Graphes aléatoires

Frédéric Havet

JCALM Février 2012

Soit $\mathcal{G}_n$ l'ensemble de tous les graphes étiquetés à n sommets. Un espace de probabilité un peu plus évolué sur l'ensemble $\mathcal{G}_n$ s'obtient en fixant un réel p entre 0 et 1 et en choisissant chaque arête avec probabilité p, ces choix étant là encore indépendants les uns des autres. Comme $1-p$ est la probabilité qu'une arête particulière ne soit pas choisie, la fonction de probabilité P qui en résulte est donnée par

$$\mathbf{Pr}(G) = p^m(1-p)^{N-m}, \quad \text{pour tout } G \in \mathcal{G}_n$$

tel que $m := e(G)$. Cet espace est dénoté par $\mathcal{G}_{n,p}$.

Voyons tout d'abord quelques propriétés simples d'un graphe aléatoire de $\mathcal{G}_{n,p}$. L'espérance de son nombre d'arêtes est clairement $\binom{n}{2}p$ par linéarité de l'espérance. La distribution des degrés d'un sommet particulier est binomiale :

$$\mathbf{Pr}(d(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

Notations asymptotiques

Dans la suite, nous considérons des espaces de probabilité $(\Omega_n, \mathbf{Pr}_n)$ qui sont définis pour tous les entiers strictement positifs n. Etant donnée une suite $(\Omega_n, \mathbf{Pr}_n)$, $n \geq 1$, d'espaces de probabilité, une propriété A est dite *presque sûrement* satisfaite si $\mathbf{Pr}_n(A_n) \to 1$ quand $n \to \infty$, avec $A_n := A \cap \Omega_n$.

Nous employons les notations asymptotiques suivantes. Si $f : \mathbb{N} \to \mathbb{R}$ et $g : \mathbb{N} \to \mathbb{R}$ sont deux fonctions telles que $g(n) > 0$ pour n suffisamment grand, nous écrivons :

$$\begin{array}{llll} f \ll g & \text{si} & f(n)/g(n) \to 0 & \text{quand } n \to \infty \\ f \gg g & \text{si} & f(n)/g(n) \to \infty & \text{quand } n \to \infty \\ f \sim g & \text{si} & f(n)/g(n) \to 1 & \text{quand } n \to \infty \end{array}$$

1 Stabilité d'un graphe aléatoire

Rappelons qu'un *stable* dans un graphe est un ensemble de sommets deux à deux indépendants. La *stabilité* d'un graphe G, dénotée $\alpha(G)$, est le cardinal d'un plus grand stable de G.

1.1 Borne sur la stabilité d'un graphe aléatoire

Nous établissons maintenant une borne fondamentale et très utile sur la stabilité des graphes aléatoires due à [6]. A moins qu'il soit spécifié autrement, log désigne le logarithme népérien (c'està-dire, relatif à la base e).

Théorème 1. *Un graphe aléatoire de $\mathcal{G}_{n,p}$ a presque sûrement une stabilité d'au plus $\lceil 2p^{-1}\log n \rceil$.*

Afin de prouver ce théorème, nous avons besoin d'un outil classique, l'Inégalité de Markov.

Proposition 2 (INÉGALITÉ DE MARKOV)**.**
Soit X une variable aléatoire positive et t un réel strictement positif. Alors

$$\mathbf{Pr}(X \geq t) \ \leq \ \frac{\mathbf{E}(X)}{t}.$$

L' Inégalité de Markov est fréquemment utilisée sous la forme suivante afin de montrer qu'un graphe aléatoire de $\mathcal{G}_{n,p}$ possède presque sûrement une propriété particulière pour une certaine valeur de p. On l'obtient en posant $X = X_n$ et $t = 1$ dans la Proposition 2.

Corollaire 3. *Soit X_n une variable aléatoire à valeurs réelles positives dans un espace de probabilité $(\Omega_n, \mathbf{Pr}_n)$, $n \geq 1$. Si $E(X_n) \to 0$ quand $n \to \infty$, alors $\mathbf{Pr}(X_n = 0) \to 1$ quand $n \to \infty$.*

Preuve du Théorème 1. Soit $G \in \mathcal{G}_{n,p}$ et soit S un ensemble de $k+1$ sommets de G, avec $k \in \mathbb{N}$. La probabilité que S soit un stable de G est $(1-p)^{\binom{k+1}{2}}$, car c'est la probabilité qu'aucune des $\binom{k+1}{2}$ paires de sommets de S ne soit une arête du graphe aléatoire G.

