11.07B

SESSION 2011

Filière BCPST

MATHÉMATIQUES

Épreuve commune aux ENS de Paris, Lyon et Cachan

Durée : 4 heures

L'usage de calculatrice est interdit.

Ce sujet traite d'un modèle de réseau ou *graphe* aléatoire. Un graphe est composé d'un ensemble de nœuds ou *sommets* et d'un ensemble de connections ou *arêtes*, chacune d'elles reliant deux sommets distincts. Les graphes ont de nombreuses applications en biologie, par exemple pour modéliser les réseaux neuronaux, les réseaux de gènes (c.-à-d. le graphe indiquant les interactions inhibitrices ou excitatrices entre protéines), ou les réseaux d'interaction en écologie ou en épidémiologie (réseaux d'interaction ou de reproduction, ou réseaux de transmission de pathogènes entre hôtes). Dans les exemples précédents, chaque sommet représente un neurone, une protéine ou un individu, et chaque arête désigne les paires de sommets reliées par le type d'interaction considéré (axones, interactions chimiques, sexuelles ou infectieuses).

Le sujet est divisé en cinq parties non-indépendantes, de longueur variable, et de difficulté dans l'ensemble croissante. Les candidats pourront traiter le sujet dans l'ordre qu'ils souhaitent.

Il est recommandé de lire l'ensemble de l'énoncé attentivement. Il est demandé de veiller au soin de la présentation, ainsi qu'à la rigueur et à la concision des raisonnements.

Notations

L'ensemble des entiers naturels est noté $\mathbb{N}$ et l'ensemble des entiers naturels non nuls $\mathbb{N}^*$.

On emploiera les abréviations v.a. pour variable(s) aléatoire(s) et i.i.d. pour indépendantes et identiquement distribuées. On utilisera également les notations $\mathrm{Var}(X)$ et $\mathrm{Cov}(X,Y)$ pour la variance de la v.a. X et la covariance des v.a. X et Y.

On utilisera la notation $\mathbb{1}_A$ pour la v.a. valant 1 si l'événement A est réalisé, et 0 sinon.

La fonction **logarithme (népérien)** sera notée **log**. La fonction exponentielle sera notée indifféremment e^x ou $\exp(x)$.

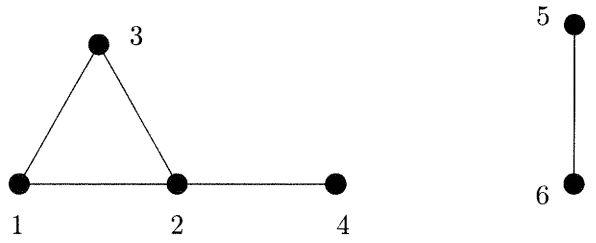
On rappelle la formule des coefficients binomiaux $C_n^k = \frac{n!}{k!(n-k)!}$ pour $0 \leqslant k \leqslant n$, où $n! = 1 \times 2 \times \ldots \times n$ si $n \geqslant 1$, et $0! = 1$. Si (u_n) et (v_n) sont deux suites réelles, on rappelle également la notation $u_n = o(v_n)$ lorsque $u_n/v_n \to 0$ quand $n \to +\infty$.

Dans tout le sujet, on posera $\mathbf{S} = \{1, 2, \ldots, n\}$ pour $n \in \mathbb{N}^*$, et on appellera $\mathbf{S}$ l'ensemble des **sommets** (ou nœuds) du graphe. Pour tout $i, j \in \mathbf{S}$ avec $i \neq j$, on note $\{i, j\}$ la paire formée par i et j. On rappelle qu'une paire n'est pas ordonnée : $\{i, j\} = \{j, i\}$. Le graphe est constitué d'un ensemble d'**arêtes**, noté $\mathbf{A}$, qui est un sous-ensemble de l'ensemble de toutes les paires $\{i, j\}$ pour $i \neq j \in \mathbf{S}$. Si $\{i, j\} \in \mathbf{A}$, on dit que l'**arête** $\{i, j\}$ **appartient au graphe**, et que les sommets i et j sont **connectés** (voir la figure ci-dessous pour une représentation graphique d'un graphe).

