$, then we have

$$\mathbf{Pr}[\omega(G) < k] = \mathbf{Pr}[Y = 0] < 2e^{-t^2/4} \leq \exp[-\Omega(\frac{n^2}{\ln^6 n})].$$

□

7.4. **Kim-Vu Polynomial concentration.** We have learned about several concentration results. The simplest one was the Chernoff's bounds which states that the sum of independent random variables is concentrated around its mean. This is useful if, say, we want to show that the number of edges of $G(n,p)$ is concentrated as the number of edges is $\sum_{ij\in\binom{[n]}{2}} \mathbb{1}_{ij\in E(G(n,p))}$. In order to show that the number of triangles in $G(n,p)$ is concentrated, we have to consider $\sum_{ijk\in\binom{[n]}{3}} \mathbb{1}_{ij}\cdot\mathbb{1}_{ik}\cdot\mathbb{1}_{jk}$. This is a polynomial of independent random variables. As the Chernoff's bound is a general tool to prove a concentration when this polynomial is of degree 1, one might ask whether there's a general tool to deal with more general polynomials. The approach of Kim and Vu is often useful. Note that as we are dealing with indicator random variables, we have $\mathbb{1}_{ij}^k = \mathbb{1}_{ij}$ for any $k \geq 1$. Hence, we assume that our polynomial f is a multilinear polynomial.

Let $H = (V, E)$ be a hypergraph and let each edge $e \in E(H)$ have a nonnegative weight $w(e)$. Let $t_i, i \in V(H)$ be mutually independent indicator random variables with $\mathbb{E}[t_i] = p_i$. Let

$$Y = \sum_{e\in E(H)} w_e \prod_{i\in e} t_i.$$

We allow $e = \emptyset$, in this case, we let $\prod_{i\in e} t_i = 1$. This yields a constant term, which is not relevant to the concentration, but we allow constant term as it is convenient for us in several ways.

Let $S \subseteq V$ be a random set obtained by adding each $i \in V$ to S independently with probability p_i. Then Y is the weighted number of hyperedge e in the restriction $H[S]$ of H to S. Let $n = |V|$ and let k be the upper bound of the size of all hyperedges, which yields the upper bound on the degree of the polynomial Y.

For $A \subseteq V$ with $|A| \leq k$, we truncate Y to Y_A as follows: for those terms $\prod_{i\in e} t_i$ with $A \subseteq e$, we set $t_i = 1$ for all $i \in A$, replacing the term by $\prod_{i\in e-A} t_i$. All other terms corresponding to e not containing A are deleted. In other words, Y_A is a partial derivative of Y with respect to the $t_i, i \in A$. Let $E_A = \mathbb{E}[Y_A]$ and let $E_i := \max\{E_A : A \in \binom{V}{i}\}$ and let $\mu = \mathbb{E}[Y]$. Let

$$E' = \max_{1\leq i\leq k} E_i \text{ and } E = \max[\mu, E'].$$

Then

Theorem 7.23 (Kim-Vu polynomial concentration)**.** *With the above hypothesis and* $a_k = 8^k k!^{1/2}$ *and* $d_k = 2e^2$*, and for any* $\lambda > 1$*, we have*

$$\mathbf{Pr}\left[|Y - \mu| > a_k(EE')^{1/2}\lambda^k\right] < d_k e^{-\lambda} n^{k-1}.$$

We will not prove this theorem. To get the intuition, imagine we are revealing each indicator variable one by one. Once $t_1, \dots, t_i$ are revealed, we are above to choose t_{i+1}. Then the difference $|\mathbb{E}[Y : t_1, \dots, t_i, t_{i+1}] - \mathbb{E}[Y : t_1, \dots, t_i]|$ are somehow related to the values E_A where $i+1 \in A \subseteq \{1, \dots, i+1\}$. This gives us some intuition that the upper bound on E' above yielding concentration is somehow natural. We consider the following simple applications of Kim-Vu polynomial concentration.

Theorem 7.24. *Let $p = n^{-\alpha}$ with $0 < \alpha < 2/3$. Fix a vertex x of $G = G(n,p)$, an let $Y = Y(x)$ be the number of triangles containing x and $\mu = \mathbb{E}[Y] \sim \frac{1}{2}n^{2-3\alpha}$. Then there exists $\varepsilon = \varepsilon(\alpha) > 0$ depending on α such that for each $\delta > 0$, we have*

$$\mathbf{Pr}[|Y - \mu| > \delta\mu] \le \exp[-C(\delta)n^{\varepsilon}].$$

Proof. The random graph $G(n,p)$ is a collection of indicator random variable $t_{ij} = \mathbb{1}_{ij \in G}$. Then

$$Y = \sum_{ij \in \binom{[n]-\{x\}}{2}} t_{xi}t_{xj}t_{ij}.$$

This is a polynomial of degree 3. For $A = \{xi\}$, we have $E_A = (n-2)p^2$ and $A = \{xi, xj, ij\}$, we have $E_A = 1$, and when $A = \emptyset$, then we have $E_A = \mu$. For all other cases, we have smaller value of E_A. Hence $E' \le \max[np^2, 1]$. If $\alpha < 1/2$, then let $\varepsilon = \frac{1}{6}(1-\alpha)$ and if $\alpha \ge 1/2$, then let $\varepsilon = \frac{1}{6}(2-3\alpha)$. Then we have $E' \sim c\mu n^{-6\varepsilon}$. By using Kim-Vu polynomial concentration with $\lambda = c'n^{\varepsilon}$ with small positive constant $c' = c'(\delta)$, we have

$$\mathbf{Pr}[|Y - \mu| > \delta\mu] \le \exp[-\Omega(n^{\varepsilon})].$$

□

8. The Poisson Paradigm

A discrete random variable X is said to have a Poisson distribution with the mean μ if $\mathbf{Pr}[X = k] = \frac{\mu^k e^{-\mu}}{k!}$ for each $k = 0, 1, 2, \ldots$.

Assume there are n events and they are all independent and each happens with probability p. When n is large and p is small, then $np = \mu$ is the expected number of events happening. The number X of events happened, actually follows the binomial distribution that $\mathbf{Pr}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$. However, if n is getting larger, then we can check

$$\lim_{n\to\infty} \mathbf{Pr}[X = k] = \frac{\mu^k e^{-\mu}}{k!}.$$

Hence, if there are many rare events and we want to compute the probability that k of them happens, we can approximate the probability by using Poisson distribution. But this is under the condition that those events are independent. What if they are not independent, but dependency is somewhat weak?

8.1. **The Janson inequality.** Consider the random graph $G(n, p)$. What is the probability that it is triangle-free? It is easy to check that if $p = o(1/n)$ then it is almost surely triangle-free and if $p = \omega(1/n)$ then it is almost surely not triangle-free. Consider the case where $p = c/n$. There are $\binom{n}{3}$ distinct sets S of size three, and let E_S be the event that S does not form a triangle, then we have $\mathbf{Pr}[E_s] = 1 - p^3$. Note that when X is the number of triangles, then the expected number $\mathbb{E}[X]$ of triangles is $\binom{n}{3}p^3$. Note that for each S, satisfying E_S is a monotonically decreasing graph property, hence, if pn tends to zero, the correction inequality yields that

$$\mathbf{Pr}[G(n,p) \text{ is triangle-free}] = \mathbf{Pr}[X = 0] \geq (1-p^3)^{\binom{n}{3}} \sim e^{-\binom{n}{3}p^3} = e^{-\mathbb{E}[X]}.$$

This provides a lower bound, but how good is this lower bound? For most of pairs $S, S' \in \binom{[n]}{3}$, then event E_S and $E_{S'}$ are independent. To be more precise, as long as they intersect at ≤ 1 points, they are independent. Hence, intuitively, the lower bound also must be close to the upper bound as the 'bad events' are 'mostly' independent.

This situation is somewhat similar to the situation in using local lemma. In the chapter of local lemma, we proved that if bad events $B_1, \ldots, B_k$ have weak dependency, then we can prove $\mathbf{Pr}[\bigwedge_{i\in[k]} \overline{B}_i] > 0$. This provides a lower bound on the events $\bigwedge_{i\in[k]} \overline{B}_i$.

Consider the following set-up. Let Ω be a finite universal set and for each $r \in \Omega$, we independently at random add r to a set R with probability p_r. This yields a random subset R of Ω. For some index set I and each $i \in I$, let A_i be a subset of Ω and B_i be the event that $A_i \subseteq R$. Let X_i be the indicator random variable for B_i and $X = \sum_{i\in I} X_i$. Then the event $\bigwedge_{i\in[k]} \overline{B}_i$ is equivalent to $X = 0$. For $i, j \in I$, we write $i \sim j$ if B_i and B_j are not pairwise independent, in other words if $A_i \cap A_j \neq \emptyset$. Then B_i is mutually independent with $\{B_j : j \in J\}$ for any $J \subseteq \{j \in I : i \not\sim j\}$. We let

$$\Delta = \sum_{i\sim j} \mathbf{Pr}[B_i \wedge B_j] \text{ and } M = \prod_{i\in I} \mathbf{Pr}[\overline{B_i}].$$

Here, Δ is the sum over ordered pairs, thus $\Delta/2$ is the sum of unordered pairs. Let $\mu = \mathbb{E}[X]$.

Definition 8.1. *Let an event $\mathcal{E}$ be an increasing event if whenever R satisfies it and $R \subseteq R'$, then R' also satisfies this. It is a decreasing event if whenever R satisfies it and $R' \subseteq R$, then R' also satisfies this.*

With this definition, the events B_i above are increasing events. With this set-up, we can prove the following theorem.

Theorem 8.2 (The Janson inequality)**.** *Assume as above. If $\mathbf{Pr}[B_i] \le \varepsilon$ for each $i \in I$, then we have*

$$M \le \mathbf{Pr}[\bigwedge_{i\in I} \overline{B_i}] \le Me^{\frac{\Delta}{2(1-\varepsilon)}}$$

and

$$\mathbf{Pr}[\bigwedge_{i\in I} \overline{B_i}] \le e^{-\mu+\Delta/2}.$$

Note that $\mathbf{Pr}[\overline{B_i}] = 1 - \mathbf{Pr}[B_i] \le e^{-\mathbf{Pr}[B_i]}$, so multiplying this all together yields $M \le e^{-\mu}$. This shows that the above inequalities are good if Δ is small.

Proof. To estimate $\mathbf{Pr}[\bigwedge_{i\in I} \overline{B_i}]$, we want to estimate

$$\mathbf{Pr}[\overline{B_i} \mid \bigwedge_{j<i} \overline{B_j}] = 1 - \mathbf{Pr}[B_i \mid \bigwedge_{j<i} \overline{B_j}]$$

and multiply them later. Fix i and let $D_0 = \bigwedge_{j<i:j\not\sim i} \overline{B_j}$ and $D_1 = \bigwedge_{j<i:j\sim i} \overline{B_j}$. Then we have

$$\begin{aligned}\mathbf{Pr}[B_i \mid D_0 \wedge D_1] &= \frac{\mathbf{Pr}[B_i \wedge D_0 \wedge D_1]}{\mathbf{Pr}[D_0 \wedge D_1]} \ge \frac{\mathbf{Pr}[B_i \wedge D_0 \wedge D_1]}{\mathbf{Pr}[D_0]} = \mathbf{Pr}[B_i \wedge D_1 \mid D_0] \\ &= \mathbf{Pr}[B_i \mid D_0] - \mathbf{Pr}[B_i \wedge \overline{D_1} \mid D_0].\end{aligned}$$

Note that B_i is independent of D_0, so $\mathbf{Pr}[B_i \mid D_0] = \mathbf{Pr}[B_i]$. Also as $B_i \wedge \overline{D_1}$ is an increasing event and D_0 is a decreasing event, so by Theorem 6.10 we have

$$\mathbf{Pr}[B_i \wedge \overline{D_1} \mid D_0] \le \mathbf{Pr}[B_i \wedge \overline{D_1}] = \mathbf{Pr}[B_i \wedge \bigvee_{j<i, j\sim i} B_j] \le \sum_{j<i, j\sim i} \mathbf{Pr}[B_i \wedge B_j]. \tag{8.1}$$

Now, this yields

$$\mathbf{Pr}[\overline{B_i} \mid D_0 \wedge D_1] \le \mathbf{Pr}[\overline{B_i}] + \sum_{j<i, j\sim i} \mathbf{Pr}[B_i \wedge B_j]. \tag{8.2}$$

Let's consider the first inequality. As $\mathbf{Pr}[\overline{B_i}] \ge 1-\varepsilon$, using the fact $1+x \le e^x$, we have

$$\mathbf{Pr}[\overline{B_i} \mid \bigwedge_{j<i} \overline{B_j}] \le \mathbf{Pr}[\overline{B_i}] \exp(\frac{1}{1-\varepsilon} \sum_{j<i, j\sim i} \mathbf{Pr}[B_i \wedge B_j]).$$

Multiplying this for all i yields that we have

$$\mathbf{Pr}[\bigwedge_{i\in I} \overline{B_i}] \le Me^{\frac{\Delta}{2(1-\varepsilon)}}.$$

For the second bound, we use (8.2) as

$$\mathbf{Pr}[\overline{B_i} \mid D_0 \wedge D_1] \le 1 - \mathbf{Pr}[B_i] + \sum_{j<i, j\sim i} \mathbf{Pr}[A_i \wedge A_j] \le \exp(-\mathbf{Pr}[B_i] + \sum_{j<i, j\sim i} \mathbf{Pr}[A_i \wedge A_j])$$

and multiplying this for all i, then we obtain the desired bound.

□

The above bound is good if Δ is small. However if $\Delta \geq 2\mu$, then the upper bound becomes useless. For those cases, the following theorem yields a better bound.