Soit A_S l'évènement que S soit un stable de G, et soit X_S la variable aléatoire caractéristique de cet évènement. Nous avons

$$\mathbf{E}(X_S) = \mathbf{Pr}(X_S = 1) = \mathbf{Pr}(A_S) = (1-p)^{\binom{k+1}{2}} \tag{1}$$

Soit X le nombre de stables de cardinal $k+1$ dans G. Alors

$$X = \sum\{X_S : \ S \subseteq V, \ |S| = k+1\}$$

et donc, par linéarité de l'espérance et (1),

$$\mathbf{E}(X) = \sum\{\mathbf{E}(X_S) : \ S \subseteq V, \ |S| = k+1\} = \binom{n}{k+1}(1-p)^{\binom{k+1}{2}}.$$

Nous majorons le membre droit à l'aide de deux inégalités élémentaires :

$$\binom{n}{k+1} \leq \frac{n^{k+1}}{(k+1)!} \quad \text{et} \quad 1-p \leq e^{-p}.$$

Cela donne la borne supérieure suivante pour $\mathbf{E}(X)$.

$$\mathbf{E}(X) \leq \frac{n^{k+1}e^{-p\binom{k+1}{2}}}{(k+1)!} = \frac{\left(ne^{-pk/2}\right)^{k+1}}{(k+1)!}. \tag{2}$$

Supposons maintenant que $k = \lceil 2p^{-1}\log n \rceil$. Alors $k \geq 2p^{-1}\log n$, donc $ne^{-pk/2} \leq 1$. Comme k croît au moins aussi vite que le logarithme de n, (2) implique que $\mathbf{E}(X) \to 0$ quand $n \to \infty$. Comme X est à valeurs entières positives, nous déduisons du Corollaire 3 que $\mathbf{Pr}(X = 0) \to 1$ quand $n \to \infty$. Par conséquent, un graphe aléatoire de $\mathcal{G}_{n,p}$ est presque sûrement de stabilité au plus $\lceil 2p^{-1}\log n \rceil$. □

Dans le cas où $p = \frac{1}{2}$, une borne sur α légèrement meilleure que celle fournie par le Théorème 1 peut être obtenue.

Théorème 4. *Si $G \in \mathcal{G}_{n,1/2}$, alors presque sûrement $\alpha(G) \leq \lceil 2\log_2 n \rceil$.*

Preuve: (Grandes lignes). Soit n un entier strictement positif. Pour $0 \leq k \leq n$, on pose $f(k) := \binom{n}{k}2^{-\binom{k}{2}}$. Soit $k^* = k^*(n)$ la plus petite valeur de k pour laquelle $f(k)$ est inférieur à un. On montre que :

$$k^* \leq \lceil 2\log_2 n \rceil \leq k^* + \log_2 k^* - 1, \tag{3}$$

$$f(k^*+1) \to 0 \ \text{ quand } \ n \to \infty.$$

Le résultat en découle. □

1.2 Concentration de la stabilité

Nous allons maintenant montrer une bien meilleure estimation sur la stabilité d'un graphe aléatoire. Celle-ci est due à Bollobás et Erdős [5] et Matula [9] indépendamment.

Théorème 5. *Soit $G \in \mathcal{G}_{n,1/2}$. Pour $0 \leq k \leq n$, posons $f(k) := \binom{n}{k} 2^{-\binom{k}{2}}$ et soit k^* la plus petite valeur de k pour laquelle $f(k)$ est inférieure à un. Alors presque sûrement $\alpha(G)$ prend une des trois valeurs k^*-2, k^*-1, k^*.*

Afin de montrer ce théorème, nous avons besoin d'un autre outil classique qui permet de connaître comment la distribution d'une variable aléatoire se concentre autour de son espérance. La *variance* d'une variable aléatoire X est définie par

$$\mathbf{Var}(X) := \mathbf{E}((X-\mathbf{E}(X))^2) = \mathbf{E}(X^2) - \mathbf{E}^2(X).$$

L'inégalité de Tchebychev borne la divergence d'une variable aléatoire par rapport à son espérance.

Théorème 6 (INÉGALITÉ DE TCHEBYCHEV)**.**
Soit X une variable aléatoire et t un réel strictement positif. Alors

$$\mathbf{Pr}(|X-\mathbf{E}(X)| \geq t) \leq \frac{\mathbf{Var}(X)}{t^2}.$$

L'Inégalité de Tchebychev est fréquemement employée sous la forme suivante obtenue en posant $X := X_n$ et $t := |\mathbf{E}(X_n)|$ et en observant que $\mathbf{Pr}(X_n = 0) \leq \mathbf{Pr}(|X_n - \mathbf{E}(X_n)| \geq |\mathbf{E}(X_n)|)$ parce que $|X_n - \mathbf{E}(X_n)| = |\mathbf{E}(X_n)|$ quand $X_n = 0$.

