2.15: Analysis of Algorithms (Duration: 3 hours)

Final Exam (Part 2)

Guidelines

- Personal course notes (lectures and exercises) are permitted. Printouts of Florent Koechlin's course notes are also permitted, as is the "Probability Reminder" document.
- It is not necessary to answer a question correctly in order to use the result in subsequent questions.
- The exam consists of 2 parts. Please use separate sheets for each part.

Part A : Probabilistic Analysis of Algorithms

Exercise 1 : Diameter of $\mathcal{G}(n, p_n)$

The graph distance $d_G(v, v')$ between two vertices v, v' in an undirected graph G is the number of edges in a shortest path connecting v and v'. By convention $d_G(v, v') = +\infty$ if v, v' are not in the same connected component of G. The diameter $\mathrm{diam}(G)$ of G is then defined by

$$\mathrm{diam}(G) = \max\{d_G(v, v'); v, v' \in G\} \in \{0, 1, 2, \dots\} \cup \{+\infty\}.$$

For each $n \geqslant 1$ let $p_n \in (0, 1)$ and consider a realization G_n of the Erdös-Rényi random graph $\mathcal{G}(n, p_n)$. Recall that this is the random graph with vertices $v_1, \dots, v_n$ where each undirected edge $\{v_i, v_j\}$ is present independently with probability p_n.

The aim of the exercise is to establish a threshold phenomenon for the property "$\mathrm{diam}(G_n) \leqslant 2$" :

$$\mathbb{P}(\mathrm{diam}(G_n) \leqslant 2) \overset{n\to+\infty}{\to} \begin{cases} 1 & \text{if } p_n \gg \sqrt{\frac{\log(n)}{n}}, \\ 0 & \text{if } p_n \ll \sqrt{\frac{\log(n)}{n}}. \end{cases}$$

The proof relies on the 1st and 2d moment methods. Recall that these can be summarized in :

$$\frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]} \overset{\text{2d moment meth.}}{\leqslant} \mathbb{P}(Z \geqslant 1) \overset{\text{1st moment meth.}}{\leqslant} \mathbb{E}[Z],$$

for any random variable $Z \in \mathbb{Z}_{\geqslant 0}$. You also can use without proof every convexity inequality you may find useful :

$$1 - u \leqslant e^{-u} \leqslant 1 - \tfrac{u}{2} \qquad \tfrac{1}{2}u \leqslant \log(1 + u) \leqslant u, \qquad \dots$$

(The above inequalities are valid for all u in a small neighborhood of 0.)

Q1. For $1 \leqslant i \neq j \leqslant n$, let

$$X_{i,j} = \mathbf{1}_{d_{G_n}(v_i, v_j) > 2}.$$

Compute $\mathbb{E}[X_{i,j}]$.

Solution : If $d_{G_n}(v_i, v_j) > 2$ then it means that

- the edge $\{v_i, v_j\}$ is not present ;
- for all $v_k \notin \{v_i, v_j\}$, the two edges $\{v_i, v_k\}$ and $\{v_k, v_j\}$ are not both present.

Therefore

$$\mathbb{E}[X_{i,j}] = \mathbb{P}(d_{G_n}(v_i, v_j) > 2) = (1 - p_n) \times (1 - p_n^2)^{n-2}.$$

Q2. Using the 1st moment method with $X = \sum_{1 \leqslant i \neq j \leqslant n} X_{i,j}$, show that [1]

$$\mathbb{P}(\mathrm{diam}(G_n) > 2) \leqslant n^2 (1 - p_n^2)^{n-2}.$$

Solution : We have that

1. You may find that this is more natural to introduce $\tilde{X} = \sum_{1 \leqslant i < j \leqslant n} X_{i,j} = X/2$. However this would simplify question Q7 below to consider a sum over $i \neq j$ instead.

$$\{\mathrm{diam}(G_n) > 2\} \Leftrightarrow \{\text{There exist } v_i, v_j \text{ such that } d_{G_n}(v_i, v_j) > 2\} \Leftrightarrow \{X \geqslant 1\}.$$

Hence

$$\begin{aligned}\mathbb{P}\left(\mathrm{diam}(G_n) > 2\right) &= \mathbb{P}(X \geqslant 1)\\ &\leqslant \mathbb{E}[X] \qquad \text{(1st moment meth.)}\\ &= \sum_{1\leqslant i\neq j\leqslant n} \mathbb{E}[X_{i,j}] \qquad \text{(linearity of expectation)},\\ &= n(n-1)(1-p_n)\times(1-p_n^2)^{n-2} \qquad \text{(Q\textbf{??})},\\ &\leqslant n^2(1-p_n^2)^{n-2}.\end{aligned}$$

Q3. Deduce that if $\sqrt{\frac{\log(n)}{n}} = o(p_n)$ then $\mathbb{P}\left(\mathrm{diam}(G_n) \leqslant 2\right) \stackrel{n\to+\infty}{\to} 1$.

