TD 07 - Graphes aléatoires (corrigé)

Exercice 1. *Graphe aléatoire bipartite*

Soit $0 < p < 1$ et $n \in \mathbb{N}^*$. On définit un graphe aléatoire non orienté $H_{2n,p}$ de la manière suivante : on se donne une famille $\{X_{i,j} \mid 1 \le i \le n,\ n+1 \le j \le 2n\}$ de v.a. i.i.d. de loi de Bernoulli de paramètre p. On pose alors $H_{2n,p} = (V, E)$ avec $V = \{1, \dots, 2n\}$ et

$$E = \{(i,j) \mid X_{i,j} = 1\} \subseteq \{1, \dots, n\} \times \{n+1, \dots, 2n\}.$$

1. Quelle est la loi du nombre d'arêtes de $H_{2n,p}$?

☞ Le nombre d'arêtes de $H_{2n,p}$ suit la loi $B(n^2, p)$.

2. Quelle est l'espérance du nombre de sommets isolés de $H_{2n,p}$?

☞ Soit N le nombre de sommets isolés. Si A_i est l'événement « le sommet i est isolé », on a par linéarité de l'espérance, on a $\mathsf{E}[N] = \sum \mathsf{P}(A_i) = 2n(1-p)^n$.

3. Dans cette question, on pose $p = c\log(n)/n$ pour un nombre réel $c > 0$.

   i. Montrer que si $c > 1$, alors :

$$\lim_{n\to\infty} \mathbf{P}\left\{H_{2n,p} \text{ a au moins un sommet isolé}\right\} = 0.$$

   ii. Montrer que si $c < 1$, alors :

$$\lim_{n\to\infty} \mathbf{P}\left\{H_{2n,p} \text{ a au moins un sommet isolé}\right\} = 1.$$

☞

i. Si $c > 1$, on a

$$\begin{aligned}\mathsf{E}[N] &= 2n(1-p)^n \\ &= 2n\left(1 - \frac{c\log n}{n}\right)^n \\ &= 2n\exp(n\log(1 - \frac{c\log n}{n})) \to 0\end{aligned}$$

et donc $\mathsf{P}(N \ge 1) \le \mathsf{E}[N] \to 0$.

ii. Si $c < 1$, on calcule

$$\mathsf{E}[N^2] = \sum_{i,j=1}^{2n} \mathsf{P}(A_i \cap A_j) = 2n(1-p)^n + 2n(n-1)(1-p)^{2n} + 2n^2(1-p)^{2n-1}$$

d'où il vient que $\mathsf{E}[N^2]/\mathsf{E}[N]^2$ tend vers 1 (la première partie tend vers 0 avec un développement limité, les deux autres chacune vers 1/2). On utilise l'inégalité de Tchebychev pour conclure que

$$\mathsf{P}(N=0) = \mathsf{P}(\mathsf{E}[N] - N \ge \mathsf{E}[N]) \le \mathsf{P}(|N - \mathsf{E}[N]| \ge \mathsf{E}[N]) \le \frac{\mathbf{Var}[N]}{\mathsf{E}[N]^2} = \frac{\mathsf{E}[N^2]}{\mathsf{E}[N]^2} - 1 \to 0.$$

4. Dans cette question, on pose $p = 1/2$. Montrer qu'il existe une constante $C > 0$ telle que

$$\lim_{n\to\infty} \mathbf{P}\left\{\text{tous les sommets de } H_{2n,p} \text{ ont un degré inférieur à } \frac{n}{2} + C\sqrt{n\log n}\right\} = 1.$$

☞ Le degré d_i du sommet i suit la loi $B(n, 1/2)$. Par l'inégalité de Chernoff I, on a donc

$$\mathsf{P}(d_i \ge \frac{n}{2} + a) \le \exp(-2a^2/n).$$

Ainsi, par la borne de l'union,

$$\mathsf{P}(\max_i d_i \ge \frac{n}{2} + a) \le 2n\exp(-2a^2/n).$$

Cette quantité tend vers 0 si $a = C\sqrt{n\log n}$ avec $2C^2 > 1$.