Corollaire 7. *Soit X_n une variable aléatoire dans un espace de probabilité $(\Omega_n, \mathbf{Pr}_n)$, $n \geq 1$. Si $\mathbf{E}(X_n) \neq 0$ et $\mathbf{Var}(X_n) \ll \mathbf{E}^2(X_n)$, alors*

$$\mathbf{Pr}(X_n = 0) \to 0 \ \textit{ quand } n \to \infty.$$

Preuve du Théorème 5. Comme dans la démonstration du Théorème 1, soit X_S la variable aléatoire caractéristique de l'évènement A_S qu'un sous-ensemble donné S de V soit un stable de G, et posons $X := \sum\{X_S : \ S \subseteq V, \ |S| = k\}$. On a donc $\mathbf{E}(X) = f(k)$. D'après le Théorème 4, presque sûrement $\alpha(G) \leq k^*$. Par conséquent, en vertu du Corollaire 7, il nous suffit de prouver que $\mathbf{Var}(X) \ll \mathbf{E}^2(X)$ quand $k = k^* - 2$. Nous supposerons donc désormais que k prend cette valeur. Par la preuve du Théorème 4, nous avons :

$$k < 2\log_2 n \ \text{ et } \ f(k) \geq n/4. \tag{4}$$

Nous allons appliquer le Corollaire 7. Pour cela, nous devons borner la la variance $\mathbf{Var}(X)$. Le concept suivant est utile à cet égard.

La *covariance* $\mathbf{Cov}(X,Y)$ de deux variables aléatoires X et Y est définie par

$$\mathbf{Cov}(X,Y) := \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y)$$

Comme X est une somme de variables aléatoires caractéristiques X_S, sa variance peut être bornée en termes de covariances comme suit.

$$\mathbf{Var}(X) \leq \mathbf{E}(X) + \sum_{S \neq T} \mathbf{Cov}(X_S, X_T) \tag{5}$$

Soient S et T deux ensembles de k sommets. La valeur de la covariance $\mathbf{Cov}(X_S, X_T)$ dépend uniquement de $|S \cap T|$. Si $|S \cap T| = 0$ ou $|S \cap T| = 1$, alors $\mathbf{Cov}(X_S, X_T) = 0$ parce qu'aucune arête n'a ses deux extrémités dans $S \cap T$. Si $|S \cap T| = i$, avec $2 \leq i \leq k-1$, alors

$$\mathbf{Cov}(X_S, X_T) \leq \mathbf{E}(X_S X_T) = \mathbf{Pr}(A_S \cap A_T) = 2^{\binom{i}{2} - 2\binom{k}{2}}$$

Il y a $\binom{n}{k}$ choix pour S, $\binom{k}{i}$ choix pour $S \cap T$, et $\binom{n-k}{k-i}$ choix pour T. Ainsi, par l'inégalité (5),

$$\mathbf{Var}(X) \le \mathbf{E}(X) + \sum_{S \neq T} \mathbf{Cov}(X_S, X_T) \le \mathbf{E}(X) + \sum_{i=2}^{k-1} \binom{n}{k}\binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2} - 2\binom{k}{2}}$$

Comme $\mathbf{E}(X) \ll \mathbf{E}^2(X)$, il reste à montrer que

$$\sum_{i=2}^{k-1} \binom{n}{k}\binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2} - 2\binom{k}{2}} \ll \mathbf{E}^2(X) = \binom{n}{k}^2 2^{-2\binom{k}{2}}$$

ou de manière équivalente que

$$\binom{n}{k}^{-1} \sum_{i=2}^{k-1} g(i) \to 0 \text{ quand } n \to \infty \tag{6}$$

avec

$$g(i) := \binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2}}$$

Nous avons :

$$g(2) = 2\binom{k}{2}\binom{n-k}{k-2} < k^2 \binom{n}{k-2}$$

et, pour $2 \le i \le k-2$,

$$\frac{g(i+1)}{g(i)} = \frac{(k-i)^2 2^i}{(i+1)(n-2k+i+1)} < \frac{k^2 2^i}{i(n-2k)} = \left(\frac{2^i}{i}\right)\left(\frac{k^2}{n-2k}\right)$$

Posons $t := \lfloor c \log_2 n \rfloor$, avec $0 < c < 1$. Alors, pour $2 \le i \le t-1$ et n suffisamment grand,

$$\frac{g(i+1)}{g(i)} < \left(\frac{2^t}{t}\right)\left(\frac{k^2}{n-2k}\right) < \left(\frac{n^c}{c \log_2 n}\right)\left(\frac{(2\log_2 n)^2}{n - 4\log_2 n}\right) = \frac{4n^c \log_2 n}{c(n - 4\log_2 n)} \le 1$$

donc

$$\binom{n}{k}^{-1} \sum_{i=2}^{t} g(i) < t \binom{n}{k}^{-1} g(2) \sim \frac{tk^4}{n^2} \to 0 \text{ quand } n \to \infty \tag{7}$$