Theorem 8.3 (The extended Janson inequality)**.** *In the above set-up and with the further assumption that $\Delta \geq \mu$, we have*

$$\mathbf{Pr}[\bigwedge_{i\in I} \overline{B_i}] \leq e^{-\frac{\mu^2}{2\Delta}}.$$

Before, we saw that $\mathrm{Var}[X] \leq \mu + \Delta$. Note that the second moment method yields that

$$\mathbf{Pr}[\bigwedge_{i\in I} \overline{B_i}] = \mathbf{Pr}[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2} \leq \frac{\mu+\Delta}{\mu^2}.$$

Suppose $\mu \to \infty$ and $\mu^2/\Delta \to \infty$. However, this is roughly $(\mu^2/\Delta)^{-1}$ while the Janson inequality yields $e^{-\mu^2/\Delta}$.

Proof. Fix a set $S \subseteq I$ of index, and order the indices in S. With this, the inequality (8.2) holds. By multiplying these for all $i \in S$ and taking logarithm, we obtain

$$-\ln\left(\mathbf{Pr}[\bigwedge_{i\in S} \overline{B_i}]\right) \geq \sum_{i\in S} \mathbf{Pr}[B_i] - \frac{1}{2} \sum_{i,j\in S, i\sim j} \mathbf{Pr}[B_i \wedge B_j].$$

For each $i \in I$, we independently add i to S with probability p which we will determine later. For this random subset S, we can compute the expectation of the above expression.

$$\mathbb{E}\left[-\ln\left(\mathbf{Pr}[\bigwedge_{i\in S} \overline{B_i}]\right)\right] \geq \mathbb{E}\left[\sum_{i\in S} \mathbf{Pr}[B_i]\right] - \frac{1}{2}\mathbb{E}\left[\sum_{i,j\in S, i\sim j} \mathbf{Pr}[B_i \wedge B_j]\right] = p\mu - \frac{p^2\Delta}{2}.$$

We let $p = \mu/\Delta$, then this p is at most 1. Then $p\mu - \frac{p^2\Delta}{2} \geq \frac{\mu^2}{2\Delta}$. Thus, this shows that there exists a set $S \subseteq I$ for which we have $-\ln\left(\mathbf{Pr}[\bigwedge_{i\in S} \overline{B_i}]\right) \geq \frac{\mu^2}{2\Delta}$, which implying

$$\mathbf{Pr}[\bigwedge_{i\in I} \overline{B_i}] \leq \mathbf{Pr}[\bigwedge_{i\in S} \overline{B_i}] \leq e^{-\frac{\mu^2}{2\Delta}}.$$

□

Forget the previous set-up and we only assume that the events B_i are monotonically increasing events. We define $i \sim j$ only when B_i and B_j are not pairwise independent, and let $\Delta = \sum_{i\sim j} \mathbf{Pr}[B_i \wedge B_j]$. Then the Janson inequality holds as the proof works same. To check that the proof works as it is, there are one point we need to check, which is that B_i is independent of D_0, meaning $\mathbf{Pr}[B_i \mid D_0] = \mathbf{Pr}[B_i]$ in the above proof. One subtlety is that our definition of $\sim$ is regarding pairwise independence, and what we require is mutual independence. In the previous set-up, as each B_i consists of independent random variables, this mutual independence was obvious. However, in the current set-up we need to check the following.

Proposition 8.4. *Let A, B, C be increasing events. If A is independent of each of B and C, then A is independent of $B \wedge C$.*

Proof. We have

$$\mathbf{Pr}[A \wedge (B \wedge C)] + \mathbf{Pr}[A \wedge (B \vee C)] = \mathbf{Pr}[A \wedge B] + \mathbf{Pr}[A \wedge C] = \mathbf{Pr}[A](\mathbf{Pr}[B] + \mathbf{Pr}[C]). \tag{8.3}$$

One the other hand, as $B \wedge C$ and $B \vee C$ are both increasing events, the FKG inequality implies that

$$\mathbf{Pr}[A \wedge (B \wedge C)] \geq \mathbf{Pr}[A]\mathbf{Pr}[B \wedge C] \text{ and } \mathbf{Pr}[A \wedge (B \vee C)] \geq \mathbf{Pr}[A]\mathbf{Pr}[B \vee C].$$

The sum of the left hand side is the left hand side of (8.3) and the sum of the right hand side is the right hand side of (8.3). Hence, we have equalities

$$\mathbf{Pr}[A \wedge (B \wedge C)] = \mathbf{Pr}[A]\mathbf{Pr}[B \wedge C] \text{ and } \mathbf{Pr}[A \wedge (B \vee C)] = \mathbf{Pr}[A]\mathbf{Pr}[B \vee C].$$

This proves the proposition. □

With this proposition, we have $\mathbf{Pr}[B_i \mid D_0] = \mathbf{Pr}[B_i]$ in the above proof. we have the Janson inequality over all increasing events (or decreasing events) when a log-supermodular measure is given over a finite distributive lattice.

Now we are back to the probability of $G(n,p)$ being triangle-free. In this case, we have $\mu = \binom{n}{3}p^3$ and $M = (1-p^3)^{\binom{n}{3}} = e^{-(1+o(1))\mu}$. Also two sets S, T in $\binom{[n]}{3}$ satisfies $S \sim T$ only when $|S \cap T| = 2$. As there are $6\binom{n}{4} = O(n^4)$ pairs S, T with $S \sim T$. Moreover, for such pairs S, T the joint probability $\mathbf{Pr}[B_S \wedge B_T]$ is p^5. Thus $\Delta = O(n^4)p^5$. Note that if $p = o(n^{-1/2})$, then $\Delta = o(\mu)$, so Janson's inequality yields that

$$\mathbf{Pr}[G(n,p) \text{ is triangle-free}] = \exp(-(1+o(1))\mu).$$

What if $p = \Omega(n^{-1/2})$? In this case, we have $\Delta \geq \mu$. So we use the extended Janson inequality, then the probability that $G(n,p)$ is triangle-free is at most $\exp[-\frac{\mu^2}{2\Delta}] \leq e^{-\Theta(n^2p)}$. On the other hand, the probability that $G(n,p)$ is triangle-free is at least the probability that it is an empty graph, which is $(1-p)^{\binom{n}{2}} \geq e^{-\Theta(n^2p)}$, so this is also tight up to a constant multiplication on the exponent. So, we have

$$\mathbf{Pr}[G(n,p) \text{ is triangle-free }] = \begin{cases} e^{-(1+o(1))n^3p^3/6} & \text{if } p = o(n^{-1/2}), \\ e^{-\Theta(n^2p)} & \text{if } p = \Omega(n^{-1/2}). \end{cases} \tag{8.4}$$

Similar result can be proved for graphs other than triangles. Recall that a graph H is strictly balanced if $\frac{e(H')}{|H'|} < \frac{e(H)}{|H|}$ holds for all subgraphs H' of H. We can prove the theorem for larger range of p with the same proof technique, but this range depends on the values f_j in the proof. Hence we just state the theorem as follows for specific p.

Theorem 8.5. *Let H be a strictly balanced graph with v vertices e edges and a automorphisms. Let $c > 0$ be fixed. For $p = cn^{-v/e}$, we have*

$$\mathbf{Pr}[G(n,p) \text{ contains no copy of } H \text{ }] = \exp(\frac{-c^e}{a} + o(1)).$$

Proof. For each $\alpha \in [\frac{1}{a}\binom{n}{v}v!]$, let A_α range over all edge sets of possible copies of H. Let B_α be the event that $G(n,p)$ contains edges of A_α. Then

$$\mu = \binom{n}{v}v!p^e/a = (1+o(1))\frac{c^e}{a} \text{ and } M = (1-p^e)^{\frac{1}{a}\binom{n}{v}v!} = e^{-\frac{c^e}{a}+o(1)}.$$

Again, we want to estimate

$$\Delta = \sum_{\alpha \sim \beta} \mathbf{Pr}[B_\alpha \wedge B_\beta].$$

If $\alpha \sim \beta$, then A_α and A_β has at least two common vertices. For each $j \geq 2$, let f_j be the maximum number of edges in the intersection of $A_\alpha \cap A_\beta$ where $\alpha \sim \beta$ and A_α and A_β intersect at j vertices. As $\alpha \neq \beta$, we have $f_j < e$. If $2 \leq j \leq v$ and $\alpha \neq \beta$ then $A_\alpha \cap A_\beta$ is a proper subgraph of H, hence strict balancedness of H implies

$$\frac{f_j}{j} < \frac{e}{v}.$$

There are $O(n^{2v-j})$ choices of α, β intersecting at j vertices. Also for such α, β intersecting at j vertices, we have

$$\mathbf{Pr}[B_\alpha \wedge B_\beta] = p^{|A_\alpha \cup A_\beta|} = p^{2e - |A_\alpha \cap A_\beta|} \leq p^{2e - f_j}.$$

Thus, we have

$$\Delta = \sum_{j=2}^{v} O(n^{2v-j}) O(n^{-\frac{v}{e}(2e-f_j)}) \leq \sum_{j=2}^{v} O(n^{\frac{vf_j}{e} - j}) \leq \sum_{j=2}^{v} o(1) = o(1).$$

Hence, Janson's inequality implies that

$$M \leq \mathbf{Pr}[\bigwedge \overline{B_\alpha}] \leq e^{-\frac{c^e}{a} + o(1)}.$$

This proves the theorem. □

8.2. **Lower tails.** In the same set-up as above, recall that $X = \sum_{i \in I} \mathbb{1}_{B_i}$ is the random variable counting the number of events B_i happened. In the example where we considered triangles in $G(n,p)$, X counts the number of triangles in the random graph. Previously, we estimated the probability that $X = 0$. How about the probability that X is smaller than $\frac{1}{2}\mathbb{E}[X]$? The following theorem provides a tool to bound such probability.

Theorem 8.6. *Assume as in Theorem 8.2. For any $0 \leq t \leq \mu$, we have*

$$\mathbf{Pr}[X \leq \mu - t] \leq e^{-\frac{t^2}{2(\mu + \Delta)}}.$$

With this, we can bound the desired probability as follows, where X is the number of triangles in $G(n,p)$ and $C = C(c)$ is a constant depending on c.

$$\mathbf{Pr}[X \leq (1-c)\mu] \leq \exp[-C \frac{n^6 p^6}{n^3 p^3 + n^4 p^5}].$$

Again these inequalities are tight up to constant multiple on the exponents by (8.4), as triangle-free case is included in the event of $X \leq (1-c)\mu$.

Proof of Theorem 8.6. Let $q \in [0,1]$ be determined later. For each $\alpha \in I$, let J_α be independent random variables which is 1 with probability q and 0 with probability $1-q$. Add each $\alpha \in I$ to a set T independently at random with probability q. Let

$$X_T = \sum_{\alpha \in T} \mathbb{1}_{B_\alpha} = \sum \mathbb{1}_{B_\alpha} J_\alpha.$$

Then we have $\mathbf{Pr}[X_T = 0 \mid X] = (1-q)^X$. Take expectation on both sides, and apply Janson's inequality, then we obtain

$$\mathbb{E}[(1-q)^X] = \mathbf{Pr}[X_T = 0] \le \sum_{S \subseteq I} \left(\mathbf{Pr}[T = S] \cdot e^{-\mu_S + \Delta_S/2} \right)$$

$$\le \exp\left[\sum_S (\mathbf{Pr}[T = S] e^{-\mu_S + \Delta_S/2}) \right] \le e^{-\mu' + \Delta'/2}$$

where $\mu_S = \mathbb{E}[X_S]$ and $\Delta_S = \sum_{\alpha \sim \beta \in S} \mathbf{Pr}[B_\alpha \wedge B_\beta]$ and $\mu' = \mathbb{E}[X_T] = q\mu$ and $\Delta' = q^2\Delta$. The penultimate inequality is by the convexity of the function $\exp(x)$.

Using Markov's inequality, we have

$$\mathbf{Pr}[X \le \mu - t] = \mathbf{Pr}[(1-q)^X \ge (1-q)^{\mu - t}] \le (1-q)^{-\mu+t} \mathbb{E}[(1-q)^X] \le (1-q)^{-\mu+t} e^{-q\mu + q^2\Delta/2}.$$

Let $1 - q = e^{-\lambda}$ with $\lambda = \frac{1}{\mu + \Delta}$, then as $\lambda \ge q \ge \lambda - \lambda^2/2$, we have

$$\mathbf{Pr}[X \le \mu - t] \le \exp[\lambda(\mu - t) - (\lambda - \frac{\lambda^2}{2})\mu + \frac{\lambda^2 \Delta}{2}] = \exp[-\lambda t + \frac{\lambda^2}{2}(\mu + \Delta)] \le \exp[-\frac{t^2}{2(\mu + \Delta)}].$$

□

Note that this proof only works for lower tails. A similar statement does not hold for upper tails. Again consider the random variable X counting triangles in $G(n,p)$ and $p = \omega(n^{-1/2})$. Once we have a clique on $2np$ vertices, then we have $X \ge \binom{2np}{3} > 2\mathbb{E}[X]$. Then the probability that $G(n,p)$ has a clique on a specific $2np$ vertices is at least $p^{\binom{2np}{2}} \ge e^{-cn^2p^2 \log(1/p)} = \omega(e^{-Cn^2p})$. Hence, one cannot expect an upper tail inequality of the form $\mathbf{Pr}[X \ge (1+c)\mathbb{E}[X]] \le e^{-Cn^2p}$. It is known that if $p = \Omega(\frac{\log n}{n})$, then we have $\mathbf{Pr}[X \ge 2\mathbb{E}[X]] = e^{-\Theta(n^2p^2 \log(1/p))}$.