Pour chaque sommet $i \in \mathbf{S}$, on appelle **voisins** de i les sommets de $\mathbf{S}$ connectés à i par une arête de $\mathbf{A}$. Pour tout sommets $i \neq j \in \mathbf{S}$, on dit que i est **relié** à j s'il existe un chemin d'arêtes de $\mathbf{A}$ joignant i à j, c.-à-d. s'il existe $n \geqslant 1$ et $i_0, i_1, \ldots, i_n \in \mathbf{S}$ tels que

$$i_0 = i, \quad i_n = j \quad \text{et} \quad \forall k \in \{0, 1, \ldots, n-1\}, \quad \{i_k, i_{k+1}\} \in \mathbf{A}.$$

Par convention, on dira également que i est relié à lui-même. Pour tout $i \in \mathbf{S}$, on appelle **composante connexe de** i, ou plus simplement **cluster de** i, et on note $\mathcal{C}_i$, l'ensemble des sommets de $\mathbf{S}$ reliés à i. On notera enfin $|\mathcal{C}_i|$ le nombre d'éléments de $\mathcal{C}_i$.



À titre d'exemple, la figure ci-dessus représente le graphe donné par

$$\mathbf{S} = \{1, 2, 3, 4, 5, 6\} \quad \text{et} \quad \mathbf{A} = \Big\{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{2, 4\}, \{5, 6\}\Big\}.$$

Les sommets sont représentés par des points et les arêtes par des segments reliant deux points.

Dans ce graphe, le sommet 1 est voisin du sommet 2, mais pas du sommet 4. Par contre, le sommet 4 est relié au sommet 1. Le cluster du sommet 1 est l'ensemble $\mathcal{C}_1 = \{1, 2, 3, 4\}$, le cluster de 5 est $\mathcal{C}_5 = \{5, 6\}$, et on a $|\mathcal{C}_1| = 4$ et $|\mathcal{C}_5| = 2$. On a également $\mathcal{C}_1 = \mathcal{C}_2 = \mathcal{C}_3 = \mathcal{C}_4$ et $\mathcal{C}_5 = \mathcal{C}_6$.

Étude asymptotique de la plus grande composante connexe dans un modèle de graphe aléatoire

Soit $n \in \mathbb{N}^*$. On considère le graphe dont l'ensemble des sommets est $\mathbf{S} = \{1, 2, \ldots, n\}$. On suppose que chaque paire $\{i, j\}$ de sommets avec $i \neq j$ appartient à $\mathbf{A}$ avec probabilité $p \in]0, 1[$, indépendamment des autre paires. En particulier, un sommet peut être voisin de n'importe quel autre sommet. On précise également qu'il n'y a pas d'arête connectant un sommet à lui-même. Pour toute paire de sommets $\{i, j\}$ avec $i \neq j$, on note $Z_{i,j}$ la variable aléatoire valant 1 si $\{i, j\} \in \mathbf{A}$ et 0 sinon.

Le but du sujet est l'étude de la taille du plus grand cluster du graphe lorsque n est grand.

I Caractérisation algorithmique de la taille d'un cluster

1) Exprimer en fonction des v.a. $Z_{i,j}$ le nombre de voisins d'un sommet $i \in \mathbf{S}$ fixé. Quel est le nombre maximal de voisins de i ? Quelle est la loi du nombre de voisins de i ? Son espérance ? Sa variance ?

On considère l'algorithme (ou procédure) suivant(e), qui consiste, au cours d'une succession d'étapes $k = 0, 1, 2, \ldots$, à assigner différents états aux sommets du graphe, parmi les trois états *neutre*, *actif* et *inactif* :

- Phase initiale (étape $k = 0$) : tous les sommets du graphe sont dans l'état *neutres*, sauf le sommet 1, qui est dans l'état *actif*. On passe ensuite à l'étape $k = 1$.
- Etape numéro $k \geqslant 1$:
  - Soit i_k le sommet *actif* de plus petit numéro.
  - Tous les voisins de i_k qui sont *neutres* passent à l'état *actif*.
  - Le sommet i_k passe à l'état *inactif*.
- On recommence l'étape précédente en remplaçant k par $k + 1$. L'algorithme s'arrête lorsqu'il ne reste plus de sommet actif.

2) Expliquer pourquoi l'ensemble des sommets inactifs à la fin de l'algorithme forme le cluster du sommet 1. Représenter graphiquement les étapes de l'algorithme sur le graphe de la figure de la page précédente (on pourra utiliser les abréviations N, A et I pour neutre, actif et inactif). À la fin de quelle étape l'algorithme s'arrête-t-il ?