Nous considérons les termes qui restent dans la somme de (6). Nous avons :

$$\begin{aligned}
\sum_{i=t+1}^{k-1} g(i) = \sum_{i=t+1}^{k-1} \binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2}} &= 2^{\binom{k}{2}} \sum_{i=t+1}^{k-1} \binom{k}{k-i}\binom{n-k}{k-i} 2^{-(k-i)(k+i-1)/2} \\
&= 2^{\binom{k}{2}} \sum_{j=1}^{k-t-1} \binom{k}{j}\binom{n-k}{j} 2^{-j(2k-j-1)/2} \\
&< 2^{\binom{k}{2}} \sum_{j=1}^{k-t-1} \left(k(n-k)2^{-(k+t)/2}\right)^j
\end{aligned}$$

Afin de majorer le membre droit, nous utilisons le fait que $k^* + \log_2 k^* - 1 \ge 2\log_2 n$ d'après (4), d'où $2^{-k/2} < (2k+4)^{1/2} n^{-1}$. Nous déduisons que pour n suffisamment grand,

$$k(n-k)2^{-(k+t)/2} < k(n-k)(2k+4)^{1/2} n^{-1} n^{-c/2} \le 1$$

et donc

$$\sum_{i=t+1}^{k-1} g(i) < 2^{\binom{k}{2}}(k-t-1)$$

En utilisant la borne (4) sur $f(k)$, nous obtenons :

$$\binom{n}{k}^{-1} \sum_{i=t+1}^{k-1} g(i) < \binom{n}{k}^{-1} 2^{\binom{k}{2}} (k-t-1) = \frac{(k-t-1)}{f(k)} \to 0 \text{ quand } n \to \infty \tag{8}$$

Les limites (7) et (8) impliquent (6). □

Le Théorème 5 peut être amélioré en un théorème de concentration en ‘deux points’.

Théorème 8. *Soit $G \in \mathcal{G}_{n,1/2}$, et soient f et k^* comme définis au Théorème 5. Alors :*

1. *soit $f(k^*) \ll 1$, auquel cas $\alpha(G)$ est presque sûrement égal à $k^* - 2$ ou $k^* - 1$,*
2. *soit $f(k^* - 1) \gg 1$, auquel cas $\alpha(G)$ est presque sûrement égal à $k^* - 1$ ou k^*.*

Observons que si $f(k^*) \ll 1$ et $f(k^* - 1) \gg 1$, alors les énoncés (i) et (ii) du Théorème 8 impliquent que $\alpha(G) = k^* - 1$ presque sûrement. C’est effectivement le cas pour la plupart des valeurs de n.

2 Nombre chromatique d’un graphe aléatoire

Une *k-coloration propre c* d’un graphe G est une application $c : V(G) \to \{1, 2, \ldots, k\}$ telle que $c(u) \neq c(v)$ pour toute arête $uv \in E(G)$. Un graphe est dit *k-colorable*, s’il est admet une k-coloration propre. Le *nombre chromatique* d’un graphe G, dénoté $\chi(G)$, est le plus petit entier k tel que G soit k-colorable. On peut également voir le nombre chromatique comme le plus petit nombre de stables nécessaires pour couvrir (ou partitionner) les sommets de G. On a donc

$$\chi(G) \geq \frac{|V(G)|}{\alpha(G)} \tag{9}$$

2.1 Ordre de grandeur du nombre chromatique

Cette inégalité et le Théorème 4 implique que si $G \in \mathcal{G}_{n,1/2}$, alors presque sûrement $\chi(G) \geq \frac{n}{2\log_2 n}$. En fait, Bollobás [4] a montré qu’en fait $\chi(G)$ est presque sûrement $\frac{n}{2\log_2 n}$

Théorème 9 (Bollobás [4])**.** *Si $G \in \mathcal{G}_{n,1/2}$, alors presque sûrement $\chi(G) \sim \frac{n}{2\log_2 n}$.*

Preuve. Nous venons de voir que $\chi(G) \geq \frac{n}{2\log_2 n}(1 + o(1))$. Il nous suffit donc de montrer l’inégalité inverse.

Posons $s = \lfloor n/\log^2 n \rfloor$. Pour tout ensemble S de s sommets le sous-graphe induit $G[S]$ a la distribution de $\mathcal{G}_{s,1/2}$. Soit $k' = k'(s) = k^*(s) - 4$ avec k^* défini comme dans la preuve du Théorème 4. Nous avons

$$k' \sim 2\log_2 s \sim 2\log_2 n.$$

En utilisant des techniques de concentration proches de la preuve du Théorème 5, on peut alors montrer que

$$\mathbf{Pr}(\alpha(G[S]) < k') < e^{-s^{2+o(1)}}.$$

Il y a $\binom{n}{s} < 2^n = 2^{s^{1+o(1)}}$ tels ensembles S. D’où

$$\mathbf{Pr}(\alpha(G[S]) < k' \text{ pour un ensemble } S \text{ de cardinal } s) < 2^{s^{1+o(1)}} e^{-s^{2+o(1)}} = o(1).$$

Ainsi, presque sûrement, tous les ensembles de s sommets de G contiennent un stable de cardinal k'.