Exercice 2. *K4*

Soit G un graphe aléatoire de loi $G_{n,p}$. L'objectif de cet exercice est de montrer qu'il y a un seuil $p_0 := n^{-2/3}$ tel que pour $p = o(p_0)$, le graphe G n'a pas de clique de taille 4 avec bonne probabilité, et que pour $p = \omega(p_0)$, le graphe G a au moins une clique de taille 4 avec bonne probabilité.

Rappels / définitions :

— Un graphe aléatoire G suit la loi $G_{n,p}$ s'il a n sommets et que chaque arrête est présente dans G avec probabilité p ;

— une clique de taille 4 est un ensemble de 4 sommets tous reliés deux à deux par des arêtes ;

— $p = o(p_0)$ signifie $\frac{p}{p_0} \to 0$ quand $n \to +\infty$;

— $p = \omega(p_0)$ signifie $\frac{p_0}{p} \to 0$ quand $n \to +\infty$.

1. Pour p quelconque, calculer $\mathbf{E}[X]$, où X est le nombre de cliques de taille 4 du graphe G.

☞ On a $\binom{n}{4}$ ensembles possibles de 4 sommets. Pour chacun de ces ensembles, on définit X_i qui vaut 1 si c'est une clique et zero sinon. On a $X = \sum_i X_i$. Et $\mathbf{E}[X_i] = \mathbf{P}\{X_i = 1\} = p^6$. D'où

$$\mathbf{E}[X] = \binom{n}{4} p^6.$$

2. Soit $p = o(p_0)$, montrer que $\Pr(X \neq 0) \to 0$ quand n tend vers l'infini.

☞ On a $\mathbf{E}[X] \le n^4 p^6 = o(n^{4-6\times 2/3}) = o(1)$. Donc $\mathbf{E}[X]$ tend vers zero quand n tend vers l'infini. Comme X est a valeur entières, positives ou nulle, on conclut que $\mathbf{P}\{X \neq 0\} = \Pr(X \geq 1) \le \mathbf{E}[X]$ tend aussi vers 0.

On suppose maintenant $p = \omega(p_0)$, et on veut montrer que $\mathbf{P}\{X = 0\} \to 0$ quand n tend vers l'infini.

3. Montrer que $\mathbf{P}\{X = 0\} \le \frac{\mathbf{Var}[X]}{\mathbf{E}[X]^2}$. Il suffira donc de montrer que $\frac{\mathbf{Var}[X]}{\mathbf{E}[X]^2} \to 0$.

☞ On utilise l'inégalité de Chebychev

$$\Pr(X = 0) \le \mathbf{P}\{|X - \mathbf{E}[X]| \ge \mathbf{E}[X]\} \le \frac{\mathbf{Var}[X]}{\mathbf{E}[X]^2}.$$

4. Soit X_i des variables aléatoires à valeur dans $0,1$ (et non indépendantes). Montrer que

$$\mathbf{Var}\left[\sum_i X_i\right] \le \mathbf{E}\left[\sum_i X_i\right] + \sum_{i\neq j} \mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])\right].$$

☞ On a

$$\begin{aligned}
\mathbf{Var}\left[\sum_i X_i\right] &= \mathbf{E}\left[\left(\sum_i (X_i - \mathbf{E}[X_i])\right)^2\right] \\
&= \mathbf{E}\left[\sum_{i,j} (X_i - \mathbf{E}[X_i]) \cdot (X_j - \mathbf{E}[X_j])\right] \\
&= \sum_i \mathbf{E}\left[(X_i - \mathbf{E}[X_i])^2\right] + \sum_{i\neq j} \mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])\right].
\end{aligned}$$

On observe ensuite que $\mathbf{E}\left[(X_i - \mathbf{E}[X_i])^2\right] = \mathbf{E}[X_i^2] - \mathbf{E}[X_i]^2 \le \mathbf{E}[X_i^2]$. Mais comme X_i est à valeur dans $0,1$, on a $\mathbf{E}[X_i^2] = \mathbf{E}[X_i]$. D'où l'inégalité.

