21-738 Homework 1

Wilson Pan

February 2, 2026

Problem 1

Let $\mathcal{H}$ be an arbitrary 3-uniform hypergraph with $\epsilon\binom{n}{3}$ edges. Let S be a random subset of vertices in $\mathcal{H}$ where each vertex of $\mathcal{H}$ is included with probability p. We will choose p later to optimize the result. Let $X = |v(S)|$ then

$$\mathbb{E}[X] = np.$$

Similarly, let $Y = |e(S)|$ then

$$\mathbb{E}[Y] = \epsilon\binom{n}{3}p^3.$$

Then we can perform the alteration where we delete a vertex from edges to produce an independent set with say $\alpha(S)$ vertices then $\alpha(S) \geq X - Y$ as we could've possibly deleted vertices in another edge so it deleted at least 2 edges. So

$$\mathbb{E}[\alpha(S)] \geq \mathbb{E}[X - Y] = np - \epsilon\binom{n}{3}p^3 \geq np - \epsilon\frac{n^3}{6}p^3.$$

To optimize $f(p) = np - \epsilon\frac{n^3}{6}p^3$,

$$\frac{df}{dp} = n - \epsilon\frac{n^3}{2}p^2.$$

So it is optimized when $p = \frac{\sqrt{2}}{n\sqrt{\epsilon}}$ So

$$\mathbb{E}[\alpha(S)] \geq \sqrt{2}\epsilon^{-\frac{1}{2}} - \frac{\sqrt{2}}{3}\epsilon^{-\frac{1}{2}} = \frac{2\sqrt{2}}{3}\epsilon^{-\frac{1}{2}}$$

Since $\frac{2\sqrt{2}}{3} > \frac{1}{10}$ then we have achieved the necessary bound.

Problem 2

Let X and Y be the number of $K_{1,s}$ and $K_{s,t}$ subgraphs in G, respectively. We can count X by choosing any s-subset of $V(G)$ and choosing a vertex in the common neighbors

$$X = \sum_{T \in \binom{V(G)}{s}} \binom{n(T)}{1} \text{ where } n(T) \text{ is the common neighbor of all vertices in } T. \tag{1}$$

Similarly, we can count X by picking any vertex in $V(G)$ and picking s vertices from its neighborhood. So

$$X = \sum_{v \in V(G)} \binom{d(v)}{s} \geq n \cdot \binom{d(v) \text{ -avg}}{s}.$$

Last inequality is from $\binom{x}{s}$ being convex for $x \geq s-1$. The average degree is

$$2\frac{\epsilon\binom{n}{2}}{n} = \epsilon(n-1).$$

Using $\binom{k}{p} \geq \frac{(k-p+1)^p}{p!}$ So

$$X \geq n \cdot \binom{(d\text{-avg} - s + 1)}{s} = \frac{n}{s!}(\epsilon(n-1) - s + 1)^s.$$

Condition 1: We can choose C such that $\epsilon(n-1) - s + 1 \geq \frac{1}{2}\epsilon n$.
When we rearrange we have

$$\frac{1}{2}\epsilon n \geq \epsilon + s - 1.$$

Since $\epsilon \leq 1$, $\epsilon + s - 1 \leq s$. So we want $2s \leq \epsilon n$ then $\epsilon n > Cn^{1-\frac{1}{s}}$ and $s \geq 1$ so $Cn^{1-\frac{1}{s}} \geq C$. So we can choose $C = 2s$ and the inequality holds.
Note: This would assume $n > 2s$ as otherwise density would be greater than 1.

Consequently,

$$X \geq \frac{n}{s!}\frac{\epsilon^s n^s}{2^s}. \tag{2}$$

To count Y we can pick any subset of s vertices from $V(G)$ and choose any t vertices from the common neighbors. So

$$Y = \sum_{T \in \binom{V(G)}{s}} \binom{n(T)}{t} \geq \binom{n}{s}\binom{n(T)\text{-avg}}{t} = \binom{n}{s}\binom{\frac{\sum_{T\in\binom{V(G)}{s}} n(T)}{\binom{n}{s}}}{t} = \binom{n}{s}\binom{\frac{X}{\binom{n}{s}}}{t}.$$