Maintenant supposons qu'un graphe G ait cette propriété. On peut enlever des stables de cardinal k et leurs donner des couleurs différentes jusqu'à ce qu'il reste moins de s sommets. On peut alors donner une couleur différente à chacun des sommets restants. Cette procédure donne

$$\begin{aligned}\chi(G) &\leq \left\lceil \frac{n-s}{k'} \right\rceil + s \leq \frac{n}{k'} + s \\ &\leq \frac{n}{2\log_2 n}(1+o(1)) + o\left(\frac{n}{\log_2 n}\right) \\ &\leq \frac{n}{2\log_2 n}(1+o(1)).\end{aligned}$$

□

2.2 Concentration du nombre chromatique

Le nombre chromatique d'un graphe aléatoire, comme beaucoup d'autres paramètres de graphes, est lui aussi concentré autour de son espérance. D'ailleurs, on peut montrer cette concentration sans même connaître la valeur de l'espérance.

Théorème 10 (Shamir et Spencer [10])**.** *Soit n un entier, p un réel dans $[0,1]$. Si $G \in \mathcal{G}_{n,p}$, alors*

$$\mathbf{Pr}(|\chi(G) - \mathbf{E}(\chi(G))| > t) < 2e^{-\frac{t^2}{2}}.$$

La preuve de ce théorème est une application simple de résultats classiques sur les *martingales*. Une *martingale* est une suite $X_0, X_1, \ldots, X_p$ de variables aléatoires telles que pour $0 \leq i < p$,

$$\mathbf{E}(X_{i+1}|X_i, X_{i-1}, \ldots, X_0) = X_i.$$

La martingale qui nous intéresse est la *martingale de révélation des sommets* qui est la suivante. Nous nous plaçons dans l'espace de probabilité $\mathcal{G}_{n,p}$. Les graphes aléatoires de ce paragraphe appartiennent à cet espace. On prend un ordre quelconque des sommets $v_1, v_2, \ldots, v_n$. Soit f un paramètre de graphe. La *martingale de révélation des sommets* $X_1, \ldots, X_n$ est définie de la manière suivante :

$$X_i(H) = \mathbf{E}(f(G) \mid \text{ pour } j,k \leq i, v_i v_j \in G \text{ si et seulement si } v_i v_j \in H).$$

Observons que $X_i(H)$ est constante sur l'ensemble des graphes de $\mathcal{G}_n$ qui ont le même sous-graphe induit par $\{v_1, \ldots, v_i\}$. Autrement dit, pour trouver $X_i(H)$ on découvre d'abord les arêtes entre les i premiers sommets et on voit si cela forme le graphe $H[\{v_1, \ldots, v_i\}]$. Les autres arêtes ne sont pas révélées et donc considérées comme aléatoires. $X_i(H)$ est alors l'espérance conditionnelle de $f(G)$ avec cette information.

Clairement $X_1(H) = \mathbf{E}(f(G))$ (car aucune arête n'est révélée) et $X_n(H) = f(H)$ pour tout $H \in \mathcal{G}_n$ comme toutes les arêtes ont été révélés.

Un outil fondamental sur les martingales est l'Inégalité d'Azuma [2].

Théorème 11 (Inégalité d'Azuma)**.** *Soit $X_1, \ldots, X_p$ une martingale telle que $X_1 = 0$ et $|X_{i+1} - X_i| \leq 1$ pour tout $1 \leq i < p$. Pour tout réel t strictement positif,*

$$\mathbf{Pr}(X_p > t) < e^{-\frac{t^2}{2p-2}}.$$

Un paramètre de graphe f est *lipschitzien* (sur les sommets) si quels que soient deux graphes G et H qui diffèrent d'un seul sommet alors $|f(H) - f(H')| \leq 1$. De nombreux paramètres de graphes sont lipschitziens. C'est en particulier le cas de la stabilité et du nombre chromatique. En effet, supposons que $H = G - v$ pour un certain sommet v. Toute coloration propre de G induit une coloration propre de H. A l'inverse, toute k-coloration propre de H peut être étendue en une $(k+1)$-coloration propre de G en donnant à v une nouvelle couleur. Donc $\chi(H) \leq \chi(G) \leq \chi(H)+1$.

