Topics in Combinatorics Sheet 1

Otto Pyper

1. We can count the LHS differently. For $b \in B$, let $b_x = \{x : b \in A + x\} = \{b - a : a \in A\}$, so $|b_x| = |A|$. Then $|(A + x) \cap B| = \sum_{b \in B} \mathbb{I}_{x \in b_x}$, and hence

$$\begin{aligned}\sum_{x \in \mathbb{Z}_n} |(A + x) \cap B| &= \sum_{x \in \mathbb{Z}_n} \sum_{b \in B} \mathbb{I}_{x \in b_x} \\ &= \sum_{b \in B} \sum_{x \in \mathbb{Z}_n} \mathbb{I}_{x \in b_x} \\ &= \sum_{b \in B} |b_x| = \sum_{b \in B} |A| \\ &= |A||B|\end{aligned}$$

Hence for some x, $|(A + x) \cap B| \geq |A||B|/n$, otherwise the sum over all x is too small.

Given $|A|$, $|B|$ and n, we have for all x that $|(A + x) \cap B| \leq \max_x |(A + x) \cap B|$, and so $|A||B| \leq n \max_x |(A+x) \cap B|$, and we have the lower bound $\max \geq |A||B|/n$, or the ceiling thereof if $|A||B|/n$ is not an integer.

In the case where $|A||B|/n$ is an integer and the bound is attainable, we must then have $|(A+x) \cap B| = |A||B|/n$ for all x, else the sum is again too small.

If $|A||n$, it is very helpful to find $A \leq \mathbb{Z}_n$, as then $\{(A+x) : 0 \leq x < n/|A|\}$ partitions $\mathbb{Z}_n$ into disjoint cosets, so for B we can freely pick k elements from each coset, for any fixed $1 \leq k \leq |A|$.

In general we do not have this niceness, but it may still be the case that $|A||n$, in which case for any given $|B|$ we can minimise $\max_x |(A + x) \cap B|$ by taking A a subgroup as above and then choosing B as evenly as possible from the cosets, so that for any x, y we have $||(A + x) \cap B| - |(A + y) \cap B|| \leq 1$, which is best possible (with all $= 0$ iff n divides $|A||B|$).

In even further generality, making no assumptions about the sizes $|A|, |B|$, we might still hope the above is possible - that for any x, y we have $||(A + x) \cap B| - |(A + y) \cap B|| \leq 1$.

The plan will be to space A 'as evenly as possible' across $\mathbb{Z}_n$, and take B to be a consecutive set of $|B|$ integers. Henceforth write $|A| = a$, $|B| = b$. As there is some non-empty intersection, we also assume wlog that $0 \in A \cap B$. Let $|A \cap B| = c$, and we assume further that this is the maximum size. Then for any x, we need $\lfloor ab/n \rfloor = c - 1 \leq |(A + x) \cap B| \leq c$. If there are y intersections of size $c - 1$, and $n - y$ of size c, then $y(c - 1) + (n - y)c = ab$, so $y = nc - ab = n - ab - \lfloor ab/n \rfloor$.

A set X is spaced 'as evenly as possible' if for all $x, y \in X$, $|x - y| \leq 1$ for an appropriate norm.

We will space A as evenly as possible. We find a consecutive pair of integers $s, s + 1$ such that there exist $p, q \in \mathbb{N} \geq 0$ such that $ps + q(s + 1) = n$ and $p + q = a$. In particular, $as = n - q$ and $a(s + 1) = n + p$, so $s \leq n/a \leq s + 1$. So $s = \lfloor n/a \rfloor$, $q = n - a\lfloor n/a \rfloor$.

My gut says this should be possible if we space A as evenly as possible, and then space the larger gaps as evenly as possible, and so on...

2. Choose a random subset $V \subset G$ of vertices, where for $x \in G$ we have $x \in V$ with probability p. Then for any $v, w \in G$, $\mathbb{P}[v \in V, w \in W, vw \in E(G)] = p(1 - p)m/\binom{n}{2}$. Hence

$$\begin{aligned}\mathbb{E}[\#\text{edges from } V \text{ to } W] &= \sum_{v \neq w \in V} p(1 - p)\frac{m}{\binom{n}{2}} \\ &= n(n - 1)p(1 - p)\frac{m}{\binom{n}{2}} \\ &= \frac{m}{2}\end{aligned}$$

Hence there exists some V for which at least half of the edges are between V and W.