3) On note A_k le nombre de sommets actifs à la fin de l'étape $k \geqslant 0$ de l'algorithme, et K_0 le nombre total (aléatoire) d'étapes dans l'algorithme. On pose par convention $A_k = 0$ pour $k > K_0$.

a) Que vaut A_0 ? Que vaut A_{K_0} ? Quelle est la probabilité que $A_1 = 0$?

b) Lorsque $k \leqslant K_0$, combien y a-t-il de sommets inactifs à la fin de l'étape k de l'algorithme ? Exprimer la taille $|\mathcal{C}_1|$ du cluster du sommet 1 en fonction de K_0. Quelle est la relation entre K_0 et le premier entier k tel que $A_k = 0$?

c) Pour toute v.a. N dans $\mathbb{N}^*$, on dit que la v.a. X suit la loi $\mathrm{Bin}(N, p)$ si pour tout $m \geqslant 1$, conditionnellement à $\{N = m\}$, X suit une loi binomiale de paramètres (m, p), c.-à-d.

$$\mathbb{P}(X = k \mid N = m) = C_m^k \, p^k \, (1-p)^{m-k}, \qquad \forall k \in \{0, 1, \ldots, m\}.$$

Soit $k \geqslant 1$ fixé. Montrer que, si $k \leqslant K_0$,

$$A_k = A_{k-1} + B_k - 1, \quad \text{où} \quad B_k \quad \text{est une v.a. de loi} \quad \mathrm{Bin}(n - k + 1 - A_{k-1}, p).$$

II Étude d'une approximation déterministe de la taille d'un cluster

Dans tout le reste du sujet, on supposera que p dépend de n comme $p = \dfrac{\lambda}{n}$, avec $\lambda > 0$ fixé.

4) Soit N une v.a. dans $\mathbb{N}^*$ d'espérance finie et X une v.a. de loi $\mathrm{Bin}(N, q)$, avec $q \in]0,1[$, (voir la définition à la question **3)c)**). Montrer que $\mathbb{E}(X) = \mathbb{E}(N)q$.

5) On considère la suite définie par récurrence

$$a_0 = 1 \quad \text{et} \quad a_k = \left(1 - \frac{\lambda}{n}\right) a_{k-1} + \lambda - 1 - \lambda\frac{k-1}{n} \quad \text{pour } k \geqslant 1.$$

En utilisant la question **3)c)**, expliquer pourquoi a_k peut être considéré comme une approximation de $\mathbb{E}(A_k)$, mais que les deux quantités ne sont a priori pas identiques.

6) Écrire la relation de récurrence satisfaite par $u_k = a_k + \alpha + \beta k$ pour α et β réels fixés. En choisissant convenablement α et β, montrer que

$$a_k = n - k - \left(1 - \frac{\lambda}{n}\right)^k (n-1) \qquad \text{pour tout } k \geqslant 0.$$

7) Soit la fonction $f_n(y) = n - y - (n-1)\, e^{-b_n y}$, où $b_n = -\log(1 - \frac{\lambda}{n})$.

a) Tracer le tableau de variation de f_n. Lorsque $\lambda \neq 1$, montrer que, pour tout n suffisamment grand, f_n prend sa valeur maximale en un point du même signe que $\lambda - 1$.

b) Montrer que

$$e^{-b_n \log n} = 1 - \frac{\lambda \log n}{n} + o\left(\frac{\log n}{n}\right),$$

et en déduire un équivalent de $f_n(\log n)$.

c) Soit $\alpha \in [0,1]$. Calculer $g(\alpha) = \lim\limits_{n \to +\infty} \dfrac{f_n(\alpha n)}{n}$. Dans le cas où $\lambda > 1$, tracer le tableau de variation de g sur l'intervalle $[0,1]$. Montrer qu'il existe un unique $\alpha^* \in]0,1[$ tel que $g(\alpha^*) = 0$.

8) Comme le suggère le fait que $|\mathcal{C}_1|$ est égal au premier indice k tel que $A_k = 0$, on admet *dans cette question uniquement* que la taille du cluster de 1 est proche du premier entier $k \leqslant n$ tel que $a_k \leqslant 0$. Déduire des questions précédentes des bornes sur $|\mathcal{C}_1|$ en fonction du signe de $\lambda - 1$ lorsque $n \to +\infty$. Si l'on considère que le graphe représente le réseau de transmission d'un agent pathogène infectant initialement un seul individu dans une population de n individus, que signifient ces résultats en terme de propagation de l'épidémie ?

