Probabilistic Methods in Combinatorics

Solutions to Assignment 8

Problem 1. Let n be an integer. Show that, with probability $1-o(1)$, in $G(n,1/2)$, all vertices have degree in the range $[n/2-2\sqrt{n\log n}, n/2+2\sqrt{n\log n}]$.

Solution. Let G be sampled as $G(n,1/2)$. By union bound, it suffices to show that for every vertex $v \in V(G)$, we have with probability $1-o(n^{-1})$ that

$$\deg(v) \in [n/2-2\sqrt{n\log n}, n/2+2\sqrt{n\log n}].$$

It is easy to see that $\deg(v)$ has the distribution of a Binomial random variable of size $n-1$ and probability $p=1/2$. Therefore, by Chernoff (Corollary 5.2 of the lecture notes), we have that

$$\begin{aligned}\mathbb{P}\left(\left|\deg(v)-\frac{n}{2}\right| \geq 2\sqrt{n\log n}\right) \leq \mathbb{P}\left(\left|\deg(v)-\frac{n-1}{2}\right| \geq \sqrt{n\log n}\right) &\leq 2e^{-2(\sqrt{n\log n})^2/(n-1)}\\ &\leq 2e^{-1.5\log n}\\ &= o(n^{-1}).\end{aligned}$$

Problem 2. You are presented with an $n\times n$ grid where each cell in the grid is either red or blue. You can now do the following operation as many times as you like:

Select a row/column and switch the color of each cell in that row/column.

Your goal is to maximize the number of red cells in your grid. Prove that there exists an initial configuration of the grid such that using the operation above arbitrarily many times you cannot turn more than $\frac{1}{2}+\sqrt{\frac{\ln 2}{n}}$ fraction of the cells red.

Solution. Colour each cell independently at random red or blue. We prove that with positive probability there exists no sequence of operations that turn more than $\frac{1}{2}+\sqrt{\frac{\ln 2}{n}}$ fraction of the cells red.

Observe that if we have a sequence of row and column operations, the outcome does not depend on the order of the switches. Moreover, if we have a sequence of row and column switches which involves switching some row (or column) R at least twice, that has the same effect as the sequence of switches which omits two such switches of R. Hence, we may restrict our attention to those sequences of operations which switch every row and column at most once.

The number of such sequences is precisely 2^{2n} since we are free to decide, for each row and column whether we switch it or not (and the order does not matter).

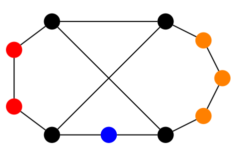
Fix a sequence S of operations. What is the probability that after performing S, the resulting grid has more than $\frac{1}{2}+\sqrt{\frac{\ln 2}{n}}$ fraction of the cells red? Note that this has the same probability as having more than $\left(\frac{1}{2}+\sqrt{\frac{\ln 2}{n}}\right)n^2$ red cells in the original grid as the distributions of the original grid and the new grid are the same. We can estimate the probability of this event by the Chernoff bound. Indeed, if we view red cells as having value $+1$ and blue cells as having value -1, then the probability of having more than $\left(\frac{1}{2}+\sqrt{\frac{\ln 2}{n}}\right)n^2$ red cells is the same as the probability of having $\sum_{1\le i\le n^2} X_i > 2\sqrt{\frac{\ln 2}{n}}\cdot n^2$ where $X_1, X_2, \dots, X_{n^2}$ are i.i.d. random variables taking values $+1$ and -1 with probability $1/2$. By Chernoff's inequality, this probability is less than $\exp\left(-\frac{(2\sqrt{\frac{\ln 2}{n}}n^2)^2}{2n^2}\right)$.

Since

$$\exp\left(-\frac{(2\sqrt{\frac{\ln 2}{n}}n^2)^2}{2n^2}\right) = \exp(-2(\ln 2)n) = 2^{-2n},$$

the union bound implies that the probability that there exists some sequence of operations which turns the original configuration into one with more than $\left(\frac{1}{2}+\sqrt{\frac{\ln 2}{n}}\right)n^2$ red cells is less than 1.