8.3. **Large deviations and disjoint families.** Recall our aims. We have bad events $B_1, \dots, B_k$ which are determined by $A_1, \dots, A_k$ where each A_i is a set of independent events. We have $B_i \sim B_j$ if and only if $A_i \cap A_j \neq \emptyset$. We wish to obtain a concentration result on $X = \sum \mathbb{1}_{B_i}$. In other words, we want to count the number of events $B_1, \dots, B_k$ happening. In many cases, especially if $\Delta = o(\mu)$, the number of (α, β) where $\alpha \sim \beta$ and B_α, B_β both occurred is likely to be much smaller than X. Hence, if we count the largest number of events $B_{i_1}, \dots, B_{i_s}$ where $i_j \not\sim i_\ell$, it is likely to be close to X. So, counting this may help us to obtain information about X. This connection from the size of independent events to X are always done on ad hoc basis for each applications. We make this 'collection of independent events' more rigorous and prove results about this. Later we see how we make connection from this to X.

Definition 8.7. *Given a selection $R \subseteq \Omega$, we call a set $J \subseteq I$ a disjoint family (disfam) if*

- *$A_j \subseteq R$ for every $j \in J$*
- *$A_i \cap A_j = \emptyset$.*

If the following holds in addition, then we call J a maximal disjoint family (maxdisfam).

- *if $j' \notin J$ and $A_{j'} \subseteq R$, then $A_{j'} \cap A_j \neq \emptyset$ for some $j \in J$.*

Recall that we write $\mu = \mathbb{E}[X]$.

Lemma 8.8. *With the above set-up and an integer s, we have*

$$\mathbf{Pr}[\exists \text{ a disfam } J \text{ with } |J| = s] \leq \frac{\mu^s}{s!}.$$

Proof. Such a probability is at most

$$\sum_{\substack{J=\{j_1,\dots,j_s\} \\ j_i \not\sim j_{i'}, i \neq i' \in [s]}} \mathbf{Pr}[\bigwedge_{j\in J} B_j] = \sum_J \prod_{j\in J} \mathbf{Pr}[B_j]$$

$$= \frac{1}{s!} \sum_{\substack{(j_1,\dots,j_s), \\ j_i \not\sim j_{i'}, i \neq i' \in [s]}} \mathbf{Pr}[B_{j_1}] \dots \mathbf{Pr}[B_{j_s}] \leq \frac{1}{s!} (\sum_{i\in I} \mathbf{Pr}[B_i])^s = \frac{\mu^s}{s!}.$$

□

This provides a good bound if $\mu^s = o(s!)$, basically if $s > \mu\alpha$ for $\alpha > e$. If s is smaller, then we can instead consider the maxdisfam. Let

$$\mu_s = \min_{j_1,\dots,j_s \in I} \sum_{i \not\sim j_\ell \text{ for all } \ell \in [s]} \mathbf{Pr}[B_i] \text{ and } \nu = \max_{j\in I} \sum_{i\sim j} \mathbf{Pr}[B_i].$$

Note that we have

$$\mu_s \geq \mu - s\nu.$$

Lemma 8.9.

$$\mathbf{Pr}[\exists \text{ a maxdisfam } J \text{ with } |J| = s] \leq \frac{\mu^s}{s!} e^{-\mu_s} e^{\Delta/2} \leq \frac{\mu^s}{s!} e^{-\mu+s\nu} e^{\Delta/2}.$$

Proof. We write $\overset{*}{\sum}$, $\overset{J}{\sum}$ and $\overset{J}{\bigwedge}$ to denote $\sum_{J \text{ disfam of } R=\Omega, |J|=s}$, $\sum_{i: i\not\sim j \forall j\in J}$ and $\bigwedge_{i: i\not\sim j \forall j\in J}$, respectively. Fix such a set J with size s, and let

$$\mu_J = \mathbb{E}[\overset{J}{\sum} \mathbf{Pr}[B_i]] \geq \mu_s$$

and let

$$\Delta_J = \sum_{\substack{i\sim\ell \\ i\not\sim j, \ell\not\sim j, \forall j\in J}} \mathbf{Pr}[B_i \wedge B_\ell] \leq \Delta = \sum_{i\sim\ell} \mathbf{Pr}[B_i \wedge B_\ell].$$

As $\mu_J \geq \mu_s$, using Janson's inequality, we have

$$\mathbf{Pr}[\overset{J}{\bigwedge} \overline{B_i}] \leq e^{-\mu_s + \Delta/2}.$$

Thus we have

$$\overset{*}{\sum} \mathbf{Pr}[J \text{ maxdisfam}] \leq e^{-\mu_s+\Delta/2} \overset{*}{\sum} \mathbf{Pr}[\bigwedge_{j\in J} B_j] \leq e^{-\mu_s+\Delta/2} \frac{\mu^s}{s!}.$$

The last inequality comes from the proof of the previous lemma. □

Note that if $\Delta = o(1)$ and $\mu_s = \mu + o(1)$, then the above lemma gives that the distribution of existence of maxdisfam is approximately same with Poisson distribution. This holds when, say, $s \leq 3\mu$ and $\nu\mu = o(1)$. Also, if s is larger, then Lemma 8.8 provides a very small upper bound.

Now we see how one can use this to some specific problems. In $G \sim G(n,p)$ and $x \in [n]$, let X denote the number of triangles containing x. Then $\mu = \mathbb{E}[X] = \binom{n-1}{2}p^3$. Basically, we want to count the edges in the neighborhood of x. One way to do this is using the maximum matching size in $E(G[N_G(x)])$ rather than the number of edges. The following theorem can be proved for all p ensuring $\mu = \omega(\ln n)$, but we only present the proof when $p = n^{-2/3+o(1)}$.

Theorem 8.10. *Let $p = n^{-2/3+o(1)}$ and let $G \sim G(n,p)$ and $x \in V(G)$. Let X be the number of triangles containing x. For a fixed $\varepsilon' > 0$, we have*

$$\mathbf{Pr}[X = (1 \pm \varepsilon')\mu] \geq 1 - o(n^{-1}).$$

Proof. Let $\varepsilon < \varepsilon'$ so that $\varepsilon\mu < \varepsilon'\mu - 50$. We use the notations used in the previous two lemmas. Let P be the Poisson random variable with the expectation μ. Note that we have $\mu = n^{o(1)}$ and if $s \leq 3\mu$, then we have $\nu = \max_{j\in I}\sum_{i\sim j}\mathbf{Pr}[B_i] = 2(n-3)p^3$, so $\nu\mu = o(1)$. Also $\Delta = \sum_{i\sim j}\mathbf{Pr}[B_i \wedge B_j] = 6\binom{n-1}{4}p^9 = o(1)$. Hence, the previous two lemmas yield

$$\mathbf{Pr}[\exists\text{maxdisfam } J, |J| \leq (1-\varepsilon)\mu] \leq (1+o(1))\mathbf{Pr}[P \leq (1-\varepsilon)\mu]$$
$$\mathbf{Pr}[\exists\text{maxdisfam } J, (1+\varepsilon)\mu \leq |J| \leq 3\mu] \leq (1+o(1))\mathbf{Pr}[(1+\varepsilon)\mu \leq P \leq 3\mu]$$
$$\mathbf{Pr}[\exists\text{disfam } J, |J| \geq 3\mu] \leq \sum_{s=3\mu}^{\infty}\frac{\mu^s}{s!} = O((1-c)^{\mu}).$$

Here c is some absolute constant. As Poisson distribution is a limit of binomial distribution, we can apply Chernoff to approximate the Poisson distribution. Then we can conclude that the first two probabilitys are $o(n^{-1})$. Hence, we can conclude that with probability at least $1 - o(n^{-1})$, every maxdisfam J has size between $(1-\varepsilon)\mu$ and $(1+\varepsilon)\mu$.

To obtain information about X from the above, we consider the probability that a vertex y has degree at least 5 in $L = G[N_G(x)]$, i.e. there exist five triangles $xyz_1, \ldots, xyz_5$ in G. This probability is at most $O(n^6p^{11}) = o(n^{-1})$. Also, the probability that L contains four disjoint $K_{1,2}$, in other words, the probability that there exists triangles xy_iz_i, xy_iz_i' for $i \in [4]$ with $x, y_1, z_1, \ldots, y_4, z_4$ distinct is at most $O(n^{12}p^{20}) = o(n^{-1})$.

Hence, with probability at least $1 - o(n^{-1})$, the graph L has maximal matching J of size between $(1-\varepsilon)\mu$ and $(1+\varepsilon)\mu$ and has maximum degree at most 4 and has no $4K_{1,2}$ as a subgraph. Let $J = \{y_1z_1, \ldots, y_sz_s\}$.

If L has 50 edges not in J, then those edges are incident to one of $\{y_1, \ldots, z_s\}$. If there are 7 edges outside J intersecting with $\{y_i, z_i\}$ for some i, then this contradicts the assumption that L has maximum degree at most 4. So there are at most 6 edges intersecting with each $\{y_i, z_i\}$. As there are more than 48 edges of L not in J, there are at least eight indices $i_1, \ldots, i_8$ such that those edges intersect with $\{y_{i_j}, z_{i_j}\}$ for each $j \in [8]$. From this, we obtain $4K_{1,2}$, a contradiction. This proves that L has at most $|J|+50$ edges. Hence, we have

$$(1-\varepsilon)\mu \leq X \leq (1+\varepsilon)\mu + 50$$

with probability at least $1 - o(n^{-1})$. □

8.4. **Counting representations.** We will use the following lemma.

Lemma 8.11 (The Borel-Cantelli Lemma)**.** *Let $B_1, \dots$ be events such that $\sum_{n=1}^{\infty} \mathbf{Pr}[B_n] < \infty$. Then, with probability 1, there exists $n_0 \in \mathbb{N}$ such that B_n is false for all $n > n_0$.*

One can consider representations of an integer as a sum of other specific integers. As an easy example, if S is a collection of powers of two, then every integer n can be expressed as a sum of distinct elements in S in a unique way. It is known that every integers are sum of four squares, and Goldbach conjectured that every even number can be written as a sum of two primes.

We consider a case related to this theme. Let $S \subseteq \mathbb{N}$ be a set of natural number, let $f(n) = f_S(n)$ be the number of representations $n = x + y$ with $x, y \in S$, $x < y$. We are interested in a set S where $f(n)$ is somewhat uniform.

Theorem 8.12 (Erdős, 1956)**.** *There is a set S and constants $c_1, c_2 > 0$ so that for all sufficiently large n, we have*

$$c_1 \ln n \leq f(n) \leq c_2 \ln n.$$

Proof. For each $x \in \mathbb{N}$, we add x to S with probability $p_x = \min[10\sqrt{\frac{\ln x}{x}}, 1]$ independently. We fix n, then $f(n)$ is a random variable with $\mu = \mathbb{E}[f(n)] = \frac{1}{2}\sum_{x+y=n, x\neq y} p_x p_y$. As long as x, y are not too small, we have $p_x p_y = \Theta(\frac{\ln n}{n})$. Some calculus yield that

$$\mu = (50 + o(1)) \ln n \int_0^1 \frac{dx}{\sqrt{x(1-x)}} = (50\pi + o(1)) \ln n.$$

For fixed n, all pairs $\{x, y\}$ with $x + y = n$ are disjoint, so we use the Chernoff bound to conclude that

$$\mathbf{Pr}[|f(n) - \mu| > 0.9\mu] < 2e^{-0.1\mu} \leq n^{-1.1}$$

for sufficiently large n. Let $c_1 = 4\pi, c_2 = 100\pi$. Let B_n be the event that $c_1 \ln n > f(n)$ or $f(n) > c_2 \ln n$ holds. As $\sum_{n\in\mathbb{N}} n^{-1.1} < \infty$, by the Borel-Cantelli Lemma, with probability 1, there exists $n_0 \in \mathbb{N}$ such that B_n fails for all $n \geq n_0$. Hence, $c_1 \ln n \leq f(n) \leq c_2 \ln n$ for all $n \geq n_0$. □

One can ask whether one can find such a set S in a more efficient way. Kolountzakis 1999 proved that there exists a recursive set S with the above property.

We can also consider similar question with k-sum. For example, let $g(n) = g_S(n)$ be the number of representation $n = x + y + z$ with $x, y, z \in S, x < y < z$. Similarly, we can define a function which is the number of representation of n into a sum of k elements of S. The following theorem also holds for $k > 3$ but we only present a proof of the case $k = 3$ using Poisson paradigm.

Theorem 8.13 (Erdős-Tetali, 1990)**.** *There exists a set S and constant c_1, c_2 such that for all sufficiently large n, we have*

$$c_1 \ln n \leq g(n) \leq c_2 \ln n.$$

Proof. For each $x \in \mathbb{N}$, we add x to S independently with probability $p_x = \min[M(\frac{\ln x}{x^2})^{1/3}, 1/2]$ for some large number M. Fix $n \in \mathbb{N}$. Then $g(n)$ is a random variable with $\mu = \mathbb{E}[g(n)] = \sum_{x+y+z=n, x<y<z} p_x p_y p_z$. Then, calculus yields that

$$\mu \sim \frac{M^3}{6} \ln n \int_{x=0}^{1} \int_{y=0}^{1-x} \frac{dxdy}{(xy(1-x-y)^{2/3})} = K \ln n$$

for some number K. As M is our choice of large number, we can assume K is also large.

We will use Lemma 8.9. Here, we have $\Delta = \sum p_x p_y p_z p_{y'} p_{z'}$ where the sum is over all tuples with $x + y + z = x + y' + z' = n$. Roughly there are n^3 terms and each term is roughly $n^{-10/3+o(1)}$, so that the sum is $o(1)$. (Readers can check that the terms coming from tuples containing small values do not contribute much.)

Fix $s \leq 3\mu = \Theta(\ln n)$. Then μ_s is the minimum possible $\sum p_x p_y p_z$ where the sum is over all x, y, z with $x + y + z = n$ that does not intersect a given set of s representations of n. As $s = \Theta(\ln n)$ is small, we can check that $\mu_s \sim \mu$.