III Quelques propriétés d'une marche aléatoire

Soit $m \in \mathbb{N}^*$ et $q \in]0,1[$ fixés. On considère une suite $(X_k)_{k \geqslant 1}$ de v.a. i.i.d. telle que la loi de $X_1 + 1$ (et donc de $X_k + 1$ pour tout $k \geqslant 1$) est binomiale de paramètres (m, q). On définit la suite $(S_k)_{k \geqslant 0}$ par

$$S_0 = 1 \quad \text{et} \quad S_k = S_0 + X_1 + X_2 + \ldots + X_k = S_{k-1} + X_k, \quad \forall k \geqslant 1.$$

Soit enfin K la v.a. définie comme le premier entier k tel que $S_k \leqslant 0$. Si $S_k > 0$ pour tout $k \geqslant 0$, on pose $K = +\infty$.

9) Justifier que K est également le premier entier k tel que $S_k = 0$.

10) Calculer $\phi(x) = \mathbb{E}[\exp(x(S_k - S_{k-1}))]$ pour $x \in \mathbb{R}$. Montrer que $\phi(x) \leqslant \psi(x)$, où $\psi(x) = \exp(-x + mq(e^x - 1))$.

11) On suppose dans cette question que $mq < 1$.

a) En utilisant la loi des grands nombres, montrer que $\mathbb{P}(K < +\infty) = 1$.

b) Trouver le point x^* où ψ est minimale. Calculer θ tel que $\psi(x^*) = e^{-\theta}$. Quel est le signe de θ ?

c) On pose $M_k = \exp(x^* S_k + \theta k)$ pour tout $k \geqslant 0$. Montrer que

$$\mathbb{E}\left[\mathbb{1}_{\{K \geqslant k-1\}}(M_k - M_{k-1})\right] = \mathbb{E}\Big[\exp(x^*(S_k - S_{k-1}) + \theta) - 1\Big]\, \mathbb{E}(\mathbb{1}_{\{K \geqslant k-1\}} M_{k-1})$$

$$\text{et} \quad \mathbb{E}\left[\mathbb{1}_{\{K = k-1\}}(M_k - M_{k-1})\right] = \mathbb{E}\Big[\exp(x^*(S_k - S_{k-1}) + \theta) - 1\Big]\, \mathbb{E}(\mathbb{1}_{\{K = k-1\}} M_{k-1}).$$

d) Montrer que $\mathbb{E}\left[\mathbb{1}_{\{K \geqslant k\}}(M_k - M_{k-1})\right] \leqslant 0$.

e) En déduire que $\mathbb{E}(M_K) \leqslant \mathbb{E}(M_0)$ (on pourra intervertir espérance et sommation sans justification).

f) Conclure que $\mathbb{P}(K \geqslant k) \leqslant \exp(x^* - \theta k)$.

12) On suppose maintenant que $mq > 1$.

a) Montrer qu'il existe un unique $y^*(mq) > 0$, ne dépendant que de mq, tel que $\psi(-y^*(mq)) = 1$.

b) Montrer que $y^*(\alpha)$ est une fonction croissante de α sur l'intervalle $]1, +\infty[$.

c) Soit $M'_k = \exp(-y^*(mq) S_k)$. On admet l'inégalité $\mathbb{E}\left[\mathbb{1}_{\{K \geqslant k\}}(M'_k - M'_{k-1})\right] \leqslant 0$ pour tout $k \geqslant 0$. En sommant cette inégalité pour k variant de 1 à $i \in \mathbb{N}^*$ fixé, montrer que

$$\mathbb{P}(K \leqslant i) \leqslant e^{-y^*(mq)} \quad \text{puis que} \quad \mathbb{P}(K < +\infty) \leqslant e^{-y^*(mq)}.$$

d) On suppose dans cette question que $S_0 = j \in \mathbb{N}^*$. Justifier que $\mathbb{P}(K < +\infty) \leqslant e^{-y^*(mq)j}$. Justifier également cette inégalité si S_0 est une v.a. indépendante de la suite $(X_i)_{i \geqslant 1}$, telle que $S_0 \geqslant j$. (On introduira, pour tout $i \leqslant j$, $K^{(i)}$ le premier entier k tel que $S_k = i$.)