Problem 3. We say that a graph H is a *subdivision of* K_k (the complete graph on k vertices) if H can be obtained from K_k by replacing each of its $\binom{k}{2}$ edges by inner vertex-disjoint paths (possibly of length 1). For example, the graph below is a subdivision of K_4:



In 1961, György Hajós conjectured that for any $k \in \mathbb{N}$ any graph with chromatic number k contains a subdivision of K_k. This conjecture was disproved by Catlin in 1979, who found counterexamples for $k \ge 7$. The goal of this exercise is to show that with probability $1-o(1)$ the random graph $G(n, 1/2)$ is a counterexample to Hajós' conjecture.

(a) Show that with probability $1-o(1)$ the random graph $G(n, 1/2)$ has chromatic number at least $n/(10\log_2 n)$.

(b) Show that with probability $1-o(1)$, for every set of $m \geq 100 \ln n$ vertices, out of the $\binom{m}{2}$ possible edges at least $\frac{1}{3}\binom{m}{2}$ are missing.

(c) Use (b) to show that with probability $1-o(1)$ the random graph $G(n,1/2)$ does not contain a subdivision of K_k for $k \geq 10\sqrt{n}$.

Solution.

(a) Let χ denote the chromatic number of $G(n,1/2)$ and let α denote the size of its largest independence set. Since $G(n,1/2)$ admits a partition of its vertex set into χ many independent sets, each of which of size at most α, it follows that $\chi \geq n/\alpha$. Therefore, it suffices to show that, w.h.p. (= with high probability, i.e. with probability $1-o(1)$), we have $\alpha \leq 10\log_2 n$. Let S be a set of vertices of size $l := \lceil 10\log_2 n \rceil$. We start by estimating the probability that in $G(n,1/2)$ the set S is an independent set. Clearly this probability is $2^{-\binom{l}{2}}$. Therefore, by the union bound, the probability that there exists a set of vertices S of size l which is an independent set in $G(n,1/2)$ is at most

$$\binom{n}{l}2^{-\binom{l}{2}} \leq \left(\frac{en}{l}\right)^l 2^{-l(l-1)/2} = \left(\frac{en}{l\cdot 2^{(l-1)/2}}\right)^l = \left(\frac{\sqrt{2}en}{l\cdot 2^{l/2}}\right)^l$$
$$\leq \left(\frac{\sqrt{2}en}{10\log_2 n \cdot 2^{5\log_2 n}}\right)^l = \left(\frac{\sqrt{2}e}{10\log_2 n \cdot n^4}\right)^l = o(1).$$

We conclude then that w.h.p. we have $\alpha < l$ in which case one has $\chi \geq n/l \geq \frac{n}{10\log_2 n}$.

(b) Let S be a set of $m \geq 100\log n$ vertices. Let X_S denote the number of edges of $G(n,1/2)$ between vertices in S. Note that X_S is a random variable which is distributed as $\mathrm{Bin}(\binom{m}{2}, 1/2)$. Therefore, by the Chernoff bounds:

$$\mathbb{P}\left[X_S > \frac{2}{3}\binom{m}{2}\right] = \mathbb{P}\left[X_S - \frac{1}{2}\binom{m}{2} > \frac{1}{6}\binom{m}{2}\right] < e^{-2\left(\frac{1}{6}\right)^2\binom{m}{2}} = e^{-m(m-1)/36}.$$

Note that in any graph, if S is a set of m vertices and T is a randomly chosen subset of S with t vertices then

$$\mathbb{E}[e(T)] = \sum_{e\in E(S)} \mathbb{P}[e \in E(T)] = e(S)\frac{\binom{m-2}{t-2}}{\binom{m}{t}} = \frac{e(S)}{\binom{m}{2}} \cdot \binom{t}{2}.$$

Therefore, if there is a set S of $m \geq 100\log n$ vertices which spans in $G(n,1/2)$ more than $\frac{2}{3}\binom{m}{2}$ edges then there is also a subset T of exactly $l := \lceil 100\log n \rceil$ vertices which spans more than $\frac{2}{3}\binom{l}{2}$ many edges. Thus, by a union bound the probability that there exists in $G(n, 1/2)$ a set of $m \geq l$ vertices which spans more than $\frac{2}{3}\binom{m}{2}$ edges is at most

$$\binom{n}{l} e^{-l(l-1)/36} \leq \left(\frac{en}{l} \cdot e^{-(l-1)/36}\right)^l \leq \left(\frac{en}{100 \log n} \cdot e^{-2 \log n}\right)^l = \left(\frac{e}{100 n \log n}\right)^l = o(1).$$

This finishes the proof of (b).