Let P be the Poisson distribution with mean μ. Again, we can conclude that for some $\varepsilon > 0$,

$$\mathbf{Pr}[\exists \text{maxdisfam } J, |J| \leq (1-\varepsilon)\mu] \leq (1+o(1))\mathbf{Pr}[P \leq (1-\varepsilon)\mu]$$

$$\mathbf{Pr}[\exists \text{maxdisfam } J, (1+\varepsilon)\mu \leq |J| \leq 3\mu] \leq (1+o(1))\mathbf{Pr}[(1+\varepsilon)\mu \leq P \leq 3\mu]$$

$$\mathbf{Pr}[\exists \text{disfam } J, |J| \geq 3\mu] \leq \sum_{s=3\mu}^{\infty} \frac{\mu^s}{s!} = o(n^{-c})$$

with $c > 1$. Here the third one can be obtained since K is large. By Borel-Cantelli Lemma, with probability 1, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, there is a maxdisfam J of size between $c_1' \ln n$ and $c_2' \ln n$ for some $c_1', c_2' > 0$.

Now we use this J to estimate $g(n)$. Let $f(n)$ be the number of representation of n as the sum of two elements of S. Then

$$\mathbb{E}[f(n)] = \frac{1}{2} \sum_{x+y=n} (xy)^{-2/3+o(1)} = n^{-1/3+o(1)}.$$

As the possible representations $\{x, y\}$ with $x + y = n$ are pairwise disjoint, by Lemma 8.8 we have

$$\mathbf{Pr}[f(n) \geq 4] \leq \frac{\mathbb{E}[f(n)]^4}{4!} = n^{-4/3+o(1)}.$$

By the Borel-Cantelli Lemma, with probability 1, we have $f(n) \leq 3$ for all sufficiently large n. So, with probability 1, there exists C such that $f(n) \leq C$ for all n.

So, there exists a set S where $f(n) \leq C$ for all n and for all sufficiently large n, there is a maxdisfam J of size at most $c_2' \ln n$. As J is maxdisfam, for every triple $x, y, z \in S$ with $x + y + z = n$ must contain at least one of these at most $3c_2' \ln n$ points consisting J. The number of triples $x, y, z \in S$ with $x + y + z = n$ for a particular x is $f(n - x)$ which is the number of representation $n - x = y + z$ is at most C. So, there are at most $3Cc_2' \ln n$ total representations of $n = x + y + z$. □

9. Codes, games and entropy

9.1. **Codes.** Imagine we send a message (a string of bits) to someone far away. As there are some noises in the channel, there is a probability p that any bit sent will be received incorrectly. We assume that the probability that a zero is received as a one or a one is received as a zero are both p, and each bits being received incorrectly are mutually independent events.

How can we improve the reliability of this system? If we repeat each bits $2n+1$ times, and 'decode' it by choosing the bits occurring the most, then we can reduce the probability of incorrect receipts. However, this makes the string longer, so it becomes less efficient.

Can we find a way so that the efficiency is not so bad while the probability of incorrect transmissions approaching to zero? Shannon's theorem states that this is possible.

Definition 9.1. *A coding scheme consists of a positive integer m, n and function $f : \{0,1\}^m \to \{0,1\}^n$ called the encoding function, and a function $g : \{0,1\}^n \to \{0,1\}^m$ decoding function.*

For given a coding scheme, a message $x \in \{0,1\}^m$ will be encoded and sent as $f(x)$ and received message $y \in \{0,1\}^n$ will be decoded as $g(y) \in \{0,1\}^m$. The rate of transmission of such a scheme is defined to be m/n. Let $E = (e_1, \ldots, e_n)$ be a random string defined by $\mathbf{Pr}[e_i = 1] = p$, $\mathbf{Pr}[e_i = 0] = 1 - p$ where the value of each e_i are mutually independent.

Definition 9.2. *We defined the probability of correct transmission as $\mathbf{Pr}[g(f(x)+E) = x]$, where x is assumed to be uniformly distributed over $\{0,1\}^m$ and independent to E, and $+$ is vector addition modulo 2.*

Let $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$ be the entropy function defined for $p \in (0,1)$. We know that

$$\binom{n}{pn} = 2^{n(H(p)+o(1))} \text{ and } \sum_{i \le pn} \binom{n}{i} = 2^{n(H(p)+o(1))}.$$

Note that the following theorem is best possible. If a coding scheme has rate of transmission larger than $1-H(p)+\varepsilon$, then it must have a significant error probability. If $f(x)$ is sent, then the obtained output y is typically of distance $(1+o(1))np$ from $f(x)$. Hence, among 2^m input words, the total size of all typical outputs is about $2^m\binom{n}{pn} = 2^{m+(1+o(1))H(p)n}$. If this is much bigger than 2^n, then there are significant overlaps between the output sets of different input words, yielding a significant error probability.

Theorem 9.3 (Shannon's theorem)**.** *Let $0 < p < 0.5$ be fixed. For $\varepsilon > 0$, there exists a coding scheme with the rate of transmission greater than $1 - H(p) - \varepsilon$ and probability of incorrect transmission less than ε.*

Before proving this theorem, consider the following concepts.

Definition 9.4. *A group code is a coding scheme in which the map $f : \{0,1\}^m \to \{0,1\}^n$ is linear, that is, $f(0) = 0$ and $f(x + x') = f(x) + f(x')$ modulo 2.*

A group code is easier coding than general coding. Once we have x, we need to 'compute' $f(x)$. But if we have a group coding, this is easier. The following theorem is stronger than Shannon's theorem, so we prove this theorem instead.

Theorem 9.5. *Let $0 < p < 0.5$ be fixed. For $\varepsilon > 0$, there exists a group code with the rate of transmission greater than $1 - H(p) - \varepsilon$ and probability of incorrect transmission less than ε.*

Proof. Additions, subtractions here are all under modulo 2. Let $\delta > 0$ be small so that $p + \delta < 0.5$ and $H(p + \delta) < H(p) + \varepsilon/2$. Let n be large and $m = n(1 - H(p) + \varepsilon)$.

Let $u_i \in \{0,1\}^m$ be the vector with a 1 in the i-th position and rest of the position zeros. Let $f(u_1), \ldots, f(u_m)$ be chosen independently uniformly at random from $\{0,1\}^n$, and for each $x \in \{0,1\}^m$ we let

$$f(x) = \sum_{i \in [m]} x_i f(u_i).$$

Define the decoding function $g : \{0,1\}^n \to \{0,1\}^m$ by setting $g(y) = x$ if x is the unique vector in $\{0,1\}^m$ whose image $f(x)$ has Hamming distance at most $n(p + \delta)$ from y. If there is no such x or more than one such x, then we consider the decoding to be incorrect.

There are two ways in which the decoding to be incorrect.

(1) $f(x) + E$ has Hamming distance more than $n(p + \delta)$ from $f(x)$
(2) There is some $x' \neq x$ with $f(x) + E$ has Hamming distance at most $n(p + \delta)$ from $f(x')$.

The first happens when E has at least $n(p + \delta)$ digits of 1. However, the number of 1s in E has the binomial distribution $B(n, p)$, so Chernoff implies that this happens with probability $o(1)$. For the second case, let $x \neq x' \in \{0,1\}^m$ and $z = x - x' = \sum z_i u_i \neq 0$. WLOG, assume $z_m = 1$. Fix E and let $B(E)$ be the set of vectors Hamming distance at most $n(p + \delta)$ from E. After we fix E and $f(u_1), \ldots, f(u_{m-1})$, the vector $f(u_m)$ is still uniform random vector, so $f(z)$ is also distributed uniformly after fixing E and $f(u_1), \ldots, f(u_{m-1})$. Thus $\mathbf{Pr}[f(z) \in B(E)] = |B(E)|2^{-n}$. As $f(z) = f(x) - f(x')$, we have $f(z) \in B(E)$ if and only if $f(x) + E$ and $f(x')$ differs on at most $n(p+\delta)$ coordinates. So, this shows that $f(x) + E$ has Hamming distance at most $n(p + \delta)$ from $f(x')$ with probability $2^{(1+o(1))H(p+\delta)}2^{-n}$. Taking union bounds for all $x' \in \{0,1\}^m - \{x\}$, we have that the second events happens with probability at most

$$2^m 2^{(1+o(1))H(p+\delta)}2^{-n} < 2^{-n(\varepsilon/2+o(1))} = o(1).$$

Hence, the total probability for incorrect encoding is $o(1)$. So for sufficiently large n, this is at most ε. This probability can be rewritten as (note that there are $(2^n)^m$ choices of f)

$$2^{-mn} \sum_f \mathbf{Pr}_{x,E}[g(f(x) + E) \neq x] \leq \varepsilon.$$

This implies that there exists a function f such that $\mathbf{Pr}_{x,E}[g(f(x) + E) \neq x] \leq \varepsilon$, so there exists a specific coding scheme with probability of incorrect coding less than ε. □

9.2. **Liar game.** Consider that Paul and Carole plays the following Liar game.

In each round, Paul picks a subset $S \subseteq [n]$ and asks a following question to Carole: "is your x belong to S?" Carole must say either Yes or No. At the end of the game, Paul picks a number $y \in [n]$ and Carole has to pick a number $x \in [m] - \{y\}$, which will determine whether each of her previous answers was true or false.

In this game consisting of q rounds, Paul wins if Paul Carole always end up lying more than k times, otherwise Carole wins.

One can consider the version of game where x is determined by Carole at the beginning, and Paul is trying to guess what x is while Carole can lie at most k times. If Paul can always win in this version of the game, then Paul can win the Liar game. If Paul cannot always ensure winning, then Carole wins the Liar game.

We are interested in determining who wins for given n, q, k. Consider the following equivalent ChipLiar game.

Definition 9.6. *There is a board with positions $0, 1, \ldots, k$. There are n chips labeled $1, \ldots, n$ which are initially at position k. There are q rounds. On each round Paul selects a set S of the chips. Carole can either move every chip not in S one position below or move every chip in S one position below. Chips moved one position below from the position 0 will be removed from the borad. After q rounds, Carole wins if there is more than one chip remaining on the board and Paul wins if there is one or zero chips remaining on the board.*

Basically chip i at position j represents that the answer $x = i$ has already received $k - j$ lies.

In this ChipLiar game, there is no reason to place all chips at position k at the start. More generally, for $x_0, \ldots, x_k \geq 0$, we define the $(x_0, \ldots, x_k), q$-ChipLiar game to be the above q round game with initial position consisting of x_i chips at position i. Let $B(q, j) = 2^{-q} \sum_{i=0}^{j} \binom{q}{i}$ be the probability that in q flips of a fair coin there are at most j heads.

Theorem 9.7. *If $\sum_{i=0}^{k} x_i B(q, i) > 1$, then Carole wins the $(x_0, \ldots, x_k), q$-ChipLiar game.*

Proof. Note that this is a perfect information game with no draws, so someone has a perfect strategy that always wins. So it suffices to show that every strategy of Paul is not perfect.

Fix a strategy of Paul. Now Carole plays randomly. In each round, Paul has selected a set S of chips, and Carole flips a fair coin. If it comes up with head then she move every chip in S one position below and if it comes up with tail then she move every chip not in S one position below.

For each chip c, let I_c be the indicator random variable for c remaining on the board at the end of the game, and let $X = \sum_c I_c$ be the number of chips remaining on the board at the end of the game.

For each chip c, it is moved in each round with probability $1/2$. So, if it starts with the position i, then it remains on the board at the end of the game with the probability $B(q, i)$. By linearity of expectation, we have

$$\mathbb{E}[X] = \sum_{i=0}^{k} x_i B(q, i) > 1.$$

Then, $X > 1$ must occur with positive probability. That is, Paul cannot always win. This proves that no strategy of Paul is perfect, so Carole must have a winning strategy. $\square$

You might consider this proof not so efficient, as it does not tell you what is the strategy. However, we can apply derandomization technique here. For a specific board with y_i chips on i-th position for each i and ℓ rounds remaining, we can compute the weight $\sum_i y_i B(\ell, i)$ which is same as $\mathbb{E}[Y]$ where Y is the number of the chips that would remain on the board, if Carole plays the rest of the game at random.

For each round, Carole evaluate this weight and choose a move which maximizes the weight. Then this yields a deterministic algorithm which make sure Carole wins.

The converse of the theorem is not true. Consider the $(0, 5), 5$-ChipLiar game. Here $B(5, 1) = 6/32$ and $5 \times (6/32) < 1$. Still Carole has winning strategy. The problem is that Paul has no good first move. If he selects 2 chips as S, then Carole move the two chips, leaving the $(2, 3), 4$-ChipLiar game. As $2B(4, 0) + 3B(4, 1) = 17/16 > 1$, Carole will now win.

9.3. **Entropy.** Let's consider the following simple game. Alice choose a number x in $[2^n]$ and Bob asks a Yes/No question to determine the number x. In each question, Bob obtains some more information and after n steps Bob can determine the number x.

There are 2^n possibility, and each answer the Bob gets shrink the possibility into half. For convenience, we want to transform the multiplication (shrinking) into addition by considering the bits of information. The numbers in $[2^n]$ can be expressed by n bits, and every answer to questions of Bob gives one bits of information.

Now we change the set-up. We consider the uniform probability distribution (or equivalently a random variable $X \in [2^n]$ chosen uniformly at random) and random variables $Y_1, \dots, Y_t$ where each Y_i is the answer to Bob's question. Assume that $Y_1, \dots, Y_t$ carries enough information to determine X. (For example, one can consider i-th digit in the binary expansion of X with $t = n$.) Note that each Y_i corresponds to a question by Bob, and note that X and Y_i are not independent.

We want to consider more general situation where we want to measure how much information on probability distribution (or a random variable) we get from some events. In general, if we know that an event of probability p happens, then we can shrink the universe of the probability distribution by a factor of p. It's same as getting $-\log_2(p)$ bits of information. So for a discrete event x with the probability p, we define $h(x) = -\log_2(p)$.