Last equality is from (1). Let $\overline{m} = \frac{X}{\binom{n}{s}}$ then we can lower bound X using (2) and upper bound using $\binom{n}{s} \geq \frac{n^s}{s!}$

$$\overline{m} \geq \frac{\frac{n}{s!}\frac{\epsilon^s n^s}{2^s}}{\frac{n^s}{s!}} = \frac{n\epsilon^s}{2^s}.$$

Condition 2: We want to pick $\overline{m} \geq 2t$ to ensure $\binom{\overline{m}}{t}$ maintains a factor of n^t. Such a $\overline{m}$ is possible as

$$\overline{m} > \frac{C^s}{2^s} > 2t.$$

We can simply choose C maximum of this condition and condition 1 to satisfy both.

We can now bound Y using $\binom{n}{s} \geq \frac{(n-s+1)^s}{s!} \geq \frac{(\frac{n}{2})^s}{s!}$ by assumption $n > 2x$.

$$Y \geq \binom{n}{s}\binom{\overline{m}}{t} \geq \frac{(n/2)^s}{s!} \cdot \frac{(\overline{m} - t + 1)^t}{t!}.$$

From condition 2 we have $\overline{m} - t + 1 \geq \frac{\overline{m}}{2}$ so

$$\frac{(n/2)^s}{s!} \cdot \frac{(\overline{m} - t + 1)^t}{t!} \geq \frac{(n/2)^s}{s!} \cdot \frac{(\overline{m}/2)^t}{t!} \geq \frac{(n/2)^s}{s!} \cdot \frac{1}{t!} \left(\frac{\epsilon^s n}{2^{s+1}} \right)^t.$$

The last inequality from substituting in $\overline{m}$. So rearranging we have

$$Y \geq \left(\frac{1}{s!t!2^{s+st+t}} \right) \epsilon^{st} n^{s+t}.$$

So we can let $c = \frac{1}{s!t!2^{s+st+t}}$

Problem 3

Greta's idea does not work as you can have a large independent set of size n with $n >> r$ and a small clique of size r. Then our graph is the union of the independent set, a vertex connected to every vertex in the independent set, say v and a clique of size r. When you pick the largest degree vertex which is v, then $N(v)$ has no edges between itself, so the induction step fails.

Her idea can work, we'll prove by induction that the the size of the maximal clique $\omega(G)$ satisfies

$$\omega(G) \geq \frac{n^2}{n + 2\overline{m}}.$$

Where $\overline{m}$ are the edges missing from the complete graph on n vertices.
(Base Case): For $n = n$, $\overline{m} = 0$ so our result holds true as $\omega(G) \geq 1$.
(Induction Step): Assume the result holds for graphs of fewer than n vertices, then instead of picking the largest degree vertex, we pick vertex v that has the minimum degree in the complement graph $\overline{G}$. Let $\gamma = \deg(v)$ then $|N(V)| = n - \gamma - 1$. Let G' be the subgraph induced by $N(v)$ and $\overline{m}$ be number of non edges inside $N(V)$. Then we have $n' := |v(G')| = n - \gamma - 1$. By induction the clique size in G' is at least

$$\omega(G') \geq \frac{(n')^2}{n' + 2\overline{m'}}.$$

It suffices to prove that $\omega(G') + 1 \geq \frac{n^2}{n+2\overline{m}}$.
We want to relate $\overline{m}$ with $\overline{m'}$, the total number of non edges $\overline{m}$ consist of non edges incident to v (exactly γ), non edges incident to vertices not connected to v, say X and non edges strictly inside $N(v)$. A loose but sufficient bound is

.

We know $2\overline{m} = 2\overline{m'} +$ (edges incident to $X \cup \{v\}$ in $\overline{G}$). So $2\overline{m'} \leq 2\overline{m} - \gamma(\gamma + 1)$.
We need to show

$$1 + \frac{(n - \gamma - 1)^2}{2\overline{m'} + (n - \gamma - 1)} \geq \frac{n^2}{2\overline{m} + n}.$$