13) Dans cette question également, on suppose que $mq > 1$.

a) Montrer que $\mathbb{E}[\exp(x(S_k - S_0 + k))] \leqslant \exp((e^x - 1)mqk)$ pour $k \geqslant 1$ et $x \in \mathbb{R}$.

b) En déduire que, pour tout $x > 0$, $\mathbb{P}(S_k - S_0 + k \geqslant 2mqk) \leqslant \exp\left((-2x + e^x - 1)mqk\right)$.

c) En prenant $x = \log 2$, en déduire que

$$\mathbb{P}(S_k - S_0 + k \geqslant 2mqk) \leqslant \exp\Big(-\frac{mqk}{3}\Big)$$

(on admettra l'inégalité $1 - 2\log 2 \leqslant -1/3$).

d) Montrer que, pour tout $x \leqslant 0$, $e^x \leqslant 1 + x + \dfrac{x^2}{2}$.

e) Soit $\alpha > 0$ tel que $(1-\alpha)mq > 1$. Montrer que, pour tout $x < 0$,

$$\mathbb{P}\left(S_k - S_0 + k \leqslant (1-\alpha)mqk\right) \leqslant \exp\left[(e^x - 1 - (1-\alpha)x)\, mqk\right],$$

puis en prenant $x = -\alpha$ que

$$\mathbb{P}\left(S_k - S_0 + k \leqslant (1-\alpha)mqk\right) \leqslant \exp\Big(-\frac{\alpha^2 mqk}{2}\Big).$$

IV Taille du plus grand cluster lorsque $\lambda < 1$

Le but de cette partie et de la suivante est de démontrer rigoureusement les résultats de croissance de la taille du plus grand cluster suggérés par l'approximation déterministe de la partie II. On suppose dans cette partie que $\lambda < 1$.

14) On fixe $q \in]0,1[$. Soit N une v.a. à valeurs dans $\{1, 2, \ldots, n\}$, et $(X_k)_{k\geqslant 1}$ une suite de v.a. i.i.d. indépendante de N de loi de Bernoulli de paramètre q. Construire à partir de N et (X_k) une v.a. X dont on montrera qu'elle est de loi $\mathrm{Bin}(N, q)$, où $\mathrm{Bin}(N, q)$ est définie à la question **3)c)**.

On admettra que cette v.a. X est telle que, pour toute v.a. Y de loi $\mathrm{Bin}(n-N, q)$ indépendante de la suite $(X_k)_{k\geqslant 1}$, la v.a. $X + Y$ suit la loi binomiale de paramètre (n, q).

15) a) Soit $(X_{k,l})_{k\geqslant 1,\, l\geqslant 1}$ une famille de v.a. i.i.d. de loi de Bernoulli de paramètre $\frac{\lambda}{n}$ et indépendantes de la suite $(A_k)_{k\geqslant 0}$. On pose $S_0 = 1$ et pour tout $k \geqslant 1$,

$$S_k - S_{k-1} = \begin{cases} A_k - A_{k-1} + \displaystyle\sum_{l=1}^{A_{k-1}+k-1} X_{k,l} & \text{si } k < K_0 \\ \displaystyle\sum_{l=1}^{n} X_{k,l} - 1 & \text{si } k \geqslant K_0. \end{cases}$$

En utilisant la question **3)c)**, montrer que

- $S_0 = A_0$,
- $A_k \leqslant S_k$ pour tout $k \leqslant K_0$,
- $(S_k - S_{k-1} + 1)_{k\geqslant 1}$ est une suite de v.a. i.i.d. de loi binomiale de paramètres $(n, \frac{\lambda}{n})$,
- La suite $(S_k - A_k)_{k\geqslant 0}$ est croissante tant que $k \leqslant K_0$.

b) Soit K le premier entier $k \geqslant 0$ tel que $S_k = 0$. Donner une inégalité reliant K et K_0.

16) Soit $\varepsilon > 0$ fixé. On rappelle qu'on suppose dans cette partie $\lambda < 1$.

a) En utilisant la question **11)**, montrer que $\mathbb{P}\left(|\mathcal{C}_1| > \dfrac{1+\varepsilon}{\theta}\log n\right) \leqslant \dfrac{1}{\lambda n^{1+\varepsilon}}$, où θ est défini à la question **11)b)**.

b) Pour tout $i \in \{1, 2, \ldots, n\}$, justifier l'égalité

$$\mathbb{P}\left(|\mathcal{C}_1| > \frac{1+\varepsilon}{\theta}\log n\right) = \mathbb{P}\left(|\mathcal{C}_i| > \frac{1+\varepsilon}{\theta}\log n\right).$$

En déduire que $\displaystyle\lim_{n\to+\infty} \mathbb{P}\left(\exists i \in \{1, 2, \ldots, n\} \text{ tel que } |\mathcal{C}_i| > \frac{1+\varepsilon}{\theta}\log n\right) = 0$

17) On suppose que le graphe représente un réseau de transmission d'un agent pathogène dans une grande population, et que le pathogène infecte initialement un seul individu dans une grande

population. Interpréter le résultat précédent en terme de propagation de l'épidémie. Relier cette interprétation au nombre moyen de voisins de chaque site dans le graphe.