Solution : Let $\varepsilon(n)$ be such that $\log(n) = \varepsilon(n) n p_n^2$, one has $\varepsilon(n) \to 0$. Then

$$\begin{aligned}\mathbb{P}\left(\mathrm{diam}(G_n) > 2\right) &\leqslant n^2(1-p_n^2)^{n-2} \qquad \text{(Q2)}\\ &\leqslant \exp(2\log(n) - (n-2)p_n^2)\\ &\leqslant \exp(2\varepsilon(n) n p_n^2 - (n-2)p_n^2)\\ &\leqslant \exp(-np_n^2/2) \qquad (n \text{ large enough}),\\ &= \exp\left(-\frac{\log(n)}{2\varepsilon(n)}\right)\\ &\to 0.\end{aligned}$$

Q4. Prove that if, on the opposite, $p_n = o(\sqrt{\frac{\log(n)}{n}})$ then $\mathbb{E}[X] \stackrel{n\to+\infty}{\to} +\infty$.

Solution : Assume that $p_n = \varepsilon(n)\sqrt{\frac{\log(n)}{n}}$ for some $\varepsilon(n) \to 0$. Then

$$\begin{aligned}\mathbb{E}[X] &= n(n-1)(1-p_n)(1-p_n^2)^{n-2}\\ &\geqslant \tfrac{1}{4} n^2 (1-p_n^2)^{n-2}\\ &\geqslant \tfrac{1}{4}\exp(2\log(n) - \tfrac{1}{2}(n-2)p_n^2) \qquad (\text{using } (1-p_n^2) \geqslant e^{-p_n^2/2} \text{ for large } n)\\ &= \tfrac{1}{4}\exp\left(2\log(n) - \tfrac{1}{2}(n-2)\epsilon(n)^2\log(n)/n\right)\\ &\geqslant \tfrac{1}{4}\exp\left((2-o(1))\log(n)\right)\\ &= \tfrac{1}{4} n^{2-o(1)} \to +\infty.\end{aligned}$$

In the following questions we want to bound $\mathbb{E}[X^2]$ in order to use the 2d moment method.

Q5. Let $1 \leqslant i \neq j \leqslant n$ and $1 \leqslant k \neq \ell \leqslant n$ be four pairwise distinct integers (*i.e.* $\mathrm{card}\{i,j,k,\ell\} = 4$). Justify that

$$\mathbb{P}\left(d_{G_n}(v_i, v_j) > 2, d_{G_n}(v_k, v_\ell) > 2\right) \leqslant (1-p_n)^2(1-p_n^2)^{2n-6}.$$

Solution : Let $w_1, \dots, w_{n-4}$ be the remaining vertices.

$$\begin{aligned}\mathbb{P}\left(d_{G_n}(v_i, v_j) > 2, d_{G_n}(v_k, v_\ell) > 2\right) \leqslant &\mathbb{P}\big(\text{No path } v_i \to v_j \text{ of length 2 through } v_k, v_\ell, w_1, \dots, w_{n-4},\\ &\cap \text{ No path } v_k \to v_\ell \text{ of length 2 through } w_1, \dots, w_{n-4}\big)\\ &= (1-p_n)(1-p_n^2)^{n-2} \times (1-p_n)(1-p_n^2)^{n-4},\end{aligned}$$

since the two events on the RHS of the first display are independent.

Q6. Let $1 \leqslant i \neq j \leqslant n$ and $k \notin \{i, j\}$ be three pairwise distinct integers. Justify that

$$\mathbb{P}\left(d_{G_n}(v_i, v_j) > 2, d_{G_n}(v_i, v_k) > 2\right) \leqslant \left(1 - p_n + p_n(1-p_n)^2\right)^{n-3}.$$

Solution : If $d_{G_n}(v_i, v_j) > 2$ and $d_{G_n}(v_i, v_k) > 2$ then for every remaining u, one has

- either u is not connected to v_i (occurs with proba. $1-p_n$),
- or u is connected to v_i, but connected to none of v_j, v_k (occurs with proba. $p_n(1-p_n)^2$) (otherwise $d(v_i,v_j)\leqslant 2$ or $d(v_i,v_k)\leqslant 2$),

hence the claimed inequality.

