PROJET 20

Connectivité des graphes aléatoires

sujet proposé par L. Massoulié

laurent.massoulie@inria.fr

On s'intéresse à des graphes aléatoires dits d'Erdős-Rényi. Par définition un tel graphe de paramètres $n \in \mathbb{N}$ et $p \in [0,1]$, qu'on note $\mathcal{G}(n,p)$ est constitué de n sommets, qu'on identifie à $[n] := \{1,\dots,n\}$, et pour chaque paire non orientée de sommets $(u,v) \in [n]$, $u \neq v$, l'arc (u,v) est présent dans le graphe avec probabilité p, indépendamment de la présence des autres arcs.

Un graphe est connecté si et seulement si il est constitué d'une unique composante connexe, i.e. de chaque sommet u il existe un chemin d'arcs dans le graphe le reliant à tout autre sommet v.

On s'intéresse à la probabilité que le graphe $\mathcal{G}(n,p)$ soit connecté. Un résultat dû à Erdős et Rényi établit que pour toute constante $c \in \mathbb{R}$, on a:

$$\lim_{n\to\infty} \mathbb{P}\left(\mathcal{G}(n,(\ln(n)+c)/n) \text{ connecté}\right) = e^{-e^{-c}}. \tag{20.1}$$

Ceci montre que les graphes d'Erdős-Rényi $\mathcal{G}(n,p)$ sont connectés avec probabilité approchant 1 si le degré moyen des sommets $D := (n-1)p$ vérifie $D - \ln(n) \gg 1$ (i.e. c positif et grand), tandis qu'ils sont déconnectés avec probabilité approchant 1 si $\ln(n) - D \gg 1$ (i.e. c négatif et grand en valeur absolue). On considère une constante $c \in \mathbb{R}$ fixée, on pose $p = (\ln(n)+c)/n$ et on considère une réalisation G de $\mathcal{G}(n,p)$.

On dit qu'un sommet u est isolé s'il n'a aucun voisin (en d'autres termes, si son degré est nul). Pour tout $u \in [n]$ on pose $Z_u = \mathbf{1}_{u \text{ sommet isolé de } G}$. Enfin on note $X = \sum_{u\in[n]} Z_u$ le nombre des sommets isolés dans le graphe.

Pour un entier k fixe, on note $X^{\underline{k}} = X(X-1)\cdots(X-k+1)$.

T1. Montrer que pour tout $k \in \mathbb{N}$, on a

$$\begin{aligned} \mathbb{E}(X^{\underline{k}}) &= \textstyle\sum_{u_1,\dots u_k} \mathbb{E}(Z_{u_1}\cdots Z_{u_k}) \\ &= n^{\underline{k}}\mathbb{E}(Z_1\cdots Z_k) \\ &= n^{\underline{k}}(1-p)^{k(n-k)+\binom{k}{2}} \end{aligned}$$

où la somme porte sur toutes les suites d'entiers distincts $u_1,\dots,u_k$ de $[n]$, et en déduire:

$$\lim_{n\to\infty} \mathbb{E}(X^{\underline{k}}) = e^{-kc}.$$

On admettra que cette convergence pour tout $k \in \mathbb{N}$ des moments descendants $\mathbb{E}(X^{\underline{k}})$ de X vers ceux d'une variable aléatoire de Poisson de paramètre $\lambda = e^{-c}$ lorsque $n \to \infty$ entraîne la convergence en loi de X vers la loi de Poisson de paramètre e^{-c}.