V Taille du plus grand cluster lorsque $\lambda > 1$

On suppose maintenant que $\lambda > 1$ jusqu'à la fin du problème. Dans toute cette partie, on fixe $\beta > 0$ tel que $\beta \log n$ est entier, et $\alpha > 0$ tel $(1-\alpha)^2\lambda > 1$ et $(1-\alpha)n$ est entier. On pose également

$$\gamma = \frac{(1-\alpha)\lambda - 1}{2} > 0. \qquad (*)$$

18) Dans cette question et les suivantes, la suite $(S_k)_{k\geqslant 0}$ est celle construite à la question **15)a)**.

a) Soit X une v.a. binomiale de paramètres (m, q), où $m \in \mathbb{N}^*$ et $q \in]0,1[$. Soit $j \leqslant m$. Montrer

$$\mathbb{P}(X \geqslant j) = q^j\, m(m-1)\dots(m-j+1) \sum_{i=0}^{m-j} \frac{1}{(i+1)(i+2)\dots(i+j)}\, C_{m-j}^i\, q^i\, (1-q)^{m-j-i}.$$

En déduire que $\mathbb{P}(X \geqslant j) \leqslant (mq)^j$.

b) Soit $k \in \mathbb{N}^*$ fixé. Déterminer la loi de la v.a.

$$X = \sum_{l=1}^{2\lambda(k+1)} \sum_{j=1}^{k} X_{j,l}.$$

Montrer que, si $S_k - S_0 + k < 2\lambda k$, alors on a $A_l + l \leqslant 2\lambda(k+1)$ pour tout $l \leqslant k$ et $l \leqslant K_0$, puis déduire de la question **15)a)** que, sur l'événement $\{S_k - S_0 + k < 2\lambda k\}$, on a $S_k - A_k \leqslant X$ si $k < K_0$, ou $S_{K_0} - A_{K_0} \leqslant X$ si $K_0 \leqslant k$.

c) Montrer que

$$\mathbb{P}\Big(|\mathcal{C}_1| \leqslant \beta \log n\Big) \leqslant \mathbb{P}(K \leqslant \beta \log n) + \mathbb{P}\Big(S_{\beta\log n} - S_0 \geqslant (2\lambda - 1)\beta \log n\Big)$$
$$+ \mathbb{P}\Big(S_{\beta\log n} - S_0 < (2\lambda-1)\beta\log n,\ K_0 \leqslant \beta \log n \text{ et } A_{K_0} \neq S_{K_0}\Big).$$

d) Soit $\varepsilon > 0$ fixé. Déduire des questions précédentes et des questions **12)c)** et **13)c)** que, pour tout n suffisamment grand

$$\mathbb{P}\Big(|\mathcal{C}_1| \leqslant \beta \log n\Big) \leqslant e^{-y^*(\lambda)} + \varepsilon$$

19) On rappelle que la constante γ est définie par l'équation (*) ci-dessus.

a) Déduire des questions **18)a)** et **b)** que, pour n suffisamment grand,

$$\mathbb{P}\Big(K_0 > \beta \log n,\ S_{\beta\log n} - S_0 \in]2\gamma\beta\log n, (2\lambda-1)\beta\log n[\text{ et } A_{\beta\log n} \leqslant \gamma\beta\log n\Big) \leqslant n^{-\frac{\gamma\beta\log n}{2}}$$

(on utilisera l'inégalité $2\lambda^2\beta\log n(\beta\log n + 1) \leqslant \sqrt{n}$, valide pour n assez grand).

b) Déduire de la question **13)** que

$$\mathbb{P}\Big(K_0 > \beta\log n \text{ et } A_{\beta\log n} \leqslant \gamma\beta\log n\Big) \leqslant n^{-\lambda\beta/3} + n^{-\alpha^2\lambda\beta/2} + n^{-\gamma\beta\log n/2}$$

20) Dans cette question, on se place sur l'événement $E = \{K_0 > \beta \log n \text{ et } A_{\beta \log n} > \gamma\beta \log n\}$. On note K_α le premier indice k tel que $A_k + k \geqslant \alpha n$. S'il n'existe pas de tel entier, on pose $K_\alpha = K_0$.

a) Montrer que, pour tout $k \leqslant K_\alpha$, il reste plus de $(1-\alpha)n$ sommets neutres à l'étape k de l'algorithme de la question **1)**.

