Problem Set 2

Out: 27 Sept.; Due: 7 Oct.

Notes: *Solutions are due by* **5pm on Friday October 7**. *Please submit your solutions on Gradescope by that time;* **remember to begin each problem (not problem part) on a new page.** *Late solutions will not be accepted. Solutions should preferably be typeset in LaTeX; if this poses a problem then they may be written* **neatly** *by hand and scanned. Depending on grading resources, we reserve the right to grade only a subset of the problems and check off the rest (but since you don't know which subset, you're strongly advised to do all of them!).*

Take time to write **clear** *and* **concise** *answers. None of the problems require a long solution (often not much longer than the problem statement itself!); if you find yourself writing a lot, you are either on the wrong track, confused, or giving too much detail. You are actively encouraged to form small groups (two or three people) to work through the problems, but you* **must always** *write up your solutions on your own. If you use external sources, you should cite them; and again, you must understand and formulate your solutions yourself.*

1. Probabilistic method for dominating sets

Let $G = (V, E)$ be an undirected graph, and $U \subseteq V$ any subset of its vertices. A set $D \subseteq V$ is a *dominating set* for U if for every vertex $u \in U$, either u or some neighbor of u belongs to D.

Let d be an integer, and let U be the set of all vertices of G whose degree is at least d. Use the probabilistic method to show that there exists a dominating set for U of size at most $\lfloor n \cdot \frac{\log(d+1)+1}{d+1} \rfloor$, where n is the number of vertices of G. [HINT: Start by picking a random subset S of vertices, each with probability p (a value to be chosen later). Then modify S to make it a dominating set.]

2. Locally 2-colorable graphs

Recall that a graph is *2-colorable* if we can assign colors red and green to each vertex such that the endpoints of every edge are assigned different colors. (Equivalently, the graph is bipartite.) Suppose we are told that a graph $G = (V, E)$ is "locally 2-colorable", in the sense that the induced subgraph[1] on *every* subset of $O(\log n)$ vertices is 2-colorable. Does this imply that G itself is 2-colorable? In this problem we will see that the answer is spectacularly "no": namely, we will show that there exists a graph that is locally 2-colorable but is "very far away" from being 2-colorable, in the sense that we would have to remove a constant fraction of its edges in order to make it 2-colorable. We will prove the existence of this graph using the probabilistic method.

Throughout, assume $n \geq 16$, set $p = 16/n$, and let G be a random graph from the model $\mathcal{G}_{n,p}$. The probabilities and expectations refer to the experiment of picking G at random.

(a) Write down the expected number of edges in G, and use Chebyshev's inequality to show that, with probability at least $3/4$, G has at most $10(n-1)$ edges.

(b) Now fix an arbitrary assignment of colors to the vertices. Show that the expected number of violated edges (i.e., edges with endpoints of the same color) in G is at least $4(n-2)$. Hence show that the probability there are more than $n-2$ violated edges is at least $1 - e^{-9(n-2)/8}$. [HINT: You may assume for simplicity that n is even. Chebyshev's inequality is not strong enough to get this exponential tail bound, so you will need the following Chernoff bound for the sum X of n iid 0-1-valued random variables: $\Pr[X \leq (1-\delta)\mathrm{E}X] \leq \exp(-\frac{1}{2}\delta^2 \mathrm{E}X)$.]

(c) Deduce from part (b) that, with probability at least $3/4$, G is not 2-colorable even if we delete any $n-2$ of its edges.

[1] The *induced* subgraph on a subset of vertices $V' \subseteq V$ is the graph with vertex set V' and edge set consisting of all the edges of G both of whose endpoints are in V'.

(d) Show that the expected number of cycles of length exactly k in G is at most 16^k. Deduce that the expected number of cycles of length at most $\frac{1}{8}\log n$ is at most $2\sqrt{n}$. (The log is base 2.)

(e) Use part (d) to deduce that, with probability at least $3/4$, by deleting only $8\sqrt{n}$ (suitably chosen) edges of G we can obtain a graph such that the induced subgraph on any subset of $\frac{1}{8}\log n$ vertices is 2-colorable.