Let $D' = 2\overline{m'} + n - \gamma - 1$, and substituting in our bound we have $D' \leq 2\overline{m} - \gamma(\gamma + 1) + n - \gamma - 1$. Letting $x = 2\overline{m} + n$ we have $D' \leq x - (\gamma + 1)^2$.
Our condition becomes

$$1 + \frac{(n - (\gamma + 1))^2}{x - (\gamma + 1)^2} \geq \frac{n^2}{x}.$$

Cross multiplying and comparing LHS-RHS's numerator we have

$$\begin{aligned}
Num &= [x + (n - (\gamma + 1))^2 - (\gamma + 1)^2] \cdot x - n^2[x - (\gamma + 1)^2] \\
&= \left[x + n^2 - 2n(\gamma + 1)\right] \cdot x - n^2 \left[x - (\gamma + 1)^2\right] \\
&= x^2 + xn^2 - 2xn(\gamma + 1) - xn^2 + n^2(\gamma + 1)^2 \\
&= x^2 - 2xn(\gamma + 1) + n^2(\gamma + 1)^2 \\
&= (x - n(\gamma + 1))^2 \geq 0
\end{aligned}$$

Thus the induction step holds.
A graph with density $(1 - \epsilon)$ is missing roughly $\epsilon \frac{n^2}{2}$ edges so $\overline{m} \approx \epsilon \frac{n^2}{2}$. So the clique size is at least

$$\frac{n^2}{\epsilon n^2 + n} = \frac{1}{\epsilon + \frac{1}{n}}.$$

This approaches $+\infty$ as $n \to \infty$ and $\epsilon \to 0$.

Problem 4

Let S be a random subset of vertices in G where each vertex in G is included with probability p. Then let X be the number of vertices of S and T be the number of triangles in S. We let

$$T = \sum_{\{i,j,k\} \in \binom{V(G)}{3}} I_{i,j,k} \text{ where } I_{i,j,k} = \begin{cases} 1 & \text{if } i,j,k \text{ form a triangle} \\ 0 & \text{otherwise} \end{cases}.$$

$$\mathbb{E}[X] = np \text{ and } \mathbb{E}[T] = \sum_{\{i,j,k\} \in \binom{V(G)}{3}} \Pr[I_{i,j,k}] = \binom{n}{3} p^3 \epsilon \leq \frac{n^3}{6} p^3 \epsilon.$$

$\mathbb{E}[T]$ as each triangle survives with probability p^3. So there exist a choice of vertices such that $X \leq np$ and $T \geq \frac{p^3 \epsilon^3 n^3}{6}$.
To produce U that is triangle free we can delete a vertex from each triangle so $|v(U)| \geq X - T \geq np - \frac{n^3}{6} p^3 \epsilon$. This is $\geq$ as we can delete a vertex in multiple triangles potentially.
To maximize the lower bound we can let $f(p) = np - \frac{n^3}{6} p^3 \epsilon$ then

$$f'(p) = n - \frac{p^2 \epsilon n^3}{2}.$$

So the maximizer is $p = \frac{\sqrt{2}}{n} \epsilon^{-\frac{1}{2}}$
Consequently

$$|v(U)| \geq \frac{2\sqrt{2}}{3} \epsilon^{-\frac{1}{2}}.$$

So we let $C = \frac{1}{2}$

Problem 5

Let the graph be G then assume the converse that G is not bipartite so there must exist an odd cycle and take the smallest such odd cycle say C of length k then as there exist no triangles so $k \geq 5$.

Note: There cannot be a chord in the C, this is seen if we say $C = (v_1, ..., v_k)$ then take any v_i, v_j and WLOG let $i < j$ then $v_i \not\sim v_j$ if v_i and v_j have common neighbors so $j > i + 2$. Assume this is true and $v_i \sim v_j$ then consider the two paths from v_i to v_j say P, Q then one of P or Q is even as C is of odd length and with edge from v_i to v_j we have produced a shorter cycle.

Let $G' = G \setminus C$ then every vertex in C is adjacent to at least $\frac{2n}{5} - 2$ vertices in G' as we exclude the neighbors of in C. The edges between G' and C is at least

$$k\left(\frac{2n}{5} - 2\right) = 2\left(\frac{nk}{5} - k\right) \geq 2\left(n - k\right).$$

Last inequality as $\frac{k}{5} \geq 1$. As $|G'| = n - k$ then there exist a vertex $v \in G'$ such that v is adjacent to 3 vertices, say v_1, v_2, v_3 in C by pigeonhole. Then for some two of them say v_1, v_2 there exist a path in C between v_1, v_2 such that it's odd length and less than $k - 2$ as C is odd length. Then connect such path with v then we have an odd cycle smaller than C. Thus a contradiction.