5. En déduire que $\mathbf{Var}[X] = o(\mathbf{E}[X]^2)$ et conclure.

☞ On note C_i les ensembles de 4 sommets du graphe G, et on défini X_i la variables aléatoire qui vaut 1 si C_i forme une clique et 0 sinon. On a $X = \sum_i X_i$ et d'après la question précédente $\mathbf{Var}[X] \le \mathbf{E}[X] + \sum_{i\neq j} \mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])\right]$. Fixons $i \neq j$ et considérons $\mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])\right] = \mathbf{E}[X_i X_j] - \mathbf{E}[X_i]\,\mathbf{E}[X_j]$. On a $\mathbf{E}[X_i X_j] = \mathbf{P}\{C_i \text{ et } C_j \text{ sont des cliques}\} = p^k$, où k est le nombre d'arêtes nécessaires pour que C_i et C_j soient des cliques. Ce nombre d'arêtes va dépendre nu nombre de sommets communs entre C_i et C_j. On distingue donc les cas suivants

— Si $|C_i \cap C_j| = 0$ ou $|C_i \cap C_j| = 1$, alors $k = 12$ et $\mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])\right] = 0$.

— Si $|C_i \cap C_j| = 2$, alors $k = 11$ et $\mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])\right] = p^{11}(1-p)$. Il y a $\binom{n}{4} \cdot \binom{4}{2} \cdot \binom{n-4}{2}$ tels couples (C_i, C_j).

— Si $|C_i \cap C_j| = 3$, alors $k = 9$ et $\mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])\right] = p^9(1-p^3)$. Il y a $\binom{n}{4} \cdot \binom{4}{3} \cdot \binom{n-4}{1}$ tels couples (C_i, C_j).

— Le cas $|C_i \cap C_j| = 4$ est impossible car $C_i \neq C_j$.

On a donc $\mathbf{Var}[X] \le \mathbf{E}[X] + \binom{n}{4} \cdot \binom{4}{2} \cdot \binom{n-4}{2} p^{11}(1-p) + \binom{n}{4} \cdot \binom{4}{3} \cdot \binom{n-4}{1} p^9(1-p^3)$. Chacun des trois termes de cette somme est un $o(\mathbf{E}[X]^2)$ (car $p = \omega(p_0)$), d'où la réponse à la question. On conclut grâce aux questions précédentes que $\Pr(X \neq 0) \to 0$ quand n tend vers l'infini.

Exercice 3. *Graphes connexes*

Un graphe non-orienté $G = (V, E)$ sur n sommets est *connexe* si, pour tous sommets u et v, il existe un chemin de u à v. Autrement dit, le graphe n'est pas connexe si l'on peut partitionner V en (A, B) de telle sorte qu'il n'existe aucune arête entre A et B.

1. Prouver que si $p = (2+\varepsilon)\log n/n$ avec $\varepsilon > 0$, alors la probabilité qu'un graphe choisi aléatoirement dans $G_{n,p}$ soit connexe tend vers 1 quand n tend vers l'infini.

☞

Note : $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$.

We first compute P_k the probability to have a cut (A,B) with no edge where $|A| = k$.

There are $k(n-k)$ edges in such a cut and $\binom{n}{k}$ such cuts. Then, by the union bound, we get, with $\tau = 2+\varepsilon$:

$$P_k \leq \binom{n}{k}(1-p)^{k(n-k)} \leq \left(\frac{en}{k}\right)^k \frac{1}{n}^{k(1-k/n)\tau} \leq \left(\frac{e}{kn^{1+\varepsilon}}\right)^k$$

For n sufficiently large, we can bound $P_k \leq P_1$. Then the sum of P_k is at most e/n^ε which goes to 0.