3. (i) Presumably the ε_i are chosen from $\{-1,1\}$ each with probability $1/2$, so that $\mathbb{E}[\varepsilon_i]=0$, and $\mathbb{E}[\varepsilon_i^2]=1$ (if >0, the first result does not hold). Let $X=\sum_i a_i\varepsilon_i$. Then $\mathbb{E}\sum_i a_i\varepsilon_i=\sum_i a_i\mathbb{E}[\varepsilon_i]=0$ by linearity of expectation. Similarly, since $\mathbb{E}[\varepsilon_i\varepsilon_j]=\mathbb{E}[\varepsilon_i]\mathbb{E}[\varepsilon_j]=0$ for $i\neq j$ by independence, we have that $\mathbb{E}[(\sum_i a_i\varepsilon_i)^2]=\sum_i a_i^2\mathbb{E}[\varepsilon_i^2]=\sum_i a_i^2$.

(ii) We can expand X^{2k} and take expectation.

$$X^{2k}=\sum_{i_1+\cdots+i_n=2k}\frac{(2k)!}{i_1!\cdots i_n!}(a_1\varepsilon_1)^{i_1}\cdots(a_n\varepsilon_n)^{i_n}$$

When taking expectation, we remark that if i_j is odd for any j, then $\mathbb{E}[\varepsilon_j^{i_j}]=0$ so the expectation of the entire summand is zero. Thus every i_j is even, which we may write as $i_j=2m_j$. Hence

$$\mathbb{E}[X^{2k}]=\sum_{m_1+\cdots+m_n=k}\frac{(2k)!}{(2m_1)!\cdots(2m_n)!}a_1^{2m_1}\cdots a_n^{2m_n}$$

This expression is very similar to the expansion of $\mathrm{Var}[X]^k$, the difference being the factorial coefficient in each summand:

$$(\mathrm{Var}[X])^k=\sum_{m_1+\cdots+m_n=k}\frac{k!}{m_1!\dots m_n!}a_1^{2m_1}\dots a_n^{2m_n}$$

If we write $S=\{(m_1,\dots,m_n):\sum m_j=k\}$, we can express these sums as $\sum_{s\in S}A_sp_s$ and $\sum_{s\in S}B_sp_s$ respectively. Let t be such that A_t/B_t is maximised. Then for all s, we have $p_sA_sB_t\le p_sB_sA_t$, and hence (summing over s) we have that

$$\frac{\sum_{s\in S}A_sp_s}{\sum_{s\in S}B_sp_s}\le\frac{A_t}{B_t}$$

And the constant A_t/B_t is a quotient of multinomial coefficients, which is dependent only on k.

4. We choose a random antisymmetric relation on $[N]$, by saying that for $i<j$, $(i,j)\in R$ with probability p, and $(j,i)\in R$ with probability $1-p$. That is to say we go through each subset $\{i,j\}$, and flip a coin with probability p of landing heads. If heads, we put $(i,j)\in R$. If tails, we put $(j,i)\in R$.

Then define X_S to be the indicator function of the event 'there exists $x\in[N]$ with xRs for all $s\in S$', and then define $X=\sum_{S\in[N]^{(k)}}X_S$. We will take $\mathbb{E}[X]$, so we need to know $\mathbb{E}[X_S]$.

I'm going to write down a big formula and explain it later.

$$\mathbb{P}[X_S=0]=\prod_{m=0}^{k}\left[(1-p^{k-m}(1-p)^m)^{s_{m+1}-s_m-1}\right]$$

The idea here is that we have a k-set $S=\{s_1<s_2<\cdots<s_k\}$, and we've defined $s_0=0$, $s_{k+1}=N+1$. We then take the product over all $x\in[N]\backslash S$ that we do not have xRs for all $s\in S$. This exact probability depends only on where x is relative to the s_i; so if it is between s_m and s_{m+1}, then the probability x relates to all of them is $(1-p)^mp^{k-m}$, so $1-$ that is the probability this doesn't happen.

Taking the product over all $s_m<x<s_{m+1}$ gives the probability that these all fail, hence the exponent $s_{m+1}-s_m-1$ in the multiplicand. Then taking the product over all such m gives the probability we have success for no $x\in[N]\backslash S$. We will henceforth denote this product as p_S.

So now we have

$$\begin{aligned}\mathbb{E}[X] &= \sum_{S\in[N]^{(k)}} (1-p_S) \\ &= \binom{N}{k} - \sum_{S\in[N]^{(k)}} p_S\end{aligned}$$

and the idea is to appropriately choose p and N so that the latter sum is less than 1. Then since X takes integer values, there must be some relation for which $X = \binom{N}{k}$, *i.*e. every k-set is related to by some x.