Q7. Use the previous questions to prove that if $p_n = o(\sqrt{\frac{\log(n)}{n}})$ then

$$\mathbb{P}\left(\mathrm{diam}(G_n) > 2\right) \stackrel{n\to+\infty}{\to} 1.$$

Hint : *In order to save yourself some painful calculations, you can use (without proving) that if* $p_n = o(\sqrt{\frac{\log(n)}{n}})$ *then*

$$\left(1-p_n+p_n(1-p_n)^2\right)^{n-3} \leqslant \frac{1}{n^{\varepsilon(n)}}, \qquad (1-p_n^2)^{2n-4} \geqslant \frac{1}{n^{\delta(n)}}$$

where $\varepsilon(n), \delta(n)$ *are two sequences converging to zero.*

Solution : We use the 2d moment method with $X=\sum_{1\leqslant i\neq j\leqslant n} X_{i,j}$. First, by Question Q1 we have that

$$\mathbb{E}[X]^2 = \left(n(n-1)(1-p_n)\times(1-p_n^2)^{n-2}\right)^2.$$

Besides

$$\begin{aligned}\mathbb{E}[X^2] &= \mathbb{E}\left[\left(\sum_{1\leqslant i\neq j\leqslant n} X_{i,j}\right)\left(\sum_{1\leqslant k\neq \ell\leqslant n} X_{k,\ell}\right)\right]\\ &= \sum_{\substack{1\leqslant i\neq j\leqslant n\\ 1\leqslant k\neq \ell\leqslant n\\ \text{all distinct}}} \mathbb{E}\left[X_{i,j}X_{k,\ell}\right] + \sum_{\substack{1\leqslant i\neq j\leqslant n\\ 1\leqslant i\neq k\leqslant n\\ i,j,k \text{ all distinct}}} \mathbb{E}\left[X_{i,j}X_{i,k}\right] + \sum_{1\leqslant i\neq j\leqslant n} \mathbb{E}\left[X_{i,j}X_{i,j}\right]\\ &=: S_1+S_2+S_3.\end{aligned}$$

Let us bound the three sums on the RHS :

- By Question Q5,

$$S_1 \leqslant n(n-1)(n-2)(n-3)(1-p_n)^2(1-p_n^2)^{2n-6} \leqslant n^4(1-p_n)^2(1-p_n^2)^{2n-6}.$$

- By Question Q6

$$S_2 \leqslant n(n-1)(n-2)\left(1-p_n+p_n(1-p_n)^2\right)^{n-3} \leqslant n^3\left(1-p_n+p_n(1-p_n)^2\right)^{n-3}.$$

- $S_3 = \mathbb{E}[X]$.

One has

$$\mathbb{P}(X\geqslant 1) \geqslant \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \geqslant \frac{\mathbb{E}[X]^2}{S_1+S_2+S_3}. \tag{1}$$

It is clear that $S_1 \leqslant \mathbb{E}[X]^2 + o(\mathbb{E}[X]^2)$. So it suffices to prove that

$$S_2 = o(\mathbb{E}[X]^2), \qquad S_3 = o(\mathbb{E}[X]^2).$$

Regarding S_2 :

$$\begin{aligned}\frac{S_2}{\mathbb{E}[X]^2} &\leqslant \frac{n^3\left(1-p_n+p_n(1-p_n)^2\right)^{n-3}}{\left(n(n-1)(1-p_n)\times(1-p_n^2)^{n-2}\right)^2}\\ &\leqslant \mathrm{Const}\times\frac{n^3\left(1-p_n+p_n(1-p_n)^2\right)^{n-3}}{n^4\left((1-p_n^2)^{n-2}\right)^2}\\ &\leqslant \mathrm{Const}\times\frac{n^3}{n^{\varepsilon(n)}}\times\frac{n^{\delta(n)}}{n^4}\to 0,\end{aligned}$$

using the hint (the constant is here to get rid of $(n-1)^2$ and $(1-p_n)^2$). Regarding S_3 :

$$\frac{S_3}{\mathbb{E}[X]^2} = \frac{1}{\mathbb{E}[X]} \to 0$$

using Question Q4.
Finally

$$\mathbb{P}\left(\mathrm{diam}(G_n) > 2\right) = \mathbb{P}\left(X \geqslant 1\right) \to 1$$

using eq.(**??**).

Q8. We have proved that if $p_n \gg \sqrt{\frac{\log(n)}{n}}$ then $\mathrm{diam}(G_n) \in \{1, 2\}$ with high probability. It remains to determine whether the true value is 1 or 2. Find a good condition on the sequence (p_n) to have $\mathrm{diam}(G_n) = 2$ with high probability when $n \to +\infty$.

Solution : We have that

$$\mathbb{P}\left(\mathrm{diam}(G_n) = 1\right) = \mathbb{P}\left(G_n \text{ is the complete graph}\right) = p_n^{\binom{n}{2}}.$$

Set $p_n = 1 - \epsilon_n$, we have

$$\begin{aligned} \mathbb{P}\left(\mathrm{diam}(G_n) = 1\right) &= (1-\epsilon_n)^{\binom{n}{2}} \\ &\approx \exp\left(-\frac{n(n-1)}{2}\epsilon_n\right). \end{aligned}$$

Therefore

$$\mathbb{P}\left(\mathrm{diam}(G_n) = 1\right) \to \begin{cases} 1 & \text{if } \epsilon_n \ll \frac{1}{n^2}, \\ 0 & \text{if } \epsilon_n \gg \frac{1}{n^2}. \end{cases}$$