Exercice 4. *Théorème de Mycielski*

La coloration d'un graphe G consiste à attribuer une couleur à chacun de ses sommets de manière que deux sommets reliés par une arête soient de couleurs différentes. Le nombre minimal de couleurs est appelé *nombre chromatique*, on le note $\chi(G)$.

Clairement, les graphes contenant de grandes cliques ont un grand nombre chromatique, mais la réciproque n'est pas vraie. Le but de cet exercice est de prouver le théorème de Mycielski [1] : pour tout entier $k \geq 2$, il existe un graphe G tel que G ne contient aucun triangle et avec pourtant $\chi(G) \geq k$.

1. Soit $0 < \varepsilon < \frac{1}{3}$ et soit G un graphe aléatoire avec n sommets où chaque arrête est présente indépendamment des autres avec probabilité $p = n^{\varepsilon-1}$. Montrer que quand n tend vers l'infini, la probabilité que G ait plus de $n/2$ triangles tend vers 0.

   ☞ Pour trois sommets distincts a, b et c, la probabilité d'être un triangle est p^3. Il y a $\binom{n}{3}$ ensembles de trois sommets, donc l'espérance du nombre de triangle est $\mathbf{E}[N] \leq n^3p^3$. Par Markov,

$$\mathbf{P}\left\{N \geq \frac{n}{2}\right\} \leq \frac{n^3p^3}{\frac{n}{2}} = 2n^2p^3 = 2n^{3\varepsilon-1} \to 0.$$

2. Soit $\alpha(G)$ la taille du plus grand *ensemble indépendant* de G (un ensemble indépendant est un ensemble de sommets deux à deux non adjacents). Montrer que $\chi(G) \geq \frac{n}{\alpha(G)}$.

   ☞ Par définition de $\chi(G)$, il existe un coloriage de G avec $\chi(G)$ couleurs, autrement dit une partition de V en $\chi(G)$ ensembles indépendants. Alors pour chaque tel sous-ensemble A, on a $\alpha(G) \geq |A|$. Donc $\chi(G) \times \alpha(G) \geq n$, d'où le résultat.

3. Soit $a = 3n^{1-\varepsilon}\ln n$. Montrer que :

$$\mathbf{P}\{\alpha(G) < a\} \xrightarrow[n\to\infty]{} 1\,.$$

   En déduire qu'il existe n et G de taille n tels que G a au plus $\frac{n}{2}$ triangles et $\alpha(G) < a$.

   ☞ Pour tous sommets $v_1, \ldots, v_a$ distincts, on a

$$\mathbf{P}\{\{v_1, \ldots, v_a\} \text{ est indépendant}\} = (1-p)^{\frac{a(a-1)}{2}}.$$

   Donc

$$\mathbf{P}\{\alpha(G) \geq a\} \quad \leq \quad \binom{n}{a}(1-p)^{\frac{a(a-1)}{2}} \quad < \quad n^a(1-p)^{\frac{a(a-1)}{2}} \quad < \quad n^a\, e^{-p\frac{a(a-1)}{2}} \quad \to 0.$$

   Donc (question 1 et question 3) il existe n et G de taille n tels que G a au plus $\frac{n}{2}$ triangles et $\alpha(G) < a$.

4. Soit G un tel graphe. Soit G' un graphe obtenu à partir de G en supprimant le minimum de sommets afin que G' ne contienne aucun triangle. Montrer que :

$$\chi(G') > \frac{n^\varepsilon}{6\ln n}$$

   et conclure la preuve du théorème de Mycielski.

   ☞ Par la question 2, on a

$$\chi(G') \quad \geq \quad \frac{|G'|}{\alpha(G')} \quad > \quad \frac{n}{2} \times \frac{1}{a} \quad = \quad \frac{n^\varepsilon}{6\ln n}.$$

   Ce qui conclut la preuve du théorème (pour n assez grand, on peut obtenir G' sans triangle avec $\chi(G')$ aussi grand que l'on veut).

[1] Jan Mycielski (1932–2025), mathématicien polono-américain.