Probabilistic Methods in Combinatorics

Instructor: Oliver Janzer

Solutions to Assignment 7

Problem 1. Show that $p_0 = n^{-1}$ is a weak threshold for the property that $G(n,p)$ contains K_3 (i.e. a triangle) as a subgraph.

Solution. We follow the same method as for K_4 in the course. Let X be the number of K_3's in $G(n,p)$, and for a given 3-element set S of vertices, let X_S be the indicator random variable of the event that S spans a clique; namely, $X_S = 1$ if S spans a clique, and $X_S = 0$ otherwise. Then $X = \sum_S X_S$, where the sum ranges through all 3-element subsets of the vertex set. Clearly, we have $\mathbb{E}(X_S) = p^3$, and so $\mathbb{E}(X) = \binom{n}{3}p^3$. Note that $G(n,p)$ contains a K_3 if and only if $X \geq 1$.

We verify that Items 1 and 2 in the definition of a (weak) threshold are satisfied. First, we show that Item 1 holds, so let $p < \frac{p_0}{C}$ with some constant C. By Markov's inequality, we have

$$f(p) = \mathbb{P}(X \geq 1) \leq \mathbb{E}(X) = \binom{n}{3}p^3 < n^3p^3 < n^3(p_0/C)^3 = C^{-3}.$$

Therefore, by choosing C sufficiently large, we have $f(p) < \varepsilon$, satisfying Item 1.

Now let $p > Cp_0 = Cn^{-2/3}$. We can write $f(p) = 1 - \mathbb{P}(X = 0)$. Now by Chebyshev, we bound the probability $\mathbb{P}(X = 0)$ as follows:

$$\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}(X))^2}.$$

Now our task is to calculate $\mathrm{Var}(X)$. We have

$$\mathrm{Var}(X) = \sum_{S,T} \mathrm{Cov}(X_S, X_T). \tag{1}$$

(Here, the sum is over all ordered pairs (S,T), including $S = T$.)

Now let us compute $\mathrm{Cov}(X_S, X_T)$. Note that

$$\mathrm{Cov}(X_S, X_T) = \mathbb{E}[X_SX_T] - \mathbb{E}[X_S]\mathbb{E}[X_T] = \mathbb{P}(A_S \cap A_T) - \mathbb{P}(A_S)\mathbb{P}(A_T),$$

where A_S is the event that S spans a clique. Since $\mathbb{P}(A_S) = p^3$ for every set S of size 3, we have $\mathbb{P}(A_S)\mathbb{P}(A_T) = p^6$ for all S, T. However, the term $\mathbb{P}(A_S \cap A_T)$ depends on the size of $S \cap T$. Indeed, $\mathbb{P}(A_S \cap A_T) = p^{g(S,T)}$, where $g(S,T)$ is the number of pairs of vertices contained entirely in S or entirely in T (since we need all these pairs to be edges in our random graph for the event $A_S \cap A_T$ to hold). Note that

$$g(S,T) = \begin{cases} 3 & \text{if } S = T \\ 5 & \text{if } |S \cap T| = 2 \\ 6 & \text{otherwise} \end{cases}$$

and therefore

$$\mathrm{Cov}(X_S, X_T) = \begin{cases} p^3 - p^6 & \text{if } S = T \\ p^5 - p^6 & \text{if } |S \cap T| = 2 \\ 0 & \text{otherwise.} \end{cases}$$

Also, for $k \in \{0,1,2,3\}$, the number of pairs (S,T) such that $|S \cap T| = k$ is less than n^{6-k} as $S \cup T$ occupies $6-k$ vertices. Therefore, the contribution of pairs (S,T) with $S = T$ is at most n^3p^3, the contribution of pairs (S,T) with $|S \cap T| = 2$ is at most n^4p^5, and the contribution of the other pairs is 0. Hence,

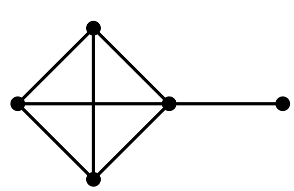
$$\mathrm{Var}(X) < n^3p^3 + n^4p^5.$$

Also, $\mathbb{E}(X) = \binom{n}{3}p^3 = \Omega(n^3p^3)$. So we get

$$\begin{aligned} P(X=0) \le \frac{\mathrm{Var}(X)}{(\mathbb{E}(X))^2} < \frac{n^3p^3 + n^4p^5}{\Omega(n^6p^6)} &= O(n^{-3}p^{-3} + n^{-2}p^{-1}) \\ &= O(C^{-3} + C^{-1}n^{-1}), \end{aligned}$$

where for the last inequality we used that $p > Cn^{-1}$. Therefore, if C is sufficiently large, then $P(X=0) \le \varepsilon$, so Item 2 is also satisfied. In conclusion, p_0 is truly a threshold.