(f) Put parts (a), (c) and (e) together to deduce that, for every sufficiently large n, there exists a graph $G = G_n$ on n vertices such that:

— The induced subgraph on any subset of $\frac{1}{8}\log n$ vertices of G_n is 2-colorable; and

— G_n is not 2-colorable, and remains not 2-colorable even after deleting any $1/20$ fraction of its edges.

3. A threshold for isolated vertices

A vertex of a graph is said to be "isolated" if no edges are incident at it. For a graph G, let $\mathcal{Q}$ be the property that G contains no isolated vertices.

(a) Show using the second moment method that the function $p_0(n) = \frac{\ln n}{n}$ is a threshold for $\mathcal{Q}$ in a random graph $G \in \mathcal{G}_{n,p}$. [Hint: Let the r.v. X denote the number of isolated vertices in G. Compute the expectation and variance of X.]

(b) Now consider $p(n) = \frac{c\ln n}{n}$ for constant c. What can you say about $\mathcal{Q}$ in random graphs $G \in \mathcal{G}_{n,p}$ when $c < 1$ and when $c > 1$?

[NOTE: It can be shown that, if $p(n) = \frac{\ln n}{n} + \frac{\alpha}{n}$ where α is a constant, the number of isolated vertices has asymptotically a Poisson distribution with mean $\lambda = \mathrm{e}^{-\alpha}$. I.e., for each fixed k, $\Pr[X = k] \to \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}$ as $n \to \infty$, where X is the number of isolated vertices. Thus the threshold in this case has width $\frac{1}{n}$.]

4. Planted cliques and cryptography

Let G be a random graph in the $\mathcal{G}_{n,1/2}$ model. Recall that the problem of finding a clique in G of size $(1+\epsilon)\log_2 n$ is widely conjectured to be intractable, for all but a tiny fraction of G's. (This is despite the fact that almost all G's contain a clique of size close to $2\log_2 n$.) In this problem we will demonstrate that, if this conjecture holds, then it holds even when we have "planted" a large clique, say of size $\frac{3}{2}\log_2 n$, in G. This suggests a novel cryptographic signature scheme: I generate a random graph G, plant a large clique in it, and release it. The above result means that nobody else can find the clique. However, at any time I can prove that the graph is mine by revealing the clique.

To make all this precise, let's define the model $\mathcal{G}'_{n,1/2}$ as follows. First, select $G \in \mathcal{G}_{n,1/2}$. Then pick a subset of $k = \frac{3}{2}\log_2 n$ vertices u.a.r. and add edges to G to make this set a clique. (As usual, we forget about such niceties as rounding k to an integer.) We shall write $\Pr$ and $\Pr'$ to denote probabilities in the $\mathcal{G}_{n,1/2}$ model and the $\mathcal{G}'_{n,1/2}$ model respectively, and similarly for expectations.

For any n-vertex graph G, let $f(G)$ denote the number of k-cliques in G, and let $\mu = \mathrm{E}(f)$ denote the expectation of $f(G)$ in the $\mathcal{G}_{n,1/2}$ model, i.e., $\mu = \binom{n}{k}2^{-\binom{k}{2}}$. Recall from Lecture 7 that, in the $\mathcal{G}_{n,1/2}$ model,

$$\frac{\mathrm{E}(f^2)}{\mu^2} = \sum_{i=0}^{k} \binom{n}{k}^{-1}\binom{k}{i}\binom{n-k}{k-i}2^{\binom{i}{2}} \equiv \sum_{i=0}^{k} f_i.$$

A straightforward but slightly tedious calculation shows that each term f_i is bounded above by an^c for constants a, c. This then implies that $\frac{\mathrm{E}(f^2)}{\mu^2} = \mathrm{O}(n^c \log n)$.

(a) Show that, for any G, $\Pr'[G] = \frac{f(G)}{\mu}\Pr[G]$.