For a given discrete random variable Y taking values in S, we can consider the expected amount of information we can get from this random variable, which is $H(X) = \mathbb{E}[h(X)] = \sum_{s \in S} p_s h(s) = \sum_{s \in S} -p_s \log_2 p_s$ where $p_s = \mathbf{Pr}[Y = s]$. So, in the above situation, if X and $(Y_1, \dots, Y_t)$ carries same information (i.e. $(Y_1, \dots, Y_t)$ are enough to determine X) then the expected information $Y_1, \dots, Y_n$ carries must be at least the expected information X carries. So, intuitively, we need $H(X) = H(Y_1, \dots, Y_t) \leq H(Y_1) + \cdots + H(Y_t)$, which requires $t \geq n$. We will later define what $H(Y_1, \dots, Y_t)$ means here and why the last inequality holds. Note that for a real $p \in (0, 1]$, the value $H(p) = -p \log_2 p + (1-p) \log_2 (1-p)$ is same as the entropy of random variable taking two values one with probability p and the other with $1 - p$.

We collect some basic inequalities. The following lemma says that uniform random variable has the largest entropy among all discrete random variables with the same range set.

Lemma 9.8. *Let X be a discrete random variable taking values in S, then $H(X) \leq \log_2 |S|$.*

Proof. Note that the function $f(x) = x \log_2 x$ is convex. So we have $\sum_{s \in S} -f(p_s) \leq -|S| f(\frac{1}{|S|}) = \log_2 |S|$. □

For two random variables X, Y not necessarily independent, let $Z = (X, Y)$ be the joint random variable. Let $H(X, Y) = H(Z)$. The below lemma shows that information we get from X and Y together is at least as large as the information on (X, Y).

Lemma 9.9 (Subadditivity)**.** *For given two random variables X, Y taking values in S, T, we have $H(X) \leq H(X, Y) \leq H(X) + H(Y)$.*

Proof. For $x \in S, y \in T$, we write

$$p_x = \mathbf{Pr}[X = x], p_y = \mathbf{Pr}[Y = y] \text{ and } p_{x,y} = \mathbf{Pr}[X = x, Y = y].$$

Then we have

$$H(X, Y) = -\sum_{x,y} p_{x,y} \log_2 p_{x,y} \geq -\sum_{x,y} p_{x,y} \log_2 p_x = -\sum_x p_x \log_2 p_x = H(X).$$

Also, we have

$$H(X)+H(Y)-H(X, Y) = \sum_{x,y}(-p_{x,y} \log_2 p_x - p_{x,y} \log_2 p_y + p_{x,y} \log_2 p_{x,y}) = \sum_{x,y} p_{x,y} \log_2 \frac{p_{x,y}}{p_x p_y}.$$

Let $f(z) = z \log_2 z$ and $z_{x,y} = \frac{p_{x,y}}{p_x p_y}$, then the convexity of f yields that

$$\sum_{x,y} p_{x,y} \log_2 \frac{p_x p_y}{p_{x,y}} \geq f(\sum_{x,y} p_x p_y z_{x,y}) = f(\sum_{x,y} p_x p_y \cdot \frac{p_{x,y}}{p_x p_y}) = f(1) = 0.$$

□

When we use $Y_1, Y_2, \ldots, Y_s$ to get more information, often Y_i is not independent of $Y_1, \ldots, Y_{i-1}$ so that the information it gives might be already partially known. We want to estimate average additional information we obtain. For this purpose, we define conditional entropy as follows. For an event E, let $H(X \mid E) = \sum_x -\mathbf{Pr}[X = x \mid E] \log_2 \mathbf{Pr}[X = x \mid E]$ and let $H(X \mid Y) = \mathbb{E}_y[H(X \mid Y = y)]$. Intuitively, this additional information we get from X knowing Y must be same as $H(X, Y) - H(Y)$. This can be shown as follows.

Lemma 9.10 (Chain rule)**.** *For given two discrete random variables X, Y, $H(X \mid Y) = H(X, Y) - H(Y)$.*

Proof. As we have $\mathbf{Pr}[X = x \mid Y = y] = p_{x,y}/p_y$, we have

$$\begin{aligned} H(X \mid Y) &= \mathbb{E}_y[H(X \mid Y = y)] = \sum_y -p_y \sum_x \mathbf{Pr}[X = x \mid Y = y] \log_2 \mathbf{Pr}[X = x \mid Y = y] \\ &= \sum_{x,y} -p_{x,y} \log_2 p_{x,y} + \sum_{x,y} p_{x,y} \log_2 p_y \\ &= H(X, Y) + \sum_y p_y \log_2 p_y = H(X, Y) - H(Y). \end{aligned}$$

The penultimate equality holds as $\sum_x p_{x,y} = p_y$. □

Also, the additional information we get from X when we just know Y is bigger than the additional information we get from X when we know both Y and Z.

Lemma 9.11 (Dropping conditions)**.** *For given discrete random variables X, Y, Z, $H(X \mid Y, Z) \leq H(X \mid Y)$*

Proof. We have $H(X \mid Y) = H(X, Y) - H(Y) \leq H(X) + H(Y) - H(Y) \leq H(X)$. Hence $H(X \mid Y, Z) \leq H(X \mid Y)$. □

Let's consider the following simple problem. Assume there are n coins, assume that the set of coins is $[n]$. Each coins have weight 1, but there are some fake coins with weights $1 + \frac{1}{2n}$. You can grab a set S of coins and weigh them, which tells you how many coins in S are fake. After weighing k times, you want to determine exactly which coins are fake. For this, we want to find a lower bound on k.

Theorem 9.12. *Let $S_1, \ldots, S_k$ be subsets of $[n]$. For any two distinct subsets $A \neq B \subseteq [n]$, there exists $i \in [k]$ such that $|S_i \cap A| \neq |S_i \cap B|$. Then $k \geq (2 - o(1))\frac{n}{\log_2 n}$.*

Proof. Most naive way of approaching this problem considering a function $f(A) = (|S_1 \cap A|, \ldots, |S_k \cap A|)$ from $2^{[n]}$ to $([n] \cup \{0\})^k$. As this is injective, we have $(n+1)^k \geq 2^n$, implying that $k \geq (1 - o(1))\frac{n}{\log_2 n}$. Here, we have the right order of magnitude, but a wrong constant. This is because each time you measure $|S_i \cap A|$, it does not shrink the possibility into $1/n$-fraction in average.

One can also consider the following approach of ignoring unlikely events. Assume $k \leq n$, as otherwise we are done. For each S_i, Chernoff's bound (Lemma 5.8) implies that there are at most $\frac{1}{100n}2^n$ subsets of $[n]$ satisfying

$$|S_i \cap A| \neq \frac{|S_i|}{2} \pm 10 \log n \sqrt{n}.$$

Let

$$\mathcal{S} = \{A \subseteq [n] : \forall i, |S_i \cap A| = \frac{|S_i|}{2} \pm 10 \log n \sqrt{n}\}.$$

Then we have that $|\mathcal{S}| \geq 2^n - k\frac{1}{100n}2^n \geq 2^{n-1}$. Now, the set $f(\mathcal{S})$ lies in $I_1 \times I_2 \times \cdots \times I_k$ where each $I_i = [\frac{|S_i|}{2} - 10 \log n \sqrt{n}, \frac{|S_i|}{2} + 10 \log n \sqrt{n}]$ is an interval of at most $20 \log n \sqrt{n}$ integers. So, we have $(20 \log n \sqrt{n})^k \geq 2^{n-1}$ implying

$$k \geq \frac{n-1}{\frac{1}{2}\log_2 n + \log\log n + \log 20} \geq (2 - o(1))\frac{n}{\log_2 n}.$$

This proof uses that S_i 'usually' shrink the possibility by the fact of $\Theta(\frac{1}{\sqrt{n}\log n})$. To deal with this problem, we can instead use the entropy.

Assume we are choosing a set $A \subseteq [n]$ uniformly at random, meaning all 2^n sets are chosen equally likely. Then $|S_i \cap A|$ is a random variable, containing some information about the random variable A.

First, the binomial random variable $Bin(m, 1/2)$ has entropy

$$H(Bin(m, \frac{1}{2})) = -\sum_{s=0}^{m} \binom{m}{s} (\frac{1}{2})^m \log_2 (\binom{m}{s} (\frac{1}{2})^m) = (\frac{1}{2} + o(1)) \log_2 m.$$

The last equality can be checked again using the Chernoff bound. (The Chernoff bound can show that the terms become negligible when s is too far from $m/2$.)

Now, as $(|S_1 \cap A|, \dots, |S_k \cap A|)$ determines A, by subadditivity we have

$$H(A) = H(|S_1 \cap A|, \dots, |S_k \cap A|) \le H(|S_1 \cap A|) + \cdots + H(|S_k \cap A|)$$
$$\le \sum_{i=1}^{k} (\frac{1}{2} + o(1)) \log_2 |S_i| \le (\frac{1}{2} + o(1)) k \log_2 n.$$

Note that each of $|S_i \cap A|$ has the same distribution as the binomial distribution $Bin(|S_i|, 1/2)$. Thus we conclude $k \ge (2 - o(1)) \frac{n}{\log_2 n}$. □

Proposition 9.13 (Shear's lemma)**.** *Let $X_1, \dots, X_n$ be random variables and let $A_1, \dots, A_k \subseteq [n]$ be subsets where each $i \in [n]$ is in at least s sets among $A_1, \dots, A_k$. Let $X_{A_i} = (X_i : i \in A_i)$. Then we have*

$$sH(X_1, \dots, X_n) \le \sum_{i \in [k]} H(X_{A_i}).$$

Proof. We use induction on s. If $s = 1$, then use subadditivity to a partition $A'_1, \dots, A'_k$ of $[n]$ where $A'_i \subseteq A_i$ holds to obtain the desired conclusion.

Assume that the proposition holds for $s - 1$. If there is A_i with $A_i = [n]$, then the result follows from the induction hypothesis. If not, choose two sets A_i and A_j. By using dropping conditions, we have

$$H[X_{A_i \setminus A_j} \mid X_{A_i \cap A_j}, X_{A_j \setminus A_i}] \le H(X_{A_i \setminus A_j} \mid X_{A_i \cap A_j}).$$

Using chain rule, we have

$$H(X_{A_i \cup A_j}) - H(X_{A_j}) \le H(X_{A_i}) - H(X_{A_i \cap A_j}).$$

Hence, we have $H(X_{A_i \cup A_j}) + H(X_{A_i \cap A_j}) \le H(X_{A_i}) + H(X_{A_j})$. So, we can modify the collections of sets by replacing A_i, A_j with $A_i \cup A_j$ and $A_i \cap A_j$ until we have a set $[n]$ without increasing the entropy. Once we have a set $[n]$, then we delete it and apply the induction hypothesis. This proves the induction. □

Using this proposition, we can prove several results.

Corollary 9.14. *Let $\mathcal{F}$ be a family of subsets of $[n]$, and let p_i be the fraction of the sets in $\mathcal{F}$ which contains i. Then we have*

$$|\mathcal{F}| \le 2^{\sum_{i=1}^{n} H(p_i)}.$$

Proof. Choose $F \in \mathcal{F}$ uniformly at random, and let $X = (X_1, \dots, X_n)$ be the $0,1$-vector such that $X_i = 1$ if $i \in F$. Then we have

$$\log_2 |\mathcal{F}| = H(X) \le \sum H(X_i) \le \sum_{i \in [n]} H(p_i).$$

This proves the corollary. □

Corollary 9.15. *Let $\mathcal{F}$ be a family of vectors in $S_1 \times \cdots \times S_n$ for some finite sets $S_1, \dots, S_n$. Let $A_1, \dots, A_m$ be a collection of subsets of $[n]$, and each $i \in [n]$ belongs to at least s sets in the collection. For each $i \in [m]$, let $\mathcal{F}_i$ be the set of all projections of the elements of $\mathcal{F}$ on $\prod_{j \in A_i} S_j$. Then*

$$|\mathcal{F}|^s \le \prod_{i \in [m]} |\mathcal{F}_i|.$$

Proof. Choose $F = (F_1, \dots, F_n) \in \mathcal{F}$ uniformly at random. For $X_{A_i} = (F_j : j \in A_i)$, the previous proposition implies

$$s \log_2 |\mathcal{F}| = sH(F) \leq \sum H(X_{A_i}) \leq \sum \log_2 |\mathcal{F}_i|.$$

□

Note that the volume of any d-dimensional measurable set in $\mathbb{R}^n$ can be approximated by the volume of standard aligned boxes in a fine grid. Using this, the previous result can prove the following.

Corollary 9.16. *Let B be a measurable body in the n-dimensional Euclidean space, let $vol(B)$ be its n-dimensional volume, and let $vol(B_i)$ denote the $(n-1)$-dimensional volume of the projection of B on the hyperplane spanned by all coordinates beside the ith one. Then we have*

$$vol(B)^{n-1} \leq \prod_{i \in [n]} vol(B_i).$$

9.4. **Graph homomorphisms and entropy.**

Definition 9.17. *A map $\phi : V(G) \to V(H)$ is a graph homomorphism if we have $\phi(u)\phi(v) \in E(H)$ for all $uv \in E(G)$. We write $H(G, H)$ to denote the set of all graph homomorphisms from G to H.*

For example, consider H be the two vertex graph where two vertices v_0 and v_1 are adjacent and v_1 has a loop to itself. Then each $f \in H(G, H)$ corresponds to an independent set $f^{-1}(v_0)$ of G. Hence, we have $|H(G, H)| = i(G)$ where $i(G)$ is the number of independent sets in G. Also it is easy to see that $Hom(G, K_q)$ is a collection of proper q-colorings of G.

Theorem 9.18 (Kahn)**.** *Let G be an n vertex d-regular bipartite graph. Then we have $i(G) \leq [i(K_{d,d})]^{\frac{n}{2d}}$.*

The above theorem can be extended to the following.