Problem 2. What is a threshold probability function $p = p(n)$ for the occurrence of the graph below as a subgraph of the random graph $G(n,p)$?



Solution. Let X be the number of copies of the graph H (depicted above). We first calculate the expectation of X.

$$\mathbb{E}[X] = n(n-1)\binom{n}{3}p^7 = \Theta(n^5p^7).$$

Seeing this, one might be tempted to guess that $p_0 = n^{-5/7}$ is a threshold function for the appearance of H. However, let Y denote the number of copies of K_4. Then $\mathbb{E}[Y] = \binom{n}{4}p^6$. It follows that the probability that $G(n,p)$ contains a K_4 is $o(1)$ if $p = o(n^{-2/3})$. As K_4 is a subgraph of H, this implies that the probability that $G(n,p)$ contains a copy of H is $o(1)$ whenever $p = o(n^{-2/3})$, contrary to the guess that if $p = \omega(p_0)$ then $G(n,p)$ contains a copy of H w.h.p. (= with high probability, i.e. with probability $1 - o(1)$). We shall see that $n^{-2/3}$ is a threshold function for the property of containing a copy of H.

Let us calculate the variance of X. For that, list all copies of H in K_n as $H_1, \dots, H_l$, and let X_i be the indicator function of the event that H_i is in G. Then $X = \sum_{i\in[l]} X_i$.

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_{i,j\in[l]} \mathrm{Cov}(X_i, X_j) = \sum_{i,j\in[l]} \left(\mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]\right) \\
&= \sum_{i,j\in[l]} \left(\mathbb{P}[H_i \cup H_j \subseteq G] - \mathbb{P}[H_i \subseteq G]\cdot\mathbb{P}[H_j \subseteq G]\right) \\
&= \sum_{i,j\in[l]} \left(p^{14-|E(H_i)\cap E(H_j)|} - p^{14}\right) \\
&= p^{14}\cdot\sum_{k=0}^{7} \sum_{i,j\in[l]:\, |E(H_i)\cap E(H_j)|=k} \left(p^{-k} - 1\right) \\
&\le p^{14}\cdot\sum_{k=1}^{7} \sum_{i,j\in[l]:\, |E(H_i)\cap E(H_j)|=k} p^{-k} \\
&= p^{14}\cdot\sum_{k=1}^{7} p^{-k}\cdot \#\{(i,j) : |E(H_i)\cap E(H_j)| = k\},
\end{aligned}$$

where the inequality follows as the pairs (i,j) for which $k = 0$ contribute 0 to the sum, and for the rest we use the inequality $p^{-k} - 1 \le p^{-k}$. Now, we wish to bound $\#\{(i,j) : |E(H_i)\cap E(H_j)| = k\}$. Firstly, we note the following upper bounds on the number of vertices that are in either H_i or H_j, provided $|E(H_i)\cap E(H_j)| = k$.

$$|V(H_i)\cup V(H_j)| \le \begin{cases} 8 & k = 1 \\ 7 & k \in \{2,3\} \\ 6 & k \in \{4,5,6\} \\ 5 & k = 7 \end{cases}$$

It follows that

$$\#\{(i,j) : |E(H_i) \cap E(H_j)| = k\} \leq \begin{cases} n^8 & k = 1 \\ n^7 & k \in \{2,3\} \\ n^6 & k \in \{4,5,6\} \\ n^5 & k = 7 \end{cases}$$

We obtain the following upper bound on $\mathrm{Var}[X]$.

$$\mathrm{Var}[X] \leq p^{14} \left(n^8 p^{-1} + 2n^7 p^{-3} + 3n^6 p^{-6} + n^5 p^{-7}\right)$$

Let $p = w(n^{-2/3})$. Now

$$\begin{aligned} \mathbb{P}[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2} &\leq O\left(\frac{p^{14}\left(n^8 p^{-1} + 2n^7 p^{-3} + 3n^6 p^{-6} + n^5 p^{-7}\right)}{n^{10} p^{14}}\right) \\ &\leq o\left(n^{-2} n^{2/3} + 2n^{-3} n^2 + 3n^{-4} n^4 + n^{-5} n^{14/3}\right) = o(1). \end{aligned}$$

(Note that the largest term corresponds to H_i and H_j intersecting in a K_4, which hints to the fact that K_4's are important.) So we have shown that if $p = o(n^{-2/3})$ then w.h.p. $G(n,p)$ does not contain a copy of H, and if $p = w(n^{-2/3})$ then w.h.p. $G(n,p)$ does contain a copy of H. This shows that $n^{-2/3}$ is a threshold function for the property of containing a copy of H.