En particulier, la v.a. B_k est plus grande qu'une v.a. binomiale de paramètres $(n(1-\alpha), \frac{\lambda}{n})$ tant que $k \leqslant K_\alpha$. On admet donc qu'il existe une suite (S'_k) pour $k \geqslant \beta \log n$ telle que $S'_{\beta \log n} > \gamma\beta \log n$ et, lorsque l'événement E est réalisé,

- $S'_{\beta \log n} = A_{\beta \log n}$,
- $A_k \geqslant S'_k$ pour tout $k \in \{\beta \log n, \beta \log n + 1, \ldots, K_\alpha\}$,
- $(S'_k - S'_{k-1} + 1)_{k \geqslant \beta \log n}$ est une suite de variables aléatoires i.i.d. de loi binomiale de paramètres $((1-\alpha)n, \frac{\lambda}{n})$.

b) Soit K' le premier indice $k \geqslant \beta \log n$ tel que $S'_k = 0$. S'il n'existe pas de tel entier, on pose $K' = +\infty$. Déduire de la question **12)d)** que

$$\mathbb{P}\left(K' < +\infty\right) \leqslant n^{-y^*((1-\alpha)\lambda)\gamma\beta}.$$

c) Donner et justifier une inégalité entre K_0 et $\min\{K', K_\alpha\}$, où $\min\{x, y\}$ désigne le plus petit des nombres x et y.

d) Soit $a \in [1, 2]$ tel que $an^{2/3}$ est entier. Montrer que, pour n suffisamment grand, lorsque l'événement

$$\{S_{an^{2/3}} - S_0 \leqslant (2\lambda - 1)an^{2/3}\}$$

est réalisé, on a $K_\alpha > an^{2/3}$.

e) Soit $\delta = \dfrac{((1-\alpha)^2\lambda - 1)a}{2}$. Déduire de la question **13)e)** que

$$\mathbb{P}\Big(S'_{an^{2/3}} - S'_{\beta \log n} \leqslant \delta n^{2/3}\Big) \leqslant \exp\Big(-\frac{\alpha^2(1-\alpha)\lambda a}{4} n^{2/3}\Big)$$

(on utilisera l'inégalité $\beta \log n \leqslant \frac{a}{2} n^{2/3}$, valable pour n assez grand).

f) Déduire des questions précédentes que

$$\mathbb{P}\Big(\Big\{K_0 < an^{2/3} \text{ ou } A_{an^{2/3}} < \delta n^{2/3}\Big\} \cap E\Big)$$
$$\leqslant n^{-y^*((1-\alpha)\lambda)\gamma\beta} + \exp\Big(-\frac{\lambda a n^{2/3}}{3}\Big) + \exp\Big(-\frac{\alpha^2(1-\alpha)\lambda a}{4} n^{2/3}\Big).$$

21) Déduire des questions **19)** et **20)** que, si l'on choisit β supérieur à une certaine constante indépendante de n, alors, lorsque n tend vers l'infini,

$$\mathbb{P}\Big(|\mathcal{C}_1| \leqslant \beta \log n \quad \text{ou} \quad \Big(|\mathcal{C}_1| \geqslant an^{2/3} \text{ et } A_{an^{2/3}} \geqslant \delta n^{2/3}\Big)\Big) \geqslant 1 - o\Big(\frac{1}{n}\Big).$$

22) On note $A_k^{(i)}$ le nombre de sommets actifs dans l'algorithme de la question **1)**, lorsqu'il démarre à l'étape 0 avec pour unique sommet actif le sommet i (au lieu du sommet 1). Montrer que

$$\lim_{n \to +\infty} \mathbb{P}\Big(\forall i \in \{1, \ldots, n\},\ |\mathcal{C}_i| \leqslant \beta \log n \quad \text{ou} \quad \Big(|\mathcal{C}_i| \geqslant an^{2/3} \text{ et } A^{(i)}_{an^{2/3}} \geqslant \delta n^{2/3}\Big)\Big) = 1.$$

23) Le but de cette question est de montrer qu'avec grande probabilité il existe au plus un seul cluster de grande taille dans le graphe.