Theorem 9.19. *Let G be an n-vertex d-regular bipartite graph, and let H be a graph with possible loops. Then $|Hom(G, H)| \leq |Hom(K_{d,d}, H)|^{\frac{n}{2d}}$.*

Proof. Let $V(G) = [n]$. We choose $\phi \in H(G, H)$ uniformly at random. For given ϕ, let $X = (x_1, \dots, x_n)$ where $x_i = \phi(i)$. Then, as ϕ is uniformly chosen, we have

$$H(X) = \log_2(|Hom(G, H)|).$$

Let $A \cup B$ be a bipartition of G with $|A| = |B|$ and for a vertex set C let $X_C = (x_i : i \in C)$. Then we have $H(X) = H(X_A) + H(X_B \mid X_A)$. As G is d-regular bipartite, each vertex in A has d neighbors in B. Hence, using Proposition 9.13, we have

$$H(X) \leq \frac{1}{d} \sum_{b \in B} H(X_{N(b)}) + \sum_{b \in B} H(X_b \mid X_A) \leq \frac{1}{d} \left(\sum_{b \in B} H(X_{N(b)}) + \sum_{b \in B} dH(X_b \mid X_{N(b)}) \right).$$

The last inequality is by dropping conditions.

Now we fix b and estimate $H(X_{N(b)}) + dH(X_b \mid X_{N(b)})$. Let $X_b^1, \dots, X_b^d$ be the independent random variables having the identical distribution to X_b. Then

$$\begin{aligned} H(X_{N(b)}) + dH(X_b \mid X_{N(b)}) &= H(X_{N(b)}) + \sum_{i\in[d]} H(X^i \mid X_{N(b)}) \\ &= H(X_{N(b)}) + H(X_b^1, \dots, X_b^d \mid X_{N(b)}) = H(X_{N(b)}, X_b^1, \dots, X_b^d). \end{aligned}$$

Here, the joint random variable $(X_{N(b)}, X_b^1, \dots, X_b^d)$ is a random homomorphisms in $Hom(K_{d,d}, H)$ in a certain distribution. To see this, $X_{N(b)}$ is identical to a certain distribution on choosing random homomorphism in left d vertices of $K_{d,d}$ and $X_b^1, \dots, X_b^d$ correspond to assigning vertices in the d different vertices on the right side. This entropy of possibly non-uniform distribution is upper bounded by the entropy of the uniform distribution over the independent sets of $K_{d,d}$ which is $\log_2 |Hom(K_{d,d}, H)|$. As the above discussion holds for all $b \in B$, we have $H(X) \le \frac{n}{2d} \log_2(|Hom(K_{d,d}, H)|)$, implying that

$$|Hom(G, H)| \le |Hom(K_{d,d}, H)|^{\frac{n}{2d}}.$$

□

This theorem yields Theorem 9.18

Definition 9.20. *Let $t(H, G) = \dfrac{|Hom(H, G)|}{|V(G)|^{|V(H)|}}$ be the homomorphism density of H in G. In other words, this is the probability that a map from $V(H)$ to $V(G)$ chosen uniformly at random yields a graph homomorpihsm.*

Regarding problems of counting homomorphisms, the following conjecture by Sidorenko is famous.

Definition 9.21. *If H is a bipartite graph, the for any G we have*

$$t(H, G) \ge t(K_2, G)^{e(H)}.$$

What this conjecture says is that among all graphs with a fixed edge density, $t(H, G)$ is minimized when G is a random graph. This conjecture is still wide open while some instances of H are known to satisfy the conjecture. For example, we know the following.

Theorem 9.22 (Sidorenko, 1993)**.** *If H is a tree, then for any G we have*

$$t(H, G) \ge t(K_2, G)^{e(H)}.$$

A proof by Conlon, Kim, Lee, Lee. Let T be a tree with a fixed ordering $r = x_0, x_1, \dots, x_t$ of all vertices of T where every x_i has an earlier neighbor. We call such an ordering appropriate. We consider the following T-branching random walk on an n-vertex graph G.

Algorithm 9.23.

Step 1. Choose $v_0 \in V(G)$ with the probability $\frac{d_G(v_0)}{2e(H)}$. (This is called the stationary distribution)

Step 2. Assume we have embedded $v_0, \dots, v_{i-1}$. Let x_j be the unique earlier neighbor of x_i, then choose v_i uniformly at random from $N_G(v_j)$.

Step 3. Repeat Step 2 until the end.

We want two aspects of this algorithm. Each edge $xy \in E(T)$ is mapped to an edge in a uniform way, and T-branching random walk distribution comes from $T - \ell$-branching random walk distribution by just adding one edge (uniformly). The rather ensures that we can multiply one '$t(K_2, G)$' term for every edge of T, yielding the desired bound $t(H, G) \geq t(K_2, G)^{e(H)}$.

Let ϕ be the resulting random homomorphism. This gives a probability distribution on $Hom(T, G)$. Note that for a specific homomorphism $\psi = (v_0, \dots, v_t)$, we have

$$\mathbf{Pr}[\phi = \psi] = \frac{d_G(v_0)}{2e(G)} \cdot (\frac{1}{d_G(v_0)})^{d_T(x_0)} \prod_{i \geq 1} (\frac{1}{d_G(v_i)})^{d_T(x_i)-1} = \frac{1}{2e(G)} \prod_{i=0}^{t} (\frac{1}{d_G(v_i)})^{d_T(x_i)-1}.$$

Note that this formula does not depend on the ordering $x_0, \dots, x_r$ of the $V(T)$. So, even if we change the ordering into another appropriate ordering, the probability distribution stays the same.

One good property of this is the following. For each $xy \in E(H)$ and an edge $uv \in E(G)$, the probability

$$\mathbf{Pr}[\phi(x) = u, \phi(y) = v] = \frac{1}{2e(G)}. \tag{9.1}$$

This is easy to see by considering x as the first vertex of the ordering and y as the second vertex of the ordering. (Remember that the ordering of T does not matter!)

Let T' be a subtree of T by deleting a leaf ℓ. Let ϕ' be a random homomorphism in $Hom(T', G)$ obtained from T'-branching random walk. Then we have

$$\mathbf{Pr}[\phi' = \psi'] = \sum_{\psi|_{V(T')} = \psi'} \mathbf{Pr}[\phi = \psi].$$

This can be easily check by considering an ordering of T where ℓ comes the last. Then obviously $\phi \mid_{V(T')}$ and ϕ' have the same distribution. Recall that each $v_0, \dots, v_t$ are all random variables.

We use induction to prove the following claim

Claim 10. *For a tree T with t edges, and $\phi_T \in Hom(T, G)$ obtained from T-branching random walk, we have $H(\phi_T) \geq t \log_2(2e(G)) - (t-1)\log_2(n)$.*

Proof. If $t = 1$, then it is trivial. Assume that the statement holds for $T' = T - \ell$ where ℓ is a leaf. We have

$$H(\phi_{T'}) \geq (t-1)\log_2(2e(G)) - (t-2)\log_2(n).$$

Let $x_0, x_1, \dots, x_t = \ell$ be an appropriate ordering where x_t is adjacent to x_j. Let $v_i = \phi(x_i)$, which is also a random variable. We have

$$\begin{aligned} H(\phi) &= H(\phi \mid_{V(T')}) + H(v_t \mid \phi \mid_{V(T')}) \\ &= H(\phi \mid_{V(T')}) + H(v_t \mid v_j) = H(\phi \mid_{V(T')}) + H(v_j, v_t) - H(v_j) \\ &\overset{(9.1)}{=} H(\phi \mid_{V(T')}) + \log_2(2e(H)) - H(v_j) \geq H(\phi \mid_{V(T')}) + \log_2(2e(H)) - H(n). \end{aligned}$$

Here, the second equality holds as the choice of v_t only depend on v_j. The last inequality holds as v_j has one of n values. By the induction hypothesis, we have

$$H(\phi_T) \geq t \log_2(2e(G)) - (t-1)\log_2(n).$$

This finishes the induction. □

Again, as we know $H(\phi) \leq \log_2 |Hom(T, G)|$, for a t-edge tree T, we have

$$|Hom(T, G)| \geq 2^{t \log_2(2e(G)) - (t-1)\log_2(n)} = t(K_2, G)^t n^{t+1}.$$

Hence, $t(T, G) \geq t(K_2, G)^t$. □

10. More on independent sets

In Theorem 9.18, we proved that $i(G) \le i(K_{d,d})^{\frac{n}{2d}}$ for bipartite d-regular graphs. Here, we will prove some theorems regarding average size of independent sets and the number of independent sets in more general class of graphs.

10.1. The hard-core model.

In order to better estimate the number of independent sets as well as the average size of independent sets, we consider the following concept.

Definition 10.1. *For given graph G, let $P_G(x) = \sum_{I \text{ an indep set}} x^{|I|}$ be the independence polynomial of G. This is also called the partition function.*

Definition 10.2. *The hard-core model with fugacity λ on G is a random independent set I drawn according to the distribution $\mathbf{Pr}[I] = \frac{\lambda^{|I|}}{P_G(\lambda)}$. In hard-core model, we let the occupancy fraction to be*

$$\alpha_G(\lambda) = \frac{1}{|G|}\mathbb{E}[|I|] = \frac{1}{|G|}\frac{\sum_I |I|\lambda^{|I|}}{P_G(\lambda)} = \frac{1}{|G|}\frac{\lambda P_G'(\lambda)}{P_G(\lambda)} = \frac{\lambda}{|G|}(\log P_G(\lambda))'.$$

Imagine a physical system with some atoms which cannot occupy spaces too closely. Some atoms in this system may leave the system, or some atoms outside the system can enter the system. By connecting nearby places in the system, one can obtain a graph. How atoms are placed corresponds to an independent set. Fugacity above indicates how easy an atom can enter to the system or leave the system. The hard-core model provides a probability distribution of the state of this system. This provides some motivation for the hard-core model.

Note that $\alpha_G(1)$ is the average size of independent set. This fact together with the equation $\alpha_G(\lambda) = \frac{1}{|G|}\frac{\lambda P_G'(\lambda)}{P_G(\lambda)}$ and some other useful property of this hard-core model allow us to yields some results regarding independent sets. We first prove the following result about average size of independent set.

Theorem 10.3. *If G is a triangle-free graph with the maximum degree at most d, then we have $\alpha_G(1) \ge (1+o(1))\frac{\log d}{d}$.*

Proof. For a positive real number z, let $W(z)$ be the unique real number satisfying $W(z)e^{W(z)} = z$. Consider an independent set I drawn from the distribution of the hard-core model with fugacity λ. We say a vertex v is occupied if $v \in I$ and uncovered if $N(v) \cap I = \emptyset$. Let

$$p_v = \mathbf{Pr}[v \in I] \text{ and } q_v = \mathbf{Pr}[N(v) \cap I = \emptyset].$$

Assume G has n vertices. Let U denote the set of uncovered vertices. Note that $v \in I$ can only happen when v is uncovered. Moreover, all indendent sets J in $G - N[v]$ corresponds to an independent set $J \cup \{v\}$ where $\lambda^{|J\cup\{v\}|} = \lambda\lambda^{|J|}$. Hence, we have $p_v = \frac{\lambda}{1+\lambda}q_v$. Let v' be a vertex of $V(G)$ chosen uniformly at random, and let $Y = |N(v') \cap U|$. Therefore, we

have

$$\begin{aligned}\alpha_G(\lambda) &= \frac{1}{n}\sum_{v\in V(G)} p_v = \frac{1}{n}\sum_{v\in V(G)} \frac{\lambda}{1+\lambda} q_v \\ &= \frac{\lambda}{(1+\lambda)n}\sum_{v\in V(G)}\sum_{j=0}^{d} \mathbf{Pr}[|N(v)\cap U| = j](1+\lambda)^{-j} \\ &= \frac{\lambda}{1+\lambda}\mathbb{E}[(1+\lambda)^{-Y}]. \end{aligned} \tag{10.1}$$

The penultimate equality holds because each of the j uncovered vertices belong to I with probability $(1+\lambda)^{-1}$ and those events are independent as G is triangle-free ($N_G(v)$ is an independent set).

Using $\alpha_G(\lambda) = \frac{1}{n}\sum_{v\in V(G)} \frac{\lambda}{1+\lambda} q_v$, we have

$$\mathbb{E}[Y] = \frac{1}{n}\sum_{v\in(G)}\sum_{u\in N(v)} q_u \le d\frac{1+\lambda}{\lambda}\alpha_G(\lambda). \tag{10.2}$$

Note that the last term in (10.1) is at least $\frac{\lambda}{1+\lambda}(1+\lambda)^{-\mathbb{E}[Y]}$ by the convexity of the function $x \mapsto (1+\lambda)^{-x}$. Hence, we have

$$\alpha_G(\lambda) \ge \frac{\lambda}{d(1+\lambda)} \min_{x\in\mathbb{R}} \left\{\max\{x, d(1+\lambda)^{-x}\}\right\}.$$

When x increases, $d(1+\lambda)^{-x}$ decrease. Thus the minimum occurs for x where $x = d(1+\lambda)^{-x}$ holds. In other words, it holds where $\log(1+\lambda)xe^{\log(1+\lambda)x} = \log(1+\lambda)d$. So, $log(1+\lambda)x = W(\log(1+\lambda)d)$. Hence, we have

$$\alpha_G(\lambda) \ge \frac{\lambda x}{d(1+\lambda)} \ge \frac{\lambda}{1+\lambda}\cdot\frac{W(d\log(1+\lambda))}{d\log(1+\lambda)}.$$

Now we plug is $\lambda = 1/\log d$ and use $W(z) \ge \log z - \log\log z$ for $z \ge e$. Then we have

$$\alpha_G(\lambda) \ge (1+o(1))\frac{\log d}{d}.$$

Now we show that $\alpha_G(\lambda)$ is monotonically increasing in terms of λ. Let's write $P_G(\lambda) = P$. Let I be an independent set drawn from the hard-core model with fugacity λ. Then we have

$$\frac{\lambda^2 P''}{P} + \frac{\lambda P'}{P} = \mathbb{E}[|I|(|I|-1) + |I|] = \mathbb{E}[|I|^2].$$

Hence, we have

$$\begin{aligned}\alpha_G'(\lambda)|G| &= \frac{d}{d\lambda}(\frac{\lambda P'}{P}) = \frac{P'}{P} + \frac{\lambda P P'' - \lambda(P')^2}{P^2} \\ &= \frac{P'}{P} + \frac{1}{\lambda}(\frac{\lambda^2 P''}{P} - (\frac{\lambda P'}{P})^2) = \frac{1}{\lambda}\left(\mathbb{E}[|I|^2] - \mathbb{E}[|I|]^2\right) \\ &= \frac{\mathrm{Var}[|I|]}{\lambda} \ge 0.\end{aligned}$$

This shows that $\alpha_G(1) \ge \alpha_G(\frac{1}{\log d})$. This concludes the theorem. □

As a corollary, we have the following bound on the ramsey number

Proposition 10.4. $R(3,k) \le (1+o(1))\frac{k^2}{\log k}$.