Remark. We could have taken a simpler approach as follows. Instead of showing directly that a copy of H exists w.h.p. when $p = \omega(n^{-2/3})$, we could have shown that $G(n,p)$ contains a copy of K_4 w.h.p. (this would involve the calculation of expectation and variance of the number of copies of K_4 and is similar to, but probably easier than, the above), and moreover we could have shown that w.h.p. every vertex has degree at least 4 (this can be done by upper-bounding the probability that a vertex has too small degree and using a union bound). Finally, we notice that if a graph contains a K_4 and every vertex has degree at least 4, then the graph contains a copy of H.

Problem 3. For $n \geq 2$ and $p \in (0,1)$ consider the random 5-partite graph $G(n,p,5)$ defined as follows. The vertex set of $G(n,p,5)$ is a union of five disjoint independent sets $V_1, \ldots, V_5$, each of size n. Moreover, for $1 \leq i < j \leq 5$, each $(v_i, v_j) \in V_i \times V_j$ is an edge in $G(n,p,5)$ independently with probability p. Find a threshold probability function $p = p(n)$ for the occurence of K_5 as a subgraph of $G(n,p,5)$.

Solution. A valid threshold probability is $p = n^{-1/2}$.

(a) Let X be the random variable counting the number of copies of K_5 in $G(n,p,5)$. Note that $\mathbb{E}[X] = n^5 p^{\binom{5}{2}}$ since there are n^5 potential copies of K_5 and each of them occurs with probability $p^{\binom{5}{2}}$. Therefore, if $p = o(n^{-1/2})$ then

$$\mathbb{E}[X] = n^5 p^{\binom{5}{2}} = o\left(n^5 n^{-\frac{1}{2}\cdot\binom{5}{2}}\right) = o(n^5 n^{-5}) = o(1)$$

and so, by Markov's inequality

$$\mathbb{P}\left[G(n,p,5) \text{ contains a copy of } K_r\right] = \mathbb{P}[X \geq 1] \leq \mathbb{E}[X] = o(1).$$

(b) By the second moment method we know that:

$$\mathbb{P}(G(n,p,5) \text{ contains a copy of } K_5) = \mathbb{P}(X \geq 1) = 1 - \mathbb{P}[X = 0] \geq 1 - \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2}.$$

Therefore, it suffices to show that $\frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2} = o(1)$. Let

$$\mathcal{S} = \{S \subseteq \cup_{i\in[5]} V_i : |S \cap V_i| = 1 \text{ for each } i \in [5]\}.$$

For each $S \in \mathcal{S}$ let X_S be the indicator random variable for the event A_S that S spans a K_5 in $G(n,p,5)$. Clearly $X = \sum_{S\in\mathcal{S}} X_S$. Therefore:

$$\mathrm{Var}[X] = \sum_{S,T\in\mathcal{S}} \mathrm{Cov}(X_S, X_T).$$

If $|S \cap T| \leq 1$ then the events A_S and A_T are independent (because no edges are shared) in which case $\mathrm{Cov}(X_S, X_T) = 0$. Hence:

$$\mathrm{Var}[X] \leq \sum_{\substack{S,T\in\mathcal{S} \\ 2\leq|S\cap T|\leq 5}} \mathrm{Cov}(X_S, X_T) \leq \sum_{\substack{S,T\in\mathcal{S} \\ 2\leq|S\cap T|\leq 5}} \mathbb{E}[X_S X_T].$$

Note that if $|S \cap T| = i$ then $\mathbb{E}[X_S X_T] = p^{2\binom{5}{2}-\binom{i}{2}} = p^{20-\binom{i}{2}}$. Therefore:

$$\sum_{\substack{S,T\in\mathcal{S} \\ 2\leq|S\cap T|\leq 5}} \mathbb{E}[X_S X_T] \leq \sum_{i=2}^{5} \binom{5}{i} n^5 (n-1)^{5-i} p^{20-\binom{i}{2}} \leq \left(n^5 p^{10}\right)^2 \sum_{i=2}^{5} \binom{5}{i} n^{-i} p^{-\binom{i}{2}}$$

since there are n^5 ways to choose $S \in \mathcal{S}$, $\binom{5}{i}$ ways to choose i parts on which S and T coincide and given these there are $(n-1)^{5-i}$ ways to choose the vertices in $T \setminus S$. So

$$\mathrm{Var}[X] \leq (n^5 p^{10})^2 \cdot O(n^{-2}p^{-1} + n^{-3}p^{-3} + n^{-4}p^{-6} + n^{-5}p^{-10}).$$

Since $\mathbb{E}[X] = n^5 p^{10}$, we have

$$\mathbb{P}(X = 0) \le \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} \le O(n^{-2}p^{-1} + n^{-3}p^{-3} + n^{-4}p^{-6} + n^{-5}p^{-10}),$$

which is $o(1)$ for $p = \omega(n^{-1/2})$.