T2. En déduire que la probabilité $\mathbb{P}(X = 0)$ que $\mathcal{G}(n, p)$ n'ait aucun sommet isolé vérifie

$$\lim_{n\to\infty} \mathbb{P}(X = 0) = e^{-e^{-c}}. \tag{20.2}$$

T3. Montrer que pour un entier $k > 1$, la probabilité que $\mathcal{G}(n, p)$ ait une composante connexe de taille k est majorée par:

$$\binom{n}{k}\mathbb{P}([k] \text{ composante connexe de } \mathcal{G}(n,p)) = \binom{n}{k}\mathbb{P}(\mathcal{G}(k,p) \text{ connecté})(1-p)^{k(n-k)}. \tag{20.3}$$

On peut alors majorer grossièrement $\mathbb{P}(\mathcal{G}(k, p)$ connecté$)$ par $T_k p^{k-1}$, où T_k est le nombre d'arbres sur $[k]$, et p^{k-1} est la probabilité que chacun des $k - 1$ arcs d'un arbre particulier sur $[k]$ est présent dans $\mathcal{G}(k, p)$.

Un théorème de Cayley établit que $T_k = k^{k-2}$; la borne donnée en (20.3) est donc à son tour majorée par:

$$\binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)}.$$

De cette majoration, on peut déduire que

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{G}(n,p) \text{ a une composante connexe de taille } k \in \{2, \dots, n-2\}) = 0.$$

T4. Etablir que cette dernière propriété, combinée avec le résultat (20.2) de la première question, implique le résultat d'Erdős-Rényi (20.1).

S1. Ecrire un programme qui détermine si un graphe non orienté G est connexe.

Réaliser une vingtaine de simulations de $\mathcal{G}(n, p)$ pour $n = 1000$ et pour chaque valeur de $p = k/n$, $k = 6.5,\ 7.0,\ 7.5,\ 8.0,\ 8.5,\ 9.0$.

S2. Comparer, pour chaque valeur de p, la fraction des graphes simulés qui sont connexes, et comparer cette fraction à $e^{-e^{-c}}$, pour $c = np - \ln(n)$. Le résultat asymptotique (20.1) d'Erdős-Rényi fournit-il une bonne approximation de $\mathbb{P}(\mathcal{G}(n, p)$ connecté$)$ pour les valeurs de n et p considérées?

On considère maintenant un graphe orienté $\mathcal{G}'(n, p)$ sur les n sommets $[n]$, où chaque arc orienté (u, v), $u \neq v$ est présent avec probabilité p, et ce indépendamment de la présence des autres arcs. Chaque sommet u du graphe a alors un degré entrant $d_u^{in} = \sum_{v\neq u} \mathbf{1}_{\text{arc } (v,u) \text{ présent}}$, et un degré sortant $d_u^{out} = \sum_{v\neq u} \mathbf{1}_{\text{arc } (u,v) \text{ présent}}$. On note alors X_{in} (respectivement, X_{out}) le nombre de sommets $u \in [n]$ de degré rentrant d_u^{in} (respectivement, sortant d_u^{out}) égal à zéro.

T5. Justifier brièvement que, pour $p = (\ln(n) + c)/n$ avec $c \in \mathbb{R}$ fixé, X_{in} et X_{out} ont une loi binomiale de paramètres $(n - 1, (1 - p)^{n-1})$, et que celle-ci tend vers la loi de Poisson de paramètre e^{-c} lorsque $n \to \infty$.

On dit qu'un graphe orienté est fortement connexe si pour toute paire de sommets distincts u, v, il existe un chemin orienté allant de u à v. Une condition nécessaire pour qu'un graphe soit fortement connexe est que chaque sommet u ait ses degrés entrant d_u^{in} et sortant d_u^{out} non nuls.

Un argumentaire semblable à celui fait pour les graphes non-orientés établit que, pour $p = (\ln(n) + c)/n$ et $c \in \mathbb{R}$ fixé, on a

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{G}'(n,p) \text{ connexe}) = \lim_{n\to\infty} \mathbb{P}(X_{out} = X_{in} = 0) = e^{-2e^{-c}}. \tag{20.4}$$

S3. Ecrire un programme déterminant si un graphe orienté est fortement connexe. Tester, pour $n = 1000$ et $p = k/n$, $k = 6.5,\ 7.0,\ 7.5,\ 8.0,\ 8.5$, la validité de la formule asymptotique (20.4) en simulant une vingtaine de graphes $\mathcal{G}'(n,p)$ pour les paramètres correspondants.