Problem 6

Part A

For any walk (u, v, w) we can pick a center v then choose 2 vertices from it's neighbors and it's possible for $u = w$ so the total number of length 2 walks is

$$\sum_{v \in V(G)} \deg(v)^2.$$

By Jensen's Inequality on $f(x) = x^2$ (f is convex)

$$\frac{1}{n} \sum_{v \in V(G)} \deg(v)^2 \geq \left(\frac{1}{n} \sum_{v \in V(G)} \deg(v) \right)^2 = d^2.$$

So

$$\sum_{v \in V(G)} \deg(v)^2 \geq nd^2.$$

Part B

Consider an arbitrary path $(v_0, v_1, v_2, v_3, v_4)$ then we can first pick v_2 then select v_0, v_1 from the 2-paths ending at v_2 and v_3, v_4 from the 2-paths starting at v_2. Notice the number of such paths in each part is the same as they're independent of one another. Let P_4 be the number length 4-paths, then we can count the 2-path starting at v then choosing a neighbor of v say u and then a neighbor of u.

$$\begin{aligned} P_4 &= \sum_{v \in V(G)} \left(\sum_{u \in N(v)} \deg(u) \right)^2 \\ &\geq \frac{1}{n} \left(\sum_{v \in V(G)} \sum_{u \in N(v)} \deg(u) \right)^2 && \text{(Jensen's Inequality on } f(x) = x^2\text{)} \end{aligned}$$

We claim

$$\sum_{v \in V(G)} \sum_{u \in N(v)} \deg(u) = \sum_{w \in V(G)} \deg(w)^2.$$

For a specific vertex w, its degree $\deg(w)$ is added to the sum every time w appears as a neighbor u of some vertex v. Since w has $\deg(w)$ neighbors, it plays the role of "neighbor u" exactly $\deg(w)$ times and it contributes $\deg(w)$ each time.

Using our previous claim and Jensen's Inequality

$$P_4 \geq \frac{1}{n} \left(\sum_{w \in V(G)} \deg(w)^2 \right)^2 \geq \frac{1}{n} \left(\frac{1}{n} \left(\sum_{w \in V(G)} \deg(w) \right)^2 \right)^2 = \frac{1}{n} \left(\frac{1}{n} (nd)^2 \right)^2 = nd^4.$$

Thus we're done.

Part C

We know the total number of walks of length 4 in G is at least nd^4. We will denote W_4 to be the number of walks of length 4 in G. We will show that the number of bad walks is small relative to the number, specifically $\leq \epsilon(d) W_4$. We will condition on when a walk is bad:

1. $v_0 = v_2$
2. $v_2 = v_4$

3. $v_0 = v_3$
4. $v_1 = v_4$
5. $v_0 = v_4$

We will iteratively remove any vertices with degree less than $\delta = \sqrt{d}$. Let the remaining subgraph be G', the total number of edges removed cannot be greater than $n\delta$ as we can remove at most n vertices. To count the number of bad walks

1. If $v_0 = v_2$ then the number of bad walks, B_1 is

$$B_1 = \sum_{v_0 \in V(G')} \deg(v_0) \left(\sum_{v_3 \sim v_0} \deg(v_3) \right)^2.$$

This is from we choose v_1 from the neighbors of v_0 and then choosing v_3 from the neighbors of v_0 and then choose v_4 from the neighbors of v_3.
Let $S(v) = \sum_{u \sim v} \deg(u)$ then from part b we've established $W_4 = \sum_{v \in V(G)} S(v)^2$. We can bound S from below by $\deg(u) \geq \delta$ for $u \in V(G')$. So

$$S(v) \geq \sqrt{d} \deg(v) \iff \deg(v) \leq \frac{1}{\sqrt{d}} S(v).$$

So

$$B_1 = \sum_{v_0 \in V(G')} \deg(v_0) S(v_0) \leq \sum_{v_0 \in V(G')} \frac{1}{\sqrt{d}} (S(v_0))^2 \leq \frac{1}{\sqrt{d}} \sum_{v_0 \in V(G')} S(v_0)^2 = \frac{1}{\sqrt{d}} W_4.$$