Problem 4. Show there is a positive constant c such that the following holds. For any n reals $a_1, \dots, a_n$ satisfying $\sum_{i=1}^n a_i^2 = 1$, if $(\varepsilon_1, \dots, \varepsilon_n)$ is a $\{-1, 1\}$-random vector obtained by choosing each ε_i randomly and independently with equal probability to be either -1 or 1, then

$$\Pr\left[\Big|\sum_{i=1}^n \varepsilon_i a_i\Big| \le 1\right] \ge c.$$

Solution. Without loss of generality, assume that $|a_1| \ge |a_2| \ge \cdots \ge |a_n|$. Let $X = \sum_{i=1}^n \varepsilon_i a_i$. We will make use of the following fact for random variables defined similarly as X. To illustrate, we work with X. For any value $x \ne 0$, we have

$$\Pr[X < 0 \,|\, |X| = x] = \Pr[X > 0 \,|\, |X| = x] = 1/2.$$

Indeed, consider a sequence $(\varepsilon_1, \dots, \varepsilon_n)$ such that $|X| = x$ and $X < 0$. Then, for the negated sequence $(-\varepsilon_1, \dots, -\varepsilon_n)$ we have $|X| = x$ and $X > 0$. Hence, we obtain a bijection between sequences of ε's for which $|X| = x$ and $X < 0$ and those for which $|X| = x$ and $X > 0$, implying the claim.

By the law of total probability, we then have

$$\Pr\left[X < 0 \,|\, |X| \le 1\right] \le 1/2.$$

The reason we do not necessarily have equality is that because it could be that $|X| = 0$.

Now, we proceed to the solution of the problem. We consider two cases:

- $|a_1| \ge \frac{1}{10}$.

  Let $Z = \sum_{i=2}^n \varepsilon_i a_i$. Since the random variables $(\varepsilon_i a_i), i \in [n]$ are independent, we have $\mathrm{Var}(Z) = \sum_{i=2}^n \mathrm{Var}(\varepsilon_i a_i) = \sum_{i=2}^n a_i^2 = 1 - a_1^2 \le 99/100$. As $\mathbb{E}[Z] = 0$, by Chebyshev's inequality $\Pr[|Z| \le 1] \ge 1 - \mathrm{Var} Z \ge 1/100$. As ε_1 and Z are independent, by the above discussion, we have

  $$\Pr\left[\varepsilon_1 \cdot Z \le 0 \,|\, |Z| \le 1\right] \ge 1/2.$$

  Note that if $|Z| \le 1$ and $\varepsilon_1 \cdot Z \le 0$, then $|X| \le \max\{|Z| - |a_1|, |a_1|\} \le 1$. We conclude that

$$\Pr[|X| \le 1] \ge \Pr[|Z| \le 1] \cdot \Pr[\varepsilon_1 \cdot Z \le 0 \,|\, |Z| \le 1] \ge 1/200.$$

- $|a_1| < \frac{1}{10}$.

  Let k be such that $\sum_{i=1}^{k} a_i^2 \ge 1/2$ and $\sum_{i=1}^{k-1} a_i^2 < 1/2$. We denote $Y = \sum_{i=1}^{k} \varepsilon_i a_i$ and $Z = \sum_{i=k+1}^{n} \varepsilon_i a_i$. Similarly as before, we have $\mathrm{Var}(Y) = \sum_{i=1}^{k} a_i^2 \le 1/2 + a_k^2 \le 1/2 + 1/100$ and $\mathrm{Var}(Z) = \sum_{i=k+1}^{n} a_i^2 \le 1/2$. By the argument above, using that Y and Z are independent, we have

  $$\Pr\left[Y \cdot Z \le 0 \,\middle|\, |Y|, |Z| \le 1\right] \ge 1/2.$$

  Therefore, we have

  $$\begin{aligned}\Pr[|X| \le 1] &\ge \Pr[|Y| \le 1 \wedge |Z| \le 1 \wedge Y \cdot Z \le 0] \\ &= \Pr[|Y| \le 1] \cdot \Pr[|Z| \le 1] \cdot \Pr[Y \cdot Z \le 0 \,|\, |Y|, |Z| \le 1] \\ &\ge (1 - \mathrm{Var} Y)(1 - \mathrm{Var} Z) \cdot \frac{1}{2} \ge 1/10.\end{aligned}$$

Hence, we may take $c = 1/200$.