On peut montrer que si f est un paramètre de graphe est lipschitzien sur les sommets, alors la martingale de révélation des sommets correspondante $X_1, \ldots, X_n$ satisfait $|X_{i+1} - X_i| \leq 1$ pour tout $1 \leq i < n$. Intuitivement, cela veut dire que si la connaissance d'un sommet ne change la valeur de f que de 1 au plus, alors révéler un sommet ne peut pas changer l'espérance de f de plus que 1. Considérons alors la martingale $X'_1, X'_2, \ldots X'_n$, définie par $X'_i = X_i - \mathbf{E}(f(G))$. On a $X'_0 = 0$ et $|X'_{i+1} - X'_i| \leq 1$ pour tout $1 \leq i < n$. On peut donc appliquer l'Inégalité d'Azuma sur cette martingale et obtenir le théorème suivant.

Théorème 12. *Soit $G \in \mathcal{G}_{n,p}$. Si f satisfait la condition de Lipschitz (sur les sommets), alors*

$$\mathbf{Pr}(|f(G) - \mathbf{E}(f(G))| > t) < 2e^{-\frac{t^2}{2n-2}}.$$

Le Théorème 10 découle directement du Théorème 12 et du fait que le nombre chromatique satisfait la condition de Lipschitz.

Remarque 13. Le Théorème 12 possède un analogue 'arête'. La condition de Lipschitz porte alors sur la différence d'une arête et la martingale de révélation des sommets est remplacée par la martingale par révélation d'arêtes.

En utilisant les martingales de manière plus astucieuse et complexe, on peut montrer que le nombre chromatique d'un graphe aléatoire de $\mathcal{G}_{n,n^{-\alpha}}$ est concentré sur quatre valeurs pour toute constante $\alpha > 5/6$. Voir le Chapitre 7 du livre d'Alon et Spencer [1].

3 Evolution des graphes aléatoires

3.1 Fonction seuil

Si Π est une propriété de graphes *monotone* (c'est-à-dire une propriété qui est préservée lorsque des arêtes sont ajoutées), une *fonction seuil* pour Π est une fonction $f(n)$ telle que :

- si $p \ll f(n)$, alors presque sûrement $G \in \mathcal{G}_{n,p}$ n'a pas la propriété Π,
- si $p \gg f(n)$, alors presque sûrement $G \in \mathcal{G}_{n,p}$ a la propriété Π.

Un phénomène remarquable des graphes aléatoires est que pour toute propriété de graphes *monotone*, il y a une fonction seuil.

Théorème 14 (Alon)**.** *Toute propriété de graphes monotone admet une fonction seuil.*

Preuve (Grandes lignes). On considère une propriété Π non-triviale, c'est-à-dire que pour large n, le graphe vide $\overline{K_n}$ n'a pas la propriété Π alors que le graphe complet K_n l'a. On pose $\mathbf{Pr}(p) := \mathbf{Pr}(G \in G_{n,p}$ a la propriété $\Pi)$.

1. On montre que, pour tout n grand fixé, $\mathbf{Pr}(p)$ est un polynôme en p croissant qui satisfait $\mathbf{Pr}(0) = 0$ et $\mathbf{Pr}(1) = 1$. On en déduit que, pour tout r, $0 \leq r \leq 1$, il y a un p, $0 \leq p \leq 1$, tel que $\mathbf{Pr}(p) = r$.

2. Supposons que $\mathbf{Pr}(p) = r$. Soit $G_i \in G_{n,p}$ des membres indépendants de $G_{n,p}$, $1 \leq i \leq k$. On montrer que :

$$\begin{aligned} \mathbf{Pr}(kp) &\geq \mathbf{Pr}(\cup_{i=1}^{k} G_i \text{ a la propriété } \Pi) \\ &\geq \mathbf{Pr}(G_i \text{ a la propriété } \Pi \text{ pour un certain } i, 1 \leq i \leq k) = 1-(1-r)^k \end{aligned}$$

3. Pour n grand, soit $f(n)$ telle que $\mathbf{Pr}(f(n)) = 1/2$, et on suppose que $k(n)$ satisfait $f(n)k(n) \leq 1$. On montre que $\mathbf{Pr}(f(n)/k(n)) \leq 1/k(n)$ et que $\mathbf{Pr}(f(n)k(n)) \geq 1 - \frac{1}{2^{k(n)}}$. On en conclut que $f(n)$ est une fonction seuil pour Π.

$\square$

Il est à noter que nous disons '*une* fonction seuil', et pas '*la* fonction seuil'. C'est parce que la fonction n'est pas unique. Des fonctions presque équivalentes, sont toutes fonctions seuils. Par exemple, nous verrons que n^{-1}, $10^{10}n^{-1}$ et $n^{-1}+n^{-2}$ sont toutes des fonctions seuils pour la propriété de contenir un triangle.