While it could have been helpful, it turns out that we didn't need p in full generality, and could have just used $p = 1/2$. This hugely simplifies the situation, as we then just have $p_S = (1-2^{-k})^{N-k}$. The latter sum is then $\binom{N}{k}(1-2^{-k})^{N-k}$, which is exponentially decaying in N and so is indeed eventually small enough.

This also gives us an upper bound on how large N needs to be, as

$$\begin{aligned}\binom{N}{k}(1-2^{-k})^{N-k} &< 1 \\ \Longleftarrow \left(\frac{eN}{k}\right)^k \left(1-2^{-k}\right)^{N-k} &< 1 \\ \iff N^k(1-2^{-k})^N &< \left(\frac{k}{e}\right)^k (1-2^{-k})^k = c_k \\ \iff k\log N + N\log(1-2^{-k}) &< \log c_k \\ \iff \frac{N}{\log N} &> -\frac{k}{\log(1-2^{-k})} + \frac{\log c_k}{\log N \log(1-2^{-k})}\end{aligned}$$

For any fixed k, the latter term vanishes as $N \to \infty$, so we can ignore this term (it is also negative, so we will obtain a stronger condition).

For any $\varepsilon > 0$, we have $N^\varepsilon > \log N$ for sufficiently large N, and hence $N/\log N < N^{1-\varepsilon}$. Thus, for any $\varepsilon > 0$ and sufficiently large k, N, and noting that $\log(1-2^{-k}) \approx -2^{-k}$, we have the sufficient condition for a solution:

$$|X| \geq k^{\frac{1}{1-\varepsilon}} 2^{k/(1-\varepsilon)}$$

or, for instance, a uniform bound $|X| \geq k^2 2^{2k}$ implies such an antisymmetric relation exists.

5. Colour each member of X red with probability p and blue with probability $1-p$. Let X_i be the indicator of the event that A_i is not monochrome, and let $X = \sum_{i=1}^r X_i$.

$\mathbb{P}[X_i = 1] = 1 - (1/2)^m - (1-1/2)^m = 1 - 2^{1-m}$, since $X_i = 1$ iff the elements of A_i are not all red and not all blue. Hence:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^r \mathbb{P}[X_i = 1] \\ &= r(1-2^{1-m}) = r - 2^{1-m}r\end{aligned}$$

So if $r < 2^{m-1}$ then $2^{1-m}r < 1$, and hence $\mathbb{E}[X] > r-1$, so there exists some colouring for which $X \geq r$; but $X \leq r$, so there exists a colouring for which every A_i contains at least one red element and at least one blue element.

Let $R(m)$ be the least r such that there exist sets $A_1, \dots, A_r$ of size m such that for every red-blue colouring there is some i with A_i monochrome. We can bound $R(m)$ recursively.

Given $R(m)$, construct $R(m)$ $m+1$-sets by extending the $A_1, \dots, A_{R(m)}$ m-sets.

Pick a set of $B = \{b_i : 1 \leq i \leq m+1\}$ such that $B \cap A_i = \emptyset$ for all i, and then take the sets $C_{i,j} = A_i \cup \{b_j\}$, altogether along with B. Then in any red-blue colouring there must be some i such that $A_i \subset C_{i,j}$ is monochrome, say blue. So then we either have a monochrome set, or every b_j has been coloured red - in which case B is monochrome.

Therefore $R(m+1) \leq R(m)(m+1) + 1$. Thus:

$$\begin{aligned}
R(m) &\leq 1 + R(m-1)m \\
&\leq 1 + m + R(m-2)(m-1)m \\
&\leq \sum_{j=0}^{k} \frac{m!}{(m-j)!} + R(m-(k+1))(m-k)(m-(k-1))\dots(m) \\
&\leq \sum_{j=0}^{m-2} + R(1)m! = \sum_{j=0}^{m-1} \frac{m!}{(m-j)!} = m! \sum_{j=0}^{m-1} \frac{1}{(m-j)!} \\
&= m! \sum_{\ell=1}^{m} \frac{1}{\ell!} \leq m!(-1 + \sum_{\ell=0}^{\infty} \frac{1}{\ell!}) \\
&\leq m!(e-1)
\end{aligned}$$

So we have an upper bound $R(m) \leq m!(e-1)$ for each m.