(b) For $\alpha > 0$, call a graph G "α-bad" if $f(G) > n^\alpha \mu$ (i.e., the number of k-cliques exceeds its mean by a factor of more than n^α). Deduce from the above that, for arbitrarily small $\epsilon > 0$ (and sufficiently large n),

$$\Pr[G \text{ is } \alpha\text{-bad}] \leq n^{-2\alpha+c+\epsilon}.$$

(c) Extend part (b) to the $\mathcal{G}'_{n,1/2}$ model by showing that

$$\Pr'[G \text{ is } \alpha\text{-bad}] = \mathrm{O}(n^{-\alpha/2+c+\epsilon}).$$

[HINT: Let $\mathcal{B}$ denote the set of α-bad graphs. Partition $\mathcal{B}$ as $\bigcup_{j=2}^{\infty} \mathcal{B}_j$, where $\mathcal{B}_j = \{G : n^{j\alpha/2}\mu < f(G) \leq n^{(j+1)\alpha/2}\mu\}$. Bound $\Pr[G \in \mathcal{B}_j]$ as in part (b), and then use part (a) to translate to a bound on $\Pr'[G \in \mathcal{B}_j]$. Finally, sum over j.]

(d) Now let $\mathcal{E}$ be an event that holds with probability $\Omega(n^{-r})$ in $\mathcal{G}'_{n,1/2}$, where $r > 0$ is some constant. Deduce from parts (c) and (a) that $\mathcal{E}$ holds in $\mathcal{G}_{n,1/2}$ with probability $\Omega(n^{-r'})$, for some other constant r' (which depends on r). [HINT: Choose α so that $\Pr'[G \text{ is } \alpha\text{-bad}] \ll n^{-r}$.]

(e) Finally, explain how the result of part (d) justifies the cryptographic signature scheme described at the beginning of the problem.

5. More on Unbalancing Lights

Let $X_1, X_2, \ldots$ be independent unbiased ± 1 coin tosses, and define $S_n = \sum_{i=1}^n X_i$. In Lecture 5 we used the fact that

$$\mathrm{E}(|S_n|) = \left(\sqrt{\frac{2}{\pi}} + o(1)\right)\sqrt{n}$$

(which follows from the Central Limit Theorem) to obtain a randomized algorithm that achieves an asymptotic expected excess of $\sqrt{\frac{2}{\pi}}\, n^{3/2}$ "on" lights over "off" lights. In this problem we consider how to derandomize this algorithm (with a slightly worse performance guarantee) using k-wise independent random variables.

(a) In preparation for later, for any $z \geq 0$ and $\alpha > 0$, show that

$$z \geq \frac{3\sqrt{3}}{2\sqrt{\alpha}}\left(z^2 - \frac{z^4}{\alpha}\right).$$

Hence, by judicious choice of α, deduce that for any random variable Z,

$$\mathrm{E}(|Z|) \geq \frac{\mathrm{E}(Z^2)^{3/2}}{\mathrm{E}(Z^4)^{1/2}}.$$

(b) Now suppose $S_n = \sum_{i=1}^n X_i$, where the X_i are unbiased ± 1 coin tosses as before, *but are only 4-wise independent*. Show using part (a) that

$$\mathrm{E}(|S_n|) \geq \sqrt{\frac{n}{3}}.$$

(c) Use part (b) to obtain a polynomial time *deterministic* algorithm that achieves an asymptotic excess of $cn^{3/2}$ "on" lights over "off" lights, for a constant c (which you should specify) that is slightly worse than $\sqrt{\frac{2}{\pi}}$.

(d) Show that 4-wise independence is really necessary here by exhibiting a 3-wise independent family of fair coin flips $\{X_i\}$ for which $\mathrm{E}(|S_n|)$ is very small. Here is one approach. Suppose $n = 2^k$, and let $v = (v_1, \ldots, v_{k+1})$ be a uniform random vector in $\{0,1\}^{k+1}$. For each integer $i \in \{0, 1, \ldots, n-1\}$, let $b_i = (i_1, i_2, \ldots, i_k, 1)$ be the binary expansion of i with a 1 appended. Define $Y_i = b_{i-1} \cdot v$ (dot product mod 2), and $X_i = (-1)^{Y_i}$. Show that the Y_i (and hence the X_i) are 3-wise independent, and that $\mathrm{E}(|S_n|) = 1$.