2. If $v_2 = v_4$, by symmetry with case 1 we have $B_2 \leq \frac{1}{\sqrt{d}} W_4$.

3. If $v_0 = v_3$ then the number of bad walks, B_3 is strictly less than number of walks where v_3 is completely free. When we constraint $v_3 = v_0$ we decrease the count by

$$\frac{\text{Ways to choose } v_3 = v_0}{\text{Total ways to choose } v_3} \leq \frac{1}{\deg(v_2)} \leq \frac{1}{\sqrt{d}}.$$

So $B_2 \leq O(\frac{1}{\sqrt{d}}) W_4$.

4. If $v_1 = v_4$ then by symmetry with case 3 we have $B_3 \leq O(\frac{1}{\sqrt{d}}) W_4$.

5. If $v_0 = v_4$ then consider a walk $v_0 \to v_1 \to v_2 \to v_3 \to v_4$ with $v_0 = v_4$, v_3 must be a neighbor of v_0. So fixing v_4 to be v_0 we decrease the count by at roughly $\frac{1}{\deg(v_0)} \leq \frac{1}{\sqrt{d}}$ so $B_4 \leq O(\frac{1}{\sqrt{d}}) W_4$.
Note: This case is likely much less since v_3 has to be a neighbor of v_0.

Thus the total number of bad walks is at most

$$B_1 + B_2 + B_3 + B_4 + B_5 \leq 5 O\left(\frac{1}{\sqrt{d}}\right) W_4 = O(\epsilon(d)) W_4.$$

So the number of good walks is at least

$$W_4 - B \geq \left(1 - \frac{C}{\sqrt{d}}\right) W_4.$$

So we're done by letting $\epsilon(d) = \frac{C}{\sqrt{d}}$

Problem 7

To show associativity we need $a \circ (b \circ c) = (a \circ b) \circ c$ for all $a, b, c \in \mathbb{F}_{p^2}$. This is true as if c is a square both sides reduce to standard multiplication. Otherwise we have

$$a \circ (b \circ c) = a \circ (b^p c) = a^p b^p c = (ab)^p c = (a \circ b) \circ c.$$

To show right distributivity we need $(a + b) \circ c = (a \circ c) + (b \circ c)$ for all $a, b, c \in \mathbb{F}_{p^2}$. If c is a square then both sides reduce to standard multiplication. Otherwise $(a+b)^p c = (a^p + b^p)c = a^p c + b^p c = (a \circ c) + (b \circ c)$

To show that $\sim$ is an equivalence relation.

1. Reflexive: $(a, b, c) \sim (a, b, c)$ as we set $\lambda = 1$
2. Symmetric: Every λ has an inverse λ^{-1} as $\mathbb{F}_{p^2}$ is a field.
3. Transitive: If $u \sim v$ (via λ) and $v \sim w$ (via μ) then $u \sim w$ (via $\lambda\mu$).

To show the lines make sense, we have $\ell_{a,b,c} = \{(x, y, z) \in P : a \circ x + b \circ y + c \circ z = 0\}$. If we take any scalar $(x \circ \lambda, y \circ \lambda, z \circ \lambda)$ we want to show this also satisfies. We have

$$\begin{aligned} a \circ (x \circ \lambda) + b \circ (y \circ \lambda) + c \circ (z \circ \lambda) &= (a \circ x) \circ \lambda + (b \circ y) \circ \lambda + (c \circ z) \circ \lambda && \text{(associativity)} \\ &= (a \circ x + b \circ y + c \circ z) \circ \lambda && \text{(right distributivity)} \\ &= 0 \circ \lambda = 0 \end{aligned}$$

To show that this structure is a projective plane, we need to show that for any two points there is a unique line through them, for any two lines there is a unique point containing both and there exist four points no three of which are collinear.
Let the two points be (x_1, y_1, z_1) and (x_2, y_2, z_2) then the line through them is $\ell_{a,b,c}$ where

$$\begin{cases} a \circ x_1 + b \circ y_1 + c \circ z_1 = 0 \\ a \circ x_2 + b \circ y_2 + c \circ z_2 = 0 \end{cases}.$$

Since our operation satisfies right distributivity we have a system of equations with respect to a, b, c. We have 3 unknowns and 2 equations so the solution space has dimension at least $3 - 2 = 1$. Thus a non-zero solution exist. Since the points are distinct then the equations are independent so the solution space has dimension exactly 1. This means all solutions are scalar multiples of each other (L and $L \circ \lambda$), which represents the same unique line.