a) Soient i et j deux sommets distincts du graphe. On suppose que $|\mathcal{C}_i|$ et $|\mathcal{C}_j|$ sont plus grands que $an^{2/3}$, et que $\mathcal{C}_i \neq \mathcal{C}_j$. Si l'on fait partir l'algorithme de la question **1)** depuis les sommets i et j, les deux algorithmes obtenus ne se sont pas encore arrêtés à l'étape $an^{2/3}$. On suppose de plus que $A^{(i)}_{an^{2/3}} \geqslant \delta n^{2/3}$ et $A^{(j)}_{an^{2/3}} \geqslant \delta n^{2/3}$. Donner une minoration du nombre de paires $\{k, l\} \notin \mathbf{A}$, où k et l sont des sommets actifs à l'étape $an^{2/3}$ des deux algorithmes.

b) En déduire que

$$\mathbb{P}\left(|\mathcal{C}_i| \geqslant an^{2/3},\ |\mathcal{C}_j| \geqslant an^{2/3},\ \mathcal{C}_i \neq \mathcal{C}_j,\ A^{(i)}_{an^{2/3}} \geqslant \delta n^{2/3},\ A^{(j)}_{an^{2/3}} \geqslant \delta n^{2/3}\right) \leqslant \exp\left(-\lambda\delta^2 n^{1/3}\right).$$

c) En utilisant la question **22)**, montrer que

$$\lim_{n\to+\infty} \mathbb{P}\Big(\text{il existe au plus un seul cluster de taille supérieure à } \beta\log n\Big) = 1.$$

24) Soit $Y_i = \mathbb{1}_{\{|\mathcal{C}_i| \leqslant \beta\log n\}}$. Pour tout $k \leqslant n$, on définit $\mathcal{C}_i^{(k)}$ comme le cluster du sommet i dans le graphe obtenu en retirant du graphe initial les k derniers sommets $n-k+1, \ldots, n$ et toutes les arêtes partant de ces sommets. On pose

$$p_k = \mathbb{P}(|\mathcal{C}_i^{(k)}| \leqslant \beta\log n).$$

En particulier, $\mathbb{E}Y_i = p_0$. On remarque que p_k est indépendant du choix de $i \leqslant n-k$, puisque la loi du graphe ne dépend pas du choix de numérotation des sommets.

a) Montrer que $p_0 \leqslant p_k$ pour tout $k \leqslant n$.

b) Montrer que $\mathbb{P}\left(|\mathcal{C}_1^{(k)}| \neq |\mathcal{C}_1| \;\middle|\; |\mathcal{C}_1^{(k)}| \leqslant \beta\log n\right) \leqslant 1 - \left(1 - \dfrac{\lambda}{n}\right)^{\beta k \log n}$.

c) En déduire que $p_k \leqslant p_0\left(1 - \frac{\lambda}{n}\right)^{-\beta k \log n}$.

d) En utilisant le fait que la loi du graphe ne change pas lorsque l'on change la numérotation des sommets, montrer que $\mathbb{P}(Y_1 = Y_2 = 1) \leqslant \mathbb{P}(\mathcal{C}_1 = \mathcal{C}_2 \text{ et } Y_1 = 1) + \displaystyle\sum_{k=1}^{\beta\log n} \mathbb{P}(|\mathcal{C}_1| = k)p_k$.

e) En déduire que $\mathbb{P}(Y_1 = Y_2 = 1) \leqslant p_0\dfrac{\beta\log n}{n} + p_0^2\left(1 - \dfrac{\lambda}{n}\right)^{-\beta^2(\log n)^2}$.

f) En déduire une majoration de $\mathrm{Cov}(Y_1, Y_2)$, puis que, pour tout n suffisamment grand, $\mathrm{Var}\left(\sum_{i=1}^n Y_i\right) \leqslant Cn(\log n)^2$ pour une constante C indépendante de n (on utilisera les inégalités $\log(1-x) \geqslant -2x$ et $e^x \leqslant 1 + 2x$, valides pour tout $x > 0$ suffisamment petit).

g) Déduire de l'inégalité de Tchebychev et des questions **18)d)** et **23)c)** que, pour tout $\varepsilon > 0$, le plus grand cluster du graphe a une taille supérieure à $(1 - e^{-y^*(\lambda)} - \varepsilon)n$ avec probabilité convergeant vers 1 quand n tend vers l'infini.

25) Répondre à la question **17)** dans le cas où $\lambda > 1$.

Fin du sujet.