6. Write $v_i = (v_{i,j})_{j=1}^n$. Consider $V = || \left(\sum_i \varepsilon_i v_i\right)^2 ||$, where the ε_i are random variables as in Q3.

$$\begin{aligned}
\mathbb{E}[V] &= \sum_{j=1}^{n} \mathbb{E}\left[\left(\sum_{i=1}^{n} \varepsilon_i v_{i,j}\right)^2\right] \\
&= \sum_{j=1}^{n}\sum_{i=1}^{n} v_{i,j}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} v_{i,j}^2 \\
&= \sum_{i=1}^{n} ||v_i||^2 \\
&= n
\end{aligned}$$

So there must exist some signs ε_i such that $|| \sum_i \varepsilon_i v_i||^2 \geq n$.

7. Choose a random ordering of the vertices of G, and then colour the vertices with the greedy algorithm - *i.e.* $c(v_1) = 1$, $c(v_n)$ is the smallest m not in $\{c(v_k) : 1 \leq k < n, v_k \in \Gamma(v_n)\}$. Let X be the number of vertices with colour 1.

Then $X = \sum_{v \in G} X_v$, where X_v is the indicator of the event that v appears before all of its neighbours in the random ordering. So $\mathbb{P}[X_v = 1] = 1/(\Gamma(v)+1) \geq 1/(d+1)$, and hence $\mathbb{E}[X] \geq n/(d+1)$. So there exists an independent set of size at least $n/(d+1)$.

If instead the average degree is d, then we have

$$d + 1 = \frac{\sum_{v \in G} \Gamma(v) + 1}{n} \geq \frac{n}{\sum_{v \in G} \frac{1}{\Gamma(v)+1}}$$

by AM≥HM. Hence

$$\mathbb{E}[X] = \sum_{v \in G} \frac{1}{\Gamma(v)+1} \geq \frac{n}{d+1}$$

which exceeds the $n/2(d+1)$ bound. In particular, there exists an independent set of size $\geq n/(d+1)$.

8. $\mathcal{A} \subset \mathcal{P}[n]$, such that $A, B \in \mathcal{A} \implies |A \Delta B| \neq 2$.

Construct a graph $G = (V, E)$ where $V = \mathcal{P}[n]$ and $AB \in E$ iff $|A\Delta B| = 2$. Let V_k denote the vertices A such that $|A| = k$. We have $|V_k|\binom{n}{k}$.

Note that each A has exactly $\binom{n}{2}$ neighbours; we can see this since there are $\binom{n}{2}$ possibilities for the symmetric difference $A\Delta B$, which will determine B completely.

Or we remark that for $|A| = k$, we have $d(A) = \binom{k}{2} + k(n-k) + \binom{n-k}{2} = \binom{n}{2}$ after a bit of algebra.

Q7 then immediately shows the existence of a set system $\mathcal{A}$ of size $\left(\binom{n}{2}+1\right)^{-1} 2^n$.

But we also want the upper bound. For this, we consider for a given set A what is the largest possible size of an independent subset of $\Gamma(A)$.

Quasi-partition $\Gamma(A)$ into classes $C_1, \ldots, C_n$, where $B \in \Gamma(A)$ lies in C_i if $i \in A\Delta B$. Each $B \in \Gamma(A)$ then lies in exactly two C_i, corresponding to the elements of $A\Delta B$.

In general we have that $B_1 \Delta B_2 = (A\Delta B_1)\Delta(A\Delta B_2)$, so if $B_1 \neq B_2 \in C_k$ then $|B_1\Delta B_2| = 2$, so they are not independent.

Hence an independent subset of $\Gamma(A)$ may only have at most one member from each C_i, and since each B lies in two classes, we can have at most $\lfloor n/2 \rfloor$ - otherwise two lie in the same class.

We now construct let $X = \mathcal{A}$, $Y = \mathcal{P}[n]\backslash A$, and count the edges between X and Y.

Every $x \in X$ has exactly $\binom{n}{2}$ edges into Y, and for each $y \in Y$ there are at most $\lfloor n/2 \rfloor$ edges into X, as we saw above.

Hence $\binom{n}{2}|X| \leq \lfloor n/2 \rfloor |Y|$. In particular:

$$\begin{aligned} |\mathcal{A}| &\leq \frac{\lfloor n/2 \rfloor (2^n - |\mathcal{A}|)}{\binom{n}{2}} \\ \implies |\mathcal{A}| &\leq \frac{\lfloor n/2 \rfloor 2^n}{\binom{n}{2} + \lfloor n/2 \rfloor} \end{aligned}$$

Which gives $|\mathcal{A}| \leq 2^n/(n+1)$ for n odd, and $|\mathcal{A}| \leq 2^n/n$ for n even. Hence in all cases we have $|\mathcal{A}| \leq n^{-1}2^n$.