(c) Assume that in all sets of $k \geq 10\sqrt{n}$ vertices, at least $\frac{1}{3}\binom{k}{2}$ of the possible $\binom{k}{2}$ edges are missing in $G(n, 1/2)$. Note that by (b) this happens w.h.p. Suppose in this case that $G(n, 1/2)$ contains a subdivision of K_k, where $k \geq 10\sqrt{n}$. We know that out of all the possible $\binom{k}{2}$ pairs of branch vertices of this subdivision of K_k (namely vertices that would correspond to a K_k before replacing its edges) at least $\frac{1}{3}\binom{k}{2}$ are not edges in $G(n, 1/2)$. Therefore, between such pairs of branch vertices, we must have a path of length at least 2. Moreover, all these paths are inner vertex-disjoint. Thus, we conclude that this subdivison of K_k has at least $k + \frac{1}{3}\binom{k}{2}$ many vertices. However, since $k \geq 10\sqrt{n}$, this number is strictly larger than n, a contradiction. It follows that w.h.p. $G(n, 1/2)$ does not contains a subdivision of K_k.

Remark. Note that this shows that there are graphs with chromatic number at least $\frac{n}{10 \log_2 n}$ that do not contain a subdivision of $K_{\lceil 10\sqrt{n} \rceil}$. In particular, Hajós' conjecture is very far from being true.

Problem 4*. Prove that the following holds for all large enough n. Let $S_1, \ldots, S_k$ be subsets of $[n] := \{1, \ldots, n\}$. If $k \leq 1.99 \frac{n}{\log_2 n}$ then there are two distinct subsets X, Y of $[n]$ such that $|X \cap S_i| = |Y \cap S_i|$ for every $1 \leq i \leq k$.

Solution. As usual, we let X be a random subset of $[n]$. Let $a = \sqrt{n \log n}$. Let $i \leq k$. By the Chernoff bound,

$$\mathbb{P}\left[\left||X \cap S_i| - \frac{1}{2}|S_i|\right| > a\right] \leq 2e^{-2a^2/|S_i|} \leq 2e^{-2a^2/n} = 2e^{-2 \log n} = \frac{2}{n^2}.$$

Hence, by the union bound,

$$\mathbb{P}\left[\left||X \cap S_i| - \frac{1}{2}|S_i|\right| > a \text{ for some } i \leq k\right] \leq \frac{2k}{n^2} \leq \frac{1}{2}. \quad (1)$$

Let $\mathcal{F}$ be the family of sets X for which $\left||X \cap S_i| - \frac{1}{2}|S_i|\right| \leq a$ for every $i \leq k$. Then, by (1), $|\mathcal{F}| \geq 2^{n-1}$. For $X \in \mathcal{F}$, let $v_X = (|X \cap S_1|, \ldots, |X \cap S_k|)$. Note that the set $\{|X \cap S_i| : X \in \mathcal{F}\}$ consists of at most $2a + 1$ values. It follows that the set $\{v_X : X \in \mathcal{F}\}$

has size at most $(2a+1)^k$. Thus, if $(2a+1)^k < 2^{n-1}$ then, by the pigeonhole principle, there are two distinct sets $X, Y \in \mathcal{F}$ for which $v_X = v_Y$, which means that $|X \cap S_i| = |Y \cap S_i|$ for every $i \leq k$, as required. So let us prove the inequality $(2a+1)^k < 2^{n-1}$. In fact, we shall prove that $(5a)^k \leq 2^n$. Since $(2a+1)^k < (5a)^k/2$, the desired inequality would follow. Note that

$$(5a)^k \leq 2^n \quad \Leftrightarrow \quad a \leq \frac{1}{5} \cdot 2^{n/k}.$$

Now, we use the assumption that $k \leq 1.99 \frac{n}{\log_2 n}$ to conclude that

$$\frac{1}{5} \cdot 2^{n/k} \geq \frac{1}{5} \cdot 2^{\log_2 n/1.99} = \frac{1}{5} \cdot n^{1/1.99} \geq \frac{1}{5} \cdot n^{0.502} \geq \sqrt{n \log n} = a.$$