Proof. Assume that G is triangle-free n-vertex graph. We will show that is has an independent set of size k.

If it has a maximum degree at least k, we are done as its neighborhood forms an independent set. If it has maximum degree at most k, then the previous theorem yields that it has an independent set of size at least $(1+o(1))\frac{\log k}{k}n$. This is at least k if $n > (1+o(1))\frac{k^2}{\log k}$. Hence, we conclude that $R(3,k) \le (1+o(1))\frac{k^2}{\log k}$. □

Theorem 10.5 (Davies, Jenssen, Perkins, Roberts)**.** *For all d-regular graphs G and all $\lambda > 0$, we have*

$$\alpha_G(\lambda) \le \alpha_{K_{d,d}}(\lambda) = \frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^d - 1}.$$

Proof. Consider an independent set I drawn from the distribution of the hard-core model with fugacity λ. We say a vertex v is occupied if $v \in I$ and uncovered if $N(v) \cap I = \emptyset$. Let

$$p_v = \mathbf{Pr}[v \in I] \text{ and } q_v = \mathbf{Pr}[N(v) \cap I = \emptyset].$$

We first prove this statement for the case when G is a triangle-free n-vertex graph. Assume G is triangle-free. Let U denote the set of uncovered vertices.

As before, (10.1) holds and (10.2) holds with equality. Hence, we have

$$\mathbb{E}[Y] = d\mathbb{E}[(1+\lambda)^{-Y}].$$

Now we do optimization over all distribution of Y (i.e. over all graphs G) we let

$$\alpha^* = \frac{\lambda}{d(1+\lambda)} \cdot \sup_Y \{\mathbb{E}[Y] : \mathbb{E}[Y] = d\mathbb{E}[(1+\lambda)^{-Y}]\}$$

where the supremum is taken over all distributions of Y which takes values in $\{0, 1, \dots, d\}$.

This is same as solving the following linear programming where $x_k = \mathbf{Pr}[Y = k]$.

$$\text{Maximize : } \sum_{k=0}^{d} k x_k,$$
$$\text{subject to : } \sum_k x_k = 1, \sum_k x_k((1+\lambda)^{-k} - \frac{k}{d}) = 0, x_k \ge 0.$$

Note that $f(x) = (1+\lambda)^{-x} - \frac{x}{d}$ is a convex decreasing function where $f(0) = 1, f(d) < 0$. For each $0 < i < d$, the convexity of f yields that $f(i) \le \frac{d-i}{d}f(0) + \frac{i}{d}f(d)$. Hence, for a function $g(z) = zf(0) + (x_i - z)f(d)$, we have

$$g(\frac{(d-i)x_i}{d}) \ge x_i f(i) \ge x_i f(d) = g(0).$$

By the intermediate value theorem, there exists z such that $g(z) = x_i f(i)$ where $0 \le z \le \frac{(d-i)x_i}{d} \le x_i$. Note that $(d-i)x_i - zd \ge 0$.

Now we replace x_i by 0 and increase x_0 by z and x_d by $x_i - z$. With this new choice of $x_0, \dots, x_d$, we have $\sum x_k = 1$ and $\sum_k x_k((1+\lambda)^{-k} - \frac{k}{d}) = 0, x_k \ge 0$ while $\sum_{k=0}^d k x_k$ increase by $(d-i)x_i - zd \ge 0$. By repeating this, we may assume that a maximum occur when $x_i = 0$ for all $i \notin \{0, d\}$. If we further solve the remaining linear program with two variables, we obtain that $x_0 = \frac{(1+\lambda)^d - 1}{2(1+\lambda)^d - 1}$ and $x_d = \frac{(1+\lambda)^d}{2(1+\lambda)^d - 1}$. This is exactly the distribution comes from a disjoint union of $K_{d,d}$.

Now we consider general (not necessarily triangle-free) case. For given independent set I, we define the free neighborhood of a vertex v to be $G[\{u \in N(v) : N(u) \cap I = \emptyset\}]$.

We draw I accorrding to the hard-core model and choose a vertex v uniformly at random. Let C be the free neighborhood of v, which is also a random variable. For a graph F, let $p_F = \mathbf{Pr}[C \text{ isomorphic to } F]$. As before, we will compute $\alpha_G(\lambda)$.

First, we have

$$\begin{aligned}\alpha_G(\lambda) &= \frac{\lambda}{1+\lambda}\mathbb{E}_v q_v = \frac{\lambda}{1+\lambda}\sum_F \left(\mathbf{Pr}[C \simeq F] \cdot \mathbf{Pr}[v \text{ uncovered} \mid C \simeq F]\right) \\ &= \frac{\lambda}{1+\lambda}\sum_F \left(\mathbf{Pr}[C \simeq F] \cdot \frac{1}{P_C(\lambda)}\right) = \frac{\lambda}{1+\lambda}\mathbb{E}_C[\frac{1}{P_C(\lambda)}].\end{aligned}$$

Here, we have the penultimate equality since v is uncovered if and only if $I \cap V(C)$ is $\emptyset$ and this happens with probability $\frac{1}{P_C(\lambda)}$. We also have

$$\begin{aligned}\alpha_G(\lambda) &= \frac{1}{dn}\sum_v \sum_{u \in N(v)} p_u = \frac{1}{d}\mathbb{E}_v[|N(v) \cap I|] \\ &= \frac{1}{d}\sum_F \mathbf{Pr}[C \simeq F]\left(\sum_k k\mathbf{Pr}\left[|V(C) \cap I| = k \mid C \simeq F\right]\right) \\ &= \frac{1}{d}\sum_F \mathbf{Pr}[C \simeq F]\mathbb{E}_{I \cap V(C)}[|I \cap V(C)|] = \frac{1}{d}\sum_F \mathbf{Pr}[C \simeq F]\frac{\lambda P_F'(\lambda)}{P_F(\lambda)} = \frac{\lambda}{d}\mathbb{E}_C[\frac{P_C'(\lambda)}{P_C(\lambda)}].\end{aligned}$$

Here, the third equality and the penultimate equality hold since C is determined by $I \cap (V(G) - N[v])$, and once we fix $I \cap (V(G) - N[v])$, the distribution of $I \cap V(C)$ is same as the hard-core model in the graph C.

Now we want to find

$$\frac{\lambda}{2(1+\lambda)}\sup_C \left\{\mathbb{E}_C[\frac{1}{P_C(\lambda)}] + \frac{\lambda+1}{d}\mathbb{E}_C[\frac{P_C'(\lambda)}{P_C(\lambda)}] : \mathbb{E}_C[\frac{1}{P_C(\lambda)}] = \frac{1+\lambda}{d}\mathbb{E}_C[\frac{P_C'(\lambda)}{P_C(\lambda)}]\right\},$$

Let p_F be the probability of having $C = F$. We want to solve the following linear programming for all graphs F with at most d vertices.

$$\begin{aligned}\text{Maximize} : &\sum_F p_F(a_F + b_F), \\ \text{subject to} : &\sum_F p_F = 1, \sum_F p_F(a_F - b_F) = 0, p_F \geq 0,\end{aligned}$$

where $a_F = \frac{1}{P_F(\lambda)}$ and $b_F = \frac{(1+\lambda)P_F'(\lambda)}{dP_F(\lambda)}$. In particular, we have $a_\emptyset = 1, b_\emptyset = 0$ and $a_{\overline{K_d}} = (1+\lambda)^{-d}$ and $b_{\overline{K_d}} = 1$.

Consider the following dual program.

$$\begin{aligned}\text{Minimize} : &\ y_1, \\ \text{subject to} : &\ y_1 + y_2(a_F - b_F) \geq a_F + b_F, y_F \geq 0.\end{aligned}$$

By setting $y_1 = \frac{2}{2-(1+\lambda)^{-d}}$ and $y_2 = 1 - y_1$, we can check that $\frac{\lambda}{2(1+\lambda)}y_1$ is exactly $\frac{\lambda(1+\lambda)^{d-1}}{2(1+\lambda)^d - 1}$ which is what we desired. So, only remaining task is that this chose is a feasible solution of the dual program. In other words, we need to show that for each F, $y_1 + y_2(a_F - b_F) \geq$

$a_F + b_F$. This is equivalent to showing $\frac{\lambda P_F'(\lambda)}{P_F(\lambda)-1} \leq \frac{\lambda d(1+\lambda)^{d-1}}{(1+\lambda)^d-1}$. For this, we prove the following claim. This shows that the desired inequality is strict for all F other than two.

Claim 11. *For each graph $F \notin \{\emptyset, \overline{K_d}\}$ with at most d vertices, we have*

$$\frac{\lambda P_F'(\lambda)}{P_F(\lambda)-1} < \frac{\lambda d(1+\lambda)^{d-1}}{(1+\lambda)^d-1} = \frac{\lambda P_{\overline{K_d}}'(\lambda)}{P_{\overline{K_d}}(\lambda)-1}.$$

Proof. Let $P_F(\lambda) = 1 + \sum_{i=1}^{d} r_i \lambda^i$, then we have $(i+1)r_{i+1} \leq (d-i)r_i$ as there are at most $d-i$ ways to extend an independent set of size i to and independent set of size $i+1$. Let $t_i = \binom{d}{i}$ then we have $(i+1)t_{i+1} = (d-i)t_i$. This shows that we have $\frac{t_{k+i}}{t_k} \geq \frac{r_{k+i}}{r_k}$ for all $0 \leq k \leq k+i \leq d$.

Now, we have

$$\begin{aligned}\lambda P_{\overline{K_d}}'(\lambda)(P_F(\lambda)-1) - \lambda P_F'(\lambda)(P_{\overline{K_d}}(\lambda)-1) &= \sum_{k=1}^{d} \left(\sum_{i=1}^{k-1} i t_i r_{k-i} - \sum_{i=1}^{k-1} i t_{k-i} r_i \right) \lambda^k \\ &= \sum_{k=1}^{d} \left(\sum_{i=1}^{\lfloor k/2 \rfloor} (k-2i)(t_{k-i} r_i - t_i r_{k-i}) \right) \lambda^k.\end{aligned}$$

However, each $t_{k-i}r_i - t_i r_{k-i}$ is nonnegative, so we have the claim. □

Note that this claim shows that each condition on the dual program holds with our choice. This yields the desired inequality. □

Now, this shows that $\frac{1}{\lambda}\alpha_G(\lambda) = \frac{d}{d\lambda}\left(\frac{1}{|G|}\log P_G(\lambda)\right)$ is maximized by $K_{d,d}$ over all d-regular graphs for all λ. So, if we integrate this nonnegative function, then $\frac{1}{|G|}\log P_G(\lambda)$ is maximized by $K_{d,d}$ over all d-regular graphs. Hence, we have

$$\frac{1}{|G|}\log P_G(\lambda) \leq \frac{1}{|K_{d,d}|}\log P_{K_{d,d}}(\lambda),$$

implying that

$$P_G(\lambda) \leq P_{K_{d,d}}(\lambda)^{\frac{|G|}{2d}}.$$

By plugging $\lambda = 1$, we have the following theorem.

Theorem 10.6 (Kahn, Zhao, Galvin-Tetali)**.** *Let G be a n vertex d-regular graph. Then we have $i(G) \leq [i(K_{d,d})]^{n/2d}$.*

11. Threshold

Recall that we defined the threshold of certain property P to be the function $r = r(n)$ where $G(n,p)$ satisfies P with high probability if $p(n)/r(n) \to \infty$ and $G(n,p)$ does not satisfies P with high probability if $p(n)/r(n) \to 0$. We wish to find some upper bound on the threshold for certain natural monotonically increasing properties P.

In $G(n,m)$-model, we choose an m-edge graph uniformly at random. As it is known that $G(n,p)$ and $G(n,m)$ model are very similar when $p = m\binom{n}{2}^{-1}$. In other words, they have the same threshold for most of the natural graph properties. So, we instead focus on $G(n,m)$-model

For this purpose, we consider the following more general set-up. Consider each pair $ij \in \binom{[n]}{2}$ as vertices of hypergraphs, and consider each minimal subgraphs in P as edges of hypergraphs. If we choose a set U of m vertices independently uniformly at random what is the probability that U contains an edge of the hypergraph? If this probability tends to one, then p is an upper bound on the threshold.

Definition 11.1. *A hypergraph is k-bounded if all edges have size at most k.*

For a hypergraph $\mathcal{H}$ and a set $S \subseteq V(\mathcal{H})$, let $\langle S\rangle = \{T \supseteq S : T \subseteq V(\mathcal{H})\}$, and we write $\langle \mathcal{H}\rangle = \bigcup_{S\in E(\mathcal{H})}\langle S\rangle$. A hypergraph is r-spread if $|\mathcal{H} \cap \langle S\rangle| \le r^{-|S|}|\mathcal{H}|$ for all $S \subseteq V(\mathcal{H})$.

Theorem 11.2 (Frankston, Kahn, Narayanan, and Park, 2020+)**.** *There is a universal constant K such that any k-bounded $2r$-spread hypergraph $\mathcal{H}$ on $[n]$ (with repeated edges allowed), a random $\frac{Kn\log k}{r}$-element subset of $[n]$ chosen uniformly at random contains an edge of $\mathcal{H}$ with high probability.*

Proof. Let C_0 be a large constant. Let $p = \frac{C}{r}$ with $C_0 \le C \le r/C_0$, and let $k' = 0.9k$ and $N = \binom{n}{np}$.