Déterminer une fonction seuil n'est ce pendant pas toujours aisé. Nous le faisons maintenant pour la propriété de contenir un certain graphe fixé, dans le cas particulier où celui-ci est harmonieux. Un graphe F est *harmonieux* si le degré moyen de chacun des sous-graphes de F n'excède pas le sien $\mathrm{Ad}(F) = 2|E(F)|/|V(F)|$. Les arbres et les graphes réguliers sont des cas particuliers de graphes harmonieux.

Théorème 15. *Soit F un graphe harmonieux non-vide à k sommets et l arêtes. Alors $n^{-k/l}$ est une fonction seuil pour la propriété de contenir F comme sous-graphe.*

Proof. Soit $G \in \mathcal{G}_{n,p}$. Pour tout k-sous-ensemble S de V, soit A_S l'évènement que $G[S]$ contienne une copie de F, et soit X_S la variable aléatoire caractéristique pour A_S. Posons

$$X := \sum\{X_S : S \subseteq V,\ |S| = k\}$$

Ainsi X est le nombre de k-sous-ensembles qui contiennent des copies de F, et n'est donc pas plus grand que le nombre total de copies de F dans G.

Nous donnons tout d'abord un encadrement de l'espérance de X. Considérons un k-sous-ensemble S de V. Si $G[S]$ contient une copie de F, il y a une bijection $f : V(F) \to S$ telle que $f(u)f(v)$ soit une arête de $G[S]$ lorsque uv est une arête de F. La probabilité que toutes ces l arêtes $f(u)f(v)$ soient présentes dans $G[S]$ est p^l. Ainsi $\mathbf{E}(X_S) = \mathbf{Pr}(A_S) \geq p^l$. D'autre part, comme il y a $k!$ bijections $f : V(F) \to S$, donc $k!$ copies possibles de F dans $G[S]$ en tout, $\mathbf{E}(X_S) \leq k!p^l$. (Il y a inégalité ici parce que des copies de F dans $G[S]$ peuvent avoir des arêtes en commun, et donc ne sont pas indépendantes.) Par linearité de l'espérance et le fait que $n^k/k^k \leq \binom{n}{k} \leq n^k/k!$, pour $n \geq k \geq 0$, il vient que

$$\frac{n^k p^l}{k^k} \leq \binom{n}{k} p^l \leq \mathbf{E}(X) \leq \binom{n}{k} k!p^l < n^k p^l \tag{10}$$

Si $p \ll n^{-k/l}$, alors $\mathbf{E}(X) < n^k p^l \to 0$ quand $n \to \infty$ et, par l'Inégalité de Markov (Propodsition 2), G contient presque sûrement aucune copie de F.

Nous majorons maintenant la variance de X à l'aide de (5). Comme précédemment, la valeur de $\mathbf{Cov}(X_S, X_T)$ dépend uniquement de $|S \cap T|$. Si $|S \cap T| = 0$ ou $|S \cap T| = 1$, alors de nouveau $\mathbf{Cov}(X_S, X_T) = 0$. Si $|S \cap T| = i$, avec $2 \leq i \leq k-1$, alors chaque copie F_S de F dans $G[S]$ intersecte chaque copie F_T de F dans $G[T]$ en i sommets. Comme F is harmonieux, l'intersection $F_S \cap F_T$ de ces deux copies de F a au plus il/k arêtes, donc leur union $F_S \cup F_T$ a au moins $2l - (il/k)$ arêtes. La probabilité que les deux copies soient présentes dans G est donc au plus $p^{2l-(il/k)}$. Comme il y a $k!$ copies possibles F_S de F dans $G[S]$ et $k!$ copies possibles F_T de F dans $G[T]$,

$$\mathbf{Cov}(X_S, X_T) \leq \mathbf{E}(X_S X_T) = \mathbf{Pr}(A_S \cap A_T) \leq (k!)^2 p^{2l-(il/k)}$$

En tout, il y a $\binom{n}{k}\binom{n-k}{k-i}$ paires (S,T) de k-sous-ensembles de $|S \cap T| = i$. Or $\binom{n}{k} \leq n^k$ et $\binom{n-k}{k-i} \leq n^{k-i}$, d'où

$$\sum_{S \neq T} \mathbf{Cov}(X_S, X_T) \leq \sum_{i=2}^{k-1} n^{2k-i}(k!)^2 p^{2l-(il/k)} = (k!)^2 \sum_{i=2}^{k-1} (n^k p^l)^2 (np^{l/k})^{-i} \tag{11}$$

Si $p \gg n^{-k/l}$, alors $(np^{l/k})^{-i} \to 0$ pour $i \geq 1$. De plus, par (10), $\mathbf{E}(X) \geq n^k p^l / k^k \to \infty$ quand $n \to \infty$. Donc $\mathbf{E}(X) \ll \mathbf{E}^2(X)$ et $(n^k p^l)^2 \leq k^{2k}\mathbf{E}^2(X)$. Les inégalités (5) et (11) donnent alors :

$$\mathbf{Var}(X) \leq \mathbf{E}(X) + \sum_{S \neq T} \mathbf{Cov}(X_S, X_T) \ll \mathbf{E}^2(X)$$

