Homework #2 CSE 525, Fall 2016
Out: 10-Oct
Due: Friday, 21-Oct, midnight

1. Show that there is a positive number c such that the following holds. For any n real numbers $a_1, a_2, \ldots, a_n$ that satisfy $a_1^2 + a_2^2 + \cdots + a_n^2 = 1$, if $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are all chosen independently and uniformly at random from the set $\{-1, 1\}$, then

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} \varepsilon_i a_i\right| \leqslant 1\right] \geqslant c\,.$$

[Hint: Look at the square...]

2. Consider the random graph model $\mathcal{G}_{n,\frac{1}{2}}$. This is a random undirected graph with n vertices, where every potential edge $\{u, v\}$ is in the graph independently with probability 1/2.

We are interested in finding a large clique in $G = (V, E) \sim \mathcal{G}_{n,\frac{1}{2}}$.

(a) Prove that the following greedy algorithm finds a clique of size $(1 - o(1)) \log_2 n$.

Algorithm: Start with $C_0 = \{v_0\}$ for some fixed vertex v_0. Now suppose we have constructed a clique C_k. If there is a vertex $v \in V \setminus C_k$ that is connected to *every* vertex in C_k, then we put $C_{k+1} = C_k \cup \{v\}$. Otherwise, we stop and output the set $C = C_k$. Prove that

$$\mathbb{P}[|C| \geqslant (1 - o(1)) \log_2 n] \geqslant 1 - o(1)\,. \tag{0.1}$$

Note that the algorithm is deterministic. The randomness here is over the choice of the graph $G \sim \mathcal{G}_{n,\frac{1}{2}}$. Also, $o(1)$ represents a function that goes to 0 as $n \to \infty$. For instance, if one proved that $\mathbb{P}[|C| \geqslant \log_2 n - 100\sqrt{\log_2 n}] \geqslant 1 - \frac{1}{n}$, it would prove (0.1).

(b) Show that the clique you found is not too much smaller than the largest clique: Let X_n be the random variable that denotes the size of the largest clique in $\mathcal{G}_{n,\frac{1}{2}}$. Show that

$$\mathbb{P}\left[X_n \leqslant (2 + o(1)) \log_2 n\right] \geqslant 1 - o(1)\,.$$

3. * Suppose you are designing the sensor deployment system for SpaceX's Mars exploration mission. In order to search for habitable land, you will deploy sensor packets from the atmosphere. Each packet is designed to disperse N sensors uniformly over one square kilometer. (So you can imagine that the N sensors are placed uniformly and independently at random in the unit square $[0, 1]^2$.)

The sensors will communicate with each other via a mesh network, but to keep energy costs down, it's important that every sensor communicates with a very small number of other sensors. There is a parameter k such that every sensor can send messages to the k closest sensors (in Euclidean distance; you may assume that the sensors are deployed on flat ground).

You can think of the resulting directed graph G: The vertices are sensors and there is an edge (u, v) between two sensors u and v if v is one of the k closest sensors to u. In order for the mesh network to be intact, it should be that G is strongly connected (in other words, between every pair of sensors, there should be a directed path in G).

(a) Prove that there is a constant c such that if $k \geqslant c \log N$, then

$$\mathbb{P}[G \text{ strongly connected}] \to 1 \quad \text{as} \quad N \to \infty$$

(b) Prove that there is a constant $c' > 0$ such that if $k < c' \log N$, then

$$\mathbb{P}[G \text{ strongly connected}] \to 0 \quad \text{as} \quad N \to \infty$$

(c) **Bonus:** Establish a sharper threshold. Prove that there is a constant c_0 such that if $k = \lceil c \log N \rceil$ and

$$c > c_0 \implies \mathbb{P}[G \text{ strongly connected}] \to 1 \quad \text{as} \quad N \to \infty$$
$$c < c_0 \implies \mathbb{P}[G \text{ strongly connected}] \to 0 \quad \text{as} \quad N \to \infty$$

4. **Bonus:** Suppose that p is an odd prime and $a_1, a_2, \ldots, a_{(p+1)/2}$ are distinct numbers modulo p. Prove that there exists an x such that the set $\{a_1 x, a_2 x, \ldots, a_{(p+1)/2} x\}$ has no gaps bigger than $5\sqrt{p}$. (If we consider the numbers as sitting on the vertices of a p-element cycle, then a "gap" is a consecutive sequence of vertices which contains no elements.)

   Hint: Use the probabilistic method applied to the set $\{a_i x + y\}$ where x, y are chosen uniformly at random modulo p.