Most natural attempt to prove this theorem is to choose a set W of certain size at random, and find a set $S \in E(\mathcal{H})$ where $S \setminus W$ is as small as possible and measure $|S \setminus W|$. However, finding best S is too difficult. Hence, we instead choose W at random and S at random, but this makes $|S \setminus W|$ not as small as we wish. So, we make compromise by choosing W at random and S at random, and choose a new set $S' \in E(\mathcal{H})$ which is better than S. We choose some $S' \subseteq W \cup S$ so that $|S' \setminus W| \le |S \setminus W|$. This choice of S' for given (W,S) will be encoded in the following function ψ, and χ will measure $|S' \setminus W|$.

Fix a map $\psi : \langle H\rangle \to H$ satisfying $\psi(Z) \subseteq Z$ for all $Z \in \langle \mathcal{H}\rangle$. For each $W \subseteq [n]$ and $S \in E(H)$, we let

$$\chi(S,W) = \psi(S \cup W) \setminus W.$$

Note that if $\chi(S,W) = \emptyset$, then it means that W contains an edge S. So, we want to ensure that $\chi(S,W)$ is small. We say that (S,W) is bad if $|\chi(S,W)| > k'$ and good otherwise.

The main part of the proof is the following lemma. By repeating this $\log k$ times, we can ensure that a randomly chosen set 'almost contains' an edge S of $\mathcal{H}$.

Lemma 11.3. *Assume $\mathcal{H}$ is r-spread on $[n]$. Assume the above set-up. Let W be chosen uniformly at random from $\binom{[n]}{np}$, then*

$$\mathbb{E}[|\{S \in H : (S,W) \text{ bad }]| \le e(H)C^{-k/3}.$$

Proof. It is enough to show that the number of bad pairs is at most $N|\mathcal{H}|C^{-k/3}$. Furthermore, let $\mathcal{H}_s = \{S \in \mathcal{H} : |S| = s\}$. It suffices to show that for each $s \in [k', k]$ that

$$|\{(S, W) \text{ bad } : S \in \mathcal{H}_s, W \subseteq [n]\}| \leq \frac{N|\mathcal{H}|}{kC^{-k/3}}.$$

We assume that $s \leq n/2$ as otherwise we have the following where m is the largest multiplicity of the edges of $\mathcal{H}$: $m \leq r^{-s}|\mathcal{H}| \leq k^{-s}m2^n$. This is contradiction as $r > C$.

We define a pair (S, W) to be pathological if there exists $T \subseteq S$ with $t = |T| > k'$ and

$$|\{S' \in \mathcal{H}_s : T \subseteq S' \subseteq S \cup W\}| \geq C^{k/2}|\mathcal{H}|r^{-t}p^{s-t}.$$

We say that the above T witnesses the pathology of (S, W).

Note that if W is randomly chosen from $\binom{[n]}{np}$, then the expected number of edges of size s containing T in $S \cup W$ is $|\mathcal{H}|r^{-t}|p^{s-t}$ as $\mathcal{H}$ is r-spread. Hence, the pathological pairs are very 'nontypical' cases. We will count pathological bad pairs and non-pathological bad pairs.

First, we count all non-pathological bad pairs by using the definition of pathological pairs. We generate all non-pathological bad pairs as follows.

Step 1. Choose a set Z of size between $[np, np + s]$ by $\sum_{i=0}^{s}\binom{n}{np+i} \leq \binom{n+s}{np+s} \leq Np^{-s}$ choices. This set Z will be our $W \cup S$ later.

Step 2. Let $S' = \psi(Z)$ and choose a subset T of S' with size $t > k'$. There are at most 2^k choices. This T will be our $S \cap S$ later.

Step 3. Choose an edge $S \in \mathcal{H}$ where $S \cap S' = T$ and $S \subseteq Z$. As we are only interested in non-pathological choices, the number of possibility is at most $C^{k/2}|\mathcal{H}|r^{-t}p^{s-t}$.

Step 4. Choose a set $S'' \subseteq S$ with at most 2^k choices.

At the end $W = Z \setminus (S \setminus S'')$ with S forms a bad pair. In total, there are

$$Np^{-s} \cdot 2^k \cdot C^{k/2}|\mathcal{H}|r^{-t}p^{s-t} \cdot 2^k \leq \frac{N|\mathcal{H}|}{C^{3k/8}}$$

non-pathological bad pairs.

In order to count pathological bad pairs, we first prove the following simple claim.

Claim 12. *If (S, W) is a pathological bad pair and T witnessing this pathology with $|T| = t$, then there exists a set $U = U(S, W)$ with $T \subseteq U \subseteq S$ with*

$$|\{X \in \mathcal{H}_s : U \subseteq X \subseteq (W \setminus S) \cup U\}| > 2^{-(s-t)}C^{k/2}|\mathcal{H}|r^{-t}p^{s-t} =: \Phi(t).$$

Proof. As T is witnessing the pathology of (S, W), there are at least $C^{k/2}|\mathcal{H}|r^{-t}p^{s-t}$ edges in $\mathcal{H}_s$ containing T. We partition them according to the intersection of the edge with S, then we have at least one U as above. □

Now, we count pathological bad pairs using the spreadness and the 'nontypicality' of pathological cases.

Step 1. Choose an edge $S \in \mathcal{H}$ with $|\mathcal{H}|$ choices.

Step 2. Choose $T \subseteq S$, there are at most 2^k choices. This set T will be a set witnessing pathology of the pair we are generating.

Step 3. Choose U with $T \subseteq U \subseteq S$ by at most 2^k choices. This U will be the set $U(S, W)$ for the pair (S, W) we construct later.

Step 4. Choose a set Y as follows: for any W with $S \cup W = S \cup Y$, (S, W) is a pathological pair with $U(S, W) = U$. We prove that there are at most $N(\frac{4}{C})^{k/2}$ choicse for this in the claim below.

Step 5. Choose a set $X \subseteq S$ which will be $S \cap W$.

Overall, the number of choices is at most

$$|\mathcal{H}| \cdot 2^{2k} \cdot N(\frac{4}{C})^{k/2} 2^k \leq \frac{N|\mathcal{H}|}{C^{3k/8}}.$$

It remains to show the following claim.

Claim 13. *There are at most $N(\frac{4}{C})^{k/2}$ choices for Y in the above Step 3.*

Proof. Choose Y from $\bigcup_{i=0}^{s} \binom{[n]\setminus S}{np-i}$ uniformly at random. Note that

$$|\bigcup_{i=0}^{s} \binom{[n] \setminus S}{np-i}| = \sum_{i=0}^{s} \binom{n-s}{np-i} \leq \binom{np}{s} \leq N.$$

Let $|U| = u$, then we have

$$|\mathcal{H}_s \cap \langle U \rangle| \leq |\mathcal{H} \cap \langle U \rangle| \leq |\mathcal{H}| r^{-u}.$$

For any $S' \in \mathcal{H}_s \cap \langle U \rangle$, we have

$$\mathbf{Pr}[Y \supseteq S' \setminus U] \leq (\frac{np}{n-s})^{s-u}.$$

So, we have

$$\mathbb{E}[|\{X \in \mathcal{H}_s : U \subseteq X \subseteq Y \cup Y\}|] \leq |\mathcal{H}| r^{-u} (\frac{np}{n-s})^{s-u}.$$

Hence, Markov's inequality implies that

$$\mathbf{Pr}[|\{X \in \mathcal{H}_s : U \subseteq X \subseteq Y \cup Y\}| > \Phi(t)] \leq \frac{|\mathcal{H}| r^{-u} (\frac{np}{n-s})^{s-u}}{\Phi(t)} \leq (\frac{4}{C})^{k/2}.$$

Hence we have at most

$$N\mathbf{Pr}[|\{X \in \mathcal{H}_s : U \subseteq X \subseteq Y \cup Y\}| > \Phi(t)] \leq N(\frac{4}{C})^{k/2}$$

choices for Y in Step 3. □

Overall, we have at mot $2\frac{N|\mathcal{H}|}{C^{3k/8}}$ bad pairs, proving the lemma. □

Assume that edges of $\mathcal{H}$ are ordered, and $\psi(Z)$ is the first edge of $\mathcal{H}$ lying inside Z.

Let $\mathcal{H}_0 = \mathcal{H}$. Let $k_i = 0.9^i k$. We choose m so that $0.9^m = \frac{\sqrt{\log k}}{k}$ and let $q = \frac{\log k}{r}$. We choose pn-subset W_i of $V(\mathcal{H}_i)$ uniformly at random, and let $\mathcal{H}_{i+1}$ be the hypergraph on the vertex set $V(\mathcal{H}_i) \setminus W$. Let $\chi_i(S, W_i) = S' \setminus W_i$ where S' is the first member of $\mathcal{H}_i$ inside S, W_i. Say $S \in \mathcal{H}_i$ is good if $|\chi_i(S, W_i)| \leq k_i$, otherwise bad. Let

$$\mathcal{H}_i = \{\chi_i(S, W_i) : S \in \mathcal{H}_i, S \text{ is good}\}$$

and we inherit the ordering in a natural way.

We say that i-th round is successful if $e(\mathcal{H}_{i+1}) > (1 - \frac{1}{2m}) e(\mathcal{H}_i)$. By the previous lemma and Markov's inequality, we know that

$$\mathbf{Pr}[i\text{-round not successful} \mid W_1, \ldots, W_{i-1} \text{ successful}] < 2mC^{-r_{i-1}/3}.$$

Note that assuming $W_1, \dots, W_{i-1}$, the hypergraph $e(\mathcal{H}_i) \geq \frac{1}{2}e(\mathcal{H})$, hence it is still r-spread (recall origial hypergraph was $2r$-spread.) Hence, the lemma ensures the above inequality. Hence, with probability at least $1 - \frac{1}{2m}\sum_{i=0}^{m} C^{-r_i/3} > 1 - \exp(-\sqrt{\log k})$, we have all m rounds successful, and get $\mathcal{H}_m$ with at least half of the edges of $\mathcal{H}$. Let $n' = |V(\mathcal{H}_m)|$.

To finish the remaining proof, we prove the following claim.

Claim 14. *Choose each vertex v in $V(\mathcal{H}_m)$ independently at random with probability $q' = qn/n'$, and let Y be the random set of chosen vertices. Then*

$$\mathbf{Pr}[Y \in \langle \mathcal{H}_m \rangle] \leq \exp^{-\sqrt{\log k}}.$$

Proof. We use Janson's inequality. Note that $\mathcal{H}$ is $k_m = 0.9^m k = \sqrt{\log k}$-bounded. Denote edges of $\mathcal{H}_m$ by S_i and let I_i be the indicator random variable of the event $S_i \subseteq Y$. Then $\mu = q'^r|\mathcal{H}_m|$ and

$$\begin{aligned}\Delta = \sum_{S_i \cap S_j \neq \emptyset} \mathbb{E}[I_i I_j] &\leq |\mathcal{H}_m| \sum_{t=1}^{k_m} \binom{k_m}{t} r^{-t} |\mathcal{H}| q^{2r-t} \\ &\leq \mu^2 \sum_{t=1}^{k_m} \binom{k_m}{t} (rq')^{-t} \leq \mu^2((1 + \frac{1}{rq'})^{k_m} - 1) \leq \mu^2 \frac{1}{\sqrt{\log k}}.\end{aligned}$$

By Janson's inequality, the probability that no events I_i occurs is at most $exp[-\frac{\mu^2}{2\Delta}] \leq exp[-\frac{1}{4}\sqrt{\log k}]$. □

As the above claim is for sets chosen randomly from all subsets of $V(\mathcal{H}_m)$, we need to make some changes for our purpose of choosing a set with fixed size. If we choose a set W of size qn from Z, then we have

$$\begin{aligned}\mathbf{Pr}[W \notin \langle \mathcal{H}_m \rangle] &\leq 2\mathbf{Pr}[|Y| \leq qn]\mathbf{Pr}[W \notin \langle \mathcal{H}_m \rangle] \\ &\leq 2\mathbf{Pr}[|Y| \leq qn]\mathbf{Pr}[Y \notin \langle \mathcal{H}_m \rangle \mid |Y| = qn] \\ &\leq \sum_{i=0}^{qn} 2\mathbf{Pr}[|Y| = i]\mathbf{Pr}[Y \notin \langle \mathcal{H}_m \rangle \mid |Y| = i] \\ &\leq 2\mathbf{Pr}[Y \notin \langle \mathcal{H}_m \rangle].\end{aligned}$$

Here, we have the penultimate inequality as choosing less vertices makes it more likely to not contain given edges. This we the above claim says that if we choose a random qn-set W_{m+1} from Z, then W_{m+1} contains an edge of $\mathcal{H}_m$ with probability at least $1 - 2exp[-\frac{1}{4}\sqrt{\log k}]$. Hence, $W_1 \cup \cdots \cup W_{m+1}$ contains an edge of $\mathcal{H}$ with probability at least $1 - (m+2)\exp[-\frac{1}{4}\sqrt{\log k}]$ which tends to 1 as k tends to infinity. As $W_1 \cup \cdots \cup W_{m+1}$ has the same distribution of choosing a random $(mp+q)$-set and $mp + q \leq \frac{3C\log k}{r}$, we obtained the desired conclusion. □

Note that many natural structures in graphs gives a well-spread hypergraphs as above. For example, if we consider the collection of all Hamilton cycles in $G(n,p)$ model, then it yields $O(n)$-spread n-uniform hypergraphs. This yields that the threshold of the existence of Hamilton cycles is $p = O(\frac{\log n}{n})$. Similarly, one can show that threshold of containing a specific copy T of bounded degree tree is $O(\frac{\log n}{n})$.