Appliquant le Corollaire 7, nous concluons que le graphe aléatoire G contient presque sûrement une copie de F. $\square$

Corollaire 16. *La fonction f définie par $f(n) = n^{-1}$ est un fonction seuil pour l'existence d'un cycle.*

La fonction seuil pour la propriété de contenir un graphe F quelconque a également été donnée. Elle dépend du *degré moyen maximum* de F, $\mathrm{Mad}(F) = \max\{\mathrm{Ad}(H) : H \subseteq F\}$. Observons que si F est harmonieux, alors $\mathrm{Mad}(F) = \mathrm{Ad}(F)$.

Théorème 17 (Erdős et Rényi)**.** *Soit F un graphe non-vide. Alors $n^{-2/\mathrm{Mad}(F)}$ est une fonction seuil pour la propriété de contenir F comme sous-graphe.*

La preuve de ce théorème utilise une méthode similaire.

3.2 Composante géante

Erdős et Rényi [7] ont montré que quand la probabilité $p := p(n)$ augmente (alors que n reste constant), un graphe aléatoire typique $G \in \mathcal{G}_{n,p}$ passe à travers un certain nombre de phases critiques au cours desquelles sa structure change brusquement. Au delà de son intérêt propre, la compréhension de ce comportement peut être d'un grande aide pour utiliser la méthode probabiliste. Nous nous contenterons ici de donner une description générale du phénomène sans entrer dans les détails techniques compliqués.

En prenant pour F un arbre à k sommets dans le Théorème 15, nous voyons que $n^{-k/(k-1)}$ est une fonction seuil que G contienne un tel arbre. Comme le nombre d'arbres non-isomorphes à k sommets est clairement inférieur à k^{k-2} (le nombre d'arbres étiquetés), ceci implique que lorsque $p \ll n^{-k/(k-1)}$, G n'a pas de composante à k sommets ou plus (car une telle composante contiendrait un arbre à k sommets), mais lorsque $p \gg n^{-k/(k-1)}$, G a de telles composantes. De plus, ces composantes sont des arbres parce que, d'après le Corollaire 16, les cycles apparaissent uniquement au seuil $p = 1/n$. Par conséquent, à $p = n^{-k/(k-1)}$, G est une forêt dont toutes les composantes ont au plus k sommets.

Ces composantes deviennent de plus en plus grande à mesure que k augmente. Par une analyse probabiliste plus sophistiquée utilisant les processus de branchement, on peut montrer que lorsque $p = c/n$ avec $c < 1$, la plus grande composante de G est de taille environ $\log n$, alors qu'à $p = 1/n$ elle a déjà un taille d'environ $n^{2/3}$, et il y a beaucoup de composantes de cette taille. Quand $p = c/n$ avec $c > 1$, se produit une autre transformation majeure, avec l'émergence d'une 'composante géante' contenant une fraction strictement positive des n sommets. Ce changement radical au seuil $p = 1/n$ possède diverses appellations comme le *Double Saut* ou le *Big Bang*.

Une autre évolution remarquable se produit au seuil $p = \log n/n$. A ce stade, G peut encore avoir des sommets isolés. Quand ceux-ci disparaissent, G devient connexe et alors, presque immédiatement, hamiltonien.

Pour un exposé détaillé sur l'évolution des graphes aléatoires, nous renvoyons le lecteur vers [3] ou [8].

References

[1] N. Alon and J. Spencer. *The Probabilistic Method.* Second edition. Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, New York, 2000.

[2] K. Azuma. Weighted sums of certain dependent random variables, *Tôhoku Math. J. (2)*, **19**, 357–367, 1967.

[3] B. Bollobás. *Random Graphs*. Second edition. Cambridge Studies in Advanced Mathematics, Vol. 73, Cambridge University Press, Cambridge, 2001.

[4] B. Bollobás. The chromatic number of random graphs, *Combinatorica* **8**, 49–55, 1988.

[5] B. Bollobás and P. Erdős Cliques in random graphs. *Math. Proc. Cambridge Philos. Soc.* **80**, 419–427, 1976.

[6] P. Erdős. Graph theory and probability. II. *Canad. J. Math.* **13**, 346–352, 1961.

[7] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5**, 17–61, 1960.

[8] S. Janson, T. Łuczak et A. Rucinski. *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, New York, 2000.

[9] D. Matula. The largest clique size in a random graph. *Technical report*, Dept. Comp. Sci., Southern Methodist University, Dallas, Texas, 1976.

[10] E. Shamir and J. H. Spencer. Sharp concentration of the chromatic number in random graphs $G_{n,p}$. *Combinatorica* **7**, 121–130, 1987.