Graphes aléatoires
E. Baur & M. Maazoun

TD 2 : Phase sous-critique v2

Exercice 1. *Application du cours*

Quelle est (avec grande probabilité) la taille de la plus grande composante connexe de $G_{n,m}$, pour $m \sim Cn^{\alpha}$, avec $\alpha < 1$ et $C > 0$?

Les deux exercices s'intéressent maintenant à la taille de la plus grande composante connexe pour $p = \frac{c}{n}$ avec $c < 1$.

Exercice 2. *Comparaison à Galton-Watson*

On veut majorer la taille de la plus grande composante connexe dans $G_{n,p}$ quand $p = \frac{c}{n}$ et $c < 1$. Pour $x \in [n]$, on explore la composante $\mathcal{C}(x)$ qui contient x de proche en proche de la manière suivante : $X_0 = 1$, et pour $i \geq 1$, X_i est le nombre de sommets dans $\mathcal{C}(x)$ à distance i de x.

1. Déduire qu'on peut construire des variables aléatoires Y_i telles que $X_i \leq Y_i$ p.s. pour tout i, où $(Y_i)_i$ est un processus de branchement correspondant à un arbre de Galton-Watson de loi $\mathrm{Bi}(n,p)$. (On rappelle qu'un processus de branchement de loi μ se construit en prenant $\xi_{i,k} \sim \mu$ i.i.d, $Y_0 = 1$ et $Y_{i+1} = \sum_{k=1}^{Y_i} \xi_{i,k}$.)
2. Montrer, pour $M \geq 1$, que $\mathbb{P}(|\mathcal{C}(x)| \geq M) \leq \mathbb{P}(\sum_{i=0}^{\infty} Y_i \geq M)$.
3. Montrer que si Y_i est processus de branchement de loi μ, $\mathbb{P}(\sum_{i=0}^{\infty} Y_i \geq M) \leq \mathbb{P}(\sum_{i=1}^{M} \beta_i \geq M-1)$, où les β_i sont i.i.d de loi μ. Dans notre cas, déduire qu'on s'est ramené à majorer $\mathbb{P}(\mathrm{Bi}(nM,p) \geq M-1)$.
4. On prend $M = \frac{3\log(n)}{(1-c)^2}$. Utiliser l'inégalité suivante (conséquence de l'inégalité de Bernstein, voir l'Exercice 4) :
$$\mathbb{P}(\mathrm{Bi}(n,p) \geq np + \lambda) \leq \exp\left(-\frac{\lambda^2}{2(np(1-p)+\lambda/3)}\right)$$
pour montrer que $\mathbb{P}(\mathcal{C}(x) \geq M) = o(\frac{1}{n})$.
5. Conclure qu'avec grande probabilité, la plus grande composante connexe de $G_{n,p}$ n'est pas plus grande que $\frac{3\log(n)}{(1-c)^2}$.

Exercice 3. *Arbres isolés*

Soit $\hat{X}_k$ le nombre de composantes isolées qui sont des arbres de taille k dans $G_{n,p}$.

1. Calculer $\mathbb{E}[\hat{X}_n]$ (fait en cours) et $\mathbb{E}[\hat{X}_n^2]$ (fait en cours uniquement pour un chemin de taille 2).
2. Pour $p = \frac{c}{n}$ avec $c < 1$, et $k = \lfloor \delta \log(n) \rfloor$, montrer par la méthode du second moment que pour $\delta = \delta(c)$ assez petit, on trouve avec grande probabilité un arbre isolé de taille k dans $G_{n,p}$.

Exercice 4. *Formule de Cayley pour le nombre d'arbres étiquetés*

Soit t_n le nombre de choix d'arêtes E sur $[n]$ tel que $([n], E)$ soit un arbre. La formule de Cayley affirme que

$$t_n = n^{n-2}.$$

La preuve suivante, dite par double comptage, est due à Jim Pitman.

1. On rappelle qu'une forêt est un graphe acyclique.
   (a) Montrer que dans un arbre T, $|V(T)| - |E(T)| = 1$. En déduire que le nombre de composantes connexes dans une forêt F est $|V(F)| - |E(F)|$.
   (b) Un enracinement d'une forêt F est le choix d'un sommet distingué par composante connexe. Une orientation d'une forêt F est le choix d'une orientation des arêtes telle que le degré sortant soit ≤ 1 partout. Montrer que pour une forêt F fixée ces notions sont en bijection (commencer par montrer qu'on peut associer une orientation à tout enracinement, faire un dessin).
2. On fixe $n \in \mathbb{N}$. Soit a_n le nombre de suites strictement croissantes

$$E_0 \subsetneq E_1 \subsetneq E_2, \ldots, E_{n-1} \subset [n] \times [n]$$

   de forêts orientées sur $[n]$ telles que
   — pour tout $i \in [\![0, n-1]\!]$, $|E_i| = i$,
   — E_{n-1} est un arbre.
   On va calculer a_n de deux manières différentes.
   (a) Combien y a-t-il de choix pour E_{n-1} ? En déduire que $a_n = t_n n!$.
   (b) Connaissant E_i, E_{i+1} s'obtient par l'ajout d'une arête orientée. Où peut-on mettre l'origine et la destination de cette arête ? Déduire qu'il y a $n(n-i-1)$ choix pour E_{i+1} sachant E_i. Déduire une autre formule pour a_n et conclure.

Exercice 5. *Inégalité de Bernstein*

Soient $X_1, \ldots, X_n$ des variables aléatoires réelles indépendantes telles que $|X_i - \mathbb{E}\, X_i| \leq 1$. On pose $S = \sum_{i=1}^n X_i$, et $\sigma^2 = \mathrm{Var}(S)$. On veut démontrer l'inégalité suivante : Pour $t \in \mathbb{R}$,

$$\mathbb{P}\left(S - \mathbb{E}\, S \geq t\right) \leq \exp\left(-\frac{t^2}{2(\sigma^2 + t/3)}\right).$$

1. Montrer qu'on peut supposer les X_i centrées.
2. Comme souvent dans ce genre d'inégalités (Bernstein, Hoeffding, Azuma,...), on commence par appliquer l'inégalité de Markov à la variable $e^{\lambda X}$ (on appelle ça l'inégalité de Chernoff), en supposant pour l'instant $\lambda > 0$ arbitraire. Montrer alors que $\mathbb{P}(S \geq t) \leq e^{-\lambda t} \prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}]$.
3. On pose $\sigma_i^2 = \mathrm{Var}(X_i)$. Développer $\mathbb{E}[e^{\lambda X_i}]$ en série, et utiliser le fait que X_i est centré et de module ≤ 1 pour obtenir

$$\mathbb{E}[e^{\lambda X_i}] \leq 1 + (e^\lambda - \lambda - 1)\sigma_i^2 \leq \exp((e^\lambda - \lambda - 1)\sigma_i^2).$$

4. Réunir 2. et 3., optimiser en λ et utiliser l'inégalité $(1+x)\log(1+x) - x \geq \frac{x^2}{2(1+x/3)}$ pour conclure.

N.B. : En remplacent X_i par $-X_i$ et en applicant l'inégalité de Bernstein encore une fois, on obtient pour $t \geq 0$

$$\mathbb{P}\left(|S - \mathbb{E}\, S| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2(\sigma^2 + t/3)}\right).$$