We'll show explicitly there exist 4 points no three of which are collinear. Let $P_0 = (1, 0, 0)$, $P_1 = (0, 1, 0)$, $P_2 = (0, 0, 1)$ and $P_3 = (1, 1, 1)$. Then the lines are $\ell_{1,0,0}$, $\ell_{0,1,0}$, $\ell_{0,0,1}$ and $\ell_{1,1,1}$.
Consider line P_1P_2 the equation $a \circ x + b \circ y + c \circ z = 0$ must satisfy $(1, 0, 0), (0, 1, 0)$. So we know $a \circ 1 = 0 \implies a = 0$ and $b \circ 1 = 0 \implies b = 0$. So the line is $z = 0$. It is clear that P_2, P_3 are not on this line. By symmetry, no combination of these three points are collinear.

To show that for any two lines they meet at a unique point, we'll first show $f(x) = M \circ x$ is bijective. To show injectivity, if $M \circ x_1 = M \circ x_2$ then if both are squares $Mx_1 = Mx_2 \implies x_1 = x_2$, if both are not squares $M^p x_1 = M^p x_2 \implies x_1 = x_2$ and the last case if one is square and the other not, we have $M^p x_1 = M x_2 \implies M^{p-1} = x_2 x_1^{-1}$, it is easy to see LHS is a square by bringing to $p^2 - 1$ power but RHS is not square so this case cannot happen. Since f is injective then it is also surjective on the finite set.
Consider the two lines

$$\begin{cases} (1) \quad a_1 \circ x + b_1 \circ y + c_1 \circ z = 0 \\ (2) \quad a_2 \circ x + b_2 \circ y + c_2 \circ z = 0 \end{cases}.$$

1. If $z = 0$, the equations become $a_1 \circ x = -b_1 \circ y$ and $a_2 \circ x = -b_2 \circ y$. If $y = 0$, then $x = 0$, which is not a valid point. Thus $y \neq 0$, and we can write $X = x \circ y^{-1}$. This yields a solution if and only if the lines share the same "slope" at infinity. Since the lines are distinct, they intersect at a unique point at infinity only if they are "parallel" in the affine sense.

2. If $z \neq 0$, we normalize to $z = 1$. The system is:

$$\begin{cases} a_1 \circ x + b_1 \circ y = -c_1 \\ a_2 \circ x + b_2 \circ y = -c_2 \end{cases}$$

We consider two subcases:

(a) Vertical Line: If $b_1 = 0$, then $a_1 \neq 0$ (otherwise the triple is all zeros). The first equation becomes $a_1 \circ x = -c_1$. Since the map $t \mapsto a_1 \circ t$ is bijective, there is a unique solution for x. We substitute this unique x into the second equation. If $b_2 \neq 0$, we solve uniquely for y. If $b_2 = 0$, the lines are parallel vertical lines, which intersect at the unique point at infinity found in the $z = 0$ case.

(b) Non-Vertical Lines: Assume $b_1, b_2 \neq 0$. Any line with $b \neq 0$ can be rewritten in the slope-intercept form $y = m \circ x + k$. We rewrite our system as:

$$\begin{cases} y = m_1 \circ x + k_1 \\ y = m_2 \circ x + k_2 \end{cases}$$

Equating the expressions for y:

$$m_1 \circ x + k_1 = m_2 \circ x + k_2 \implies m_1 \circ x - m_2 \circ x = k_2 - k_1$$

Using the right distributivity property, we get:

$$(m_1 - m_2) \circ x = k_2 - k_1$$

Let $M = m_1 - m_2$ and $K = k_2 - k_1$. Since the lines are distinct and non-vertical, their slopes are distinct, so $M \neq 0$.

We previously established that the map $f(x) = M \circ x$ is bijective for $M \neq 0$. Therefore, there exists a unique solution $x = f^{-1}(K)$. substituting this x back into either line equation yields a unique y.

Thus, in all cases, two distinct lines intersect at exactly one unique point.