Devoir maison 2

Exercice 1 **Touver un chemin dans un graphe dynamique**

On considère un graphe évoluant avec le temps de la manière suivante :

— $G_0 \sim \mathcal{G}(n,p)$ est le graphe initial ;

— à l'instant $t+1$, G_{t+1} est obtenu à partir de G_t en sélectionnant une paire (u,v) de sommets uniformément au hasard et en plaçant une arête entre ces deux sommets avec probabilité p, en n'en plaçant pas avec probabilité $1-p$, indépendamment du reste du graphe et de son évolution passée. Donc G_t et G_{t+1} diffèrent donc d'au plus une arête.

Le but de ce problème des de maintenir un chemin depuis une source S jusqu'à une destination T, alors que les changements dans le graphe sont inconnus. La seule opération possible est de *sonder* un sommet à chaque étape. Sonder u signifie obtenir l'ensemble de ses voisins $N(u)$.

1. Justifier qu'à chaque étape t, $G_t \sim \mathcal{G}(n,p)$.

Correction : Pour $t=0$, $G_0 \sim \mathcal{G}(n,p)$ par définition. Supposons que $G_{t-1} \sim \mathcal{G}(n,p)$. Alors pour toute famille $A \subseteq \mathcal{P}_2(S)$, où S est l'ensemble des sommets, et en notant E_a l'événement « a est une arête de G_t »,

$$\mathbf{P}(\cap_{a\in A} E_a) = \sum_{(u,v)\in\mathcal{P}_2(S)} \mathbf{P}(\cap_{a\in A} E_a \mid (u,v) \text{ choisie à l'étape } t)\mathbf{P}((u,v) \text{ choisie à l'étape } t)$$

Si $(u,v) \in A$, cette arête est présente avec probabilité p indépendamment des autres, et les autres arêtes aussi, puisque $G_{t-1} \sim \mathcal{G}(n,p)$. Si $(u,v) \notin A$, alors chaque arête de A est présente indépendamment des autres avec probabilité p (ce sont des arêtes de $G_{t-1} \sim \mathcal{G}(n,p)$. Dans tous les cas, $\mathbf{P}(\cap_{a\in A} E_a \mid (u,v) \text{ choisie à l'étape } t) = p^{|A|}$ et donc $\mathbf{P}(\cap_{a\in A} E_a) = p^{|A|}$: les événements sont mutuellement indépendants et de probabilité p : $G_t \sim \mathcal{G}(n,p)$.

On suppose que p dépend de n et $\frac{\ln n}{n} \ll p \ll n^{-1/2}$ ($f \ll g$ signifie $f = o(g)$ et $f \gg g$ signifie $g = o(f)$).

2. Expliquer en une phrase pourquoi on choisit $p \gg \frac{\ln n}{n}$.

Correction : D'après le DM2, $\ln n/n$ est une fonction de seuil de la connectivité : si $p \gg \frac{\ln n}{n}$, alors G_t est connexe avec forte probabilité.

Fixons u et v deux sommets et soit $T_{u,v}$ l'événement « il existe un chemin entre u et v de longueur au plus 2 ».

3. Montrer que si $p \ll n^{-1/2}$, alors $\mathbf{P}(T_{u,v}) \overset{n\to\infty}{\longrightarrow} 0$.

Correction : L'événement $T_{u,v}$ peut s'exprimer à l'aide des événements E_a :

$$T_{u,v} = E_{u,v} \cup \bigcup_{w\in S\setminus\{u,v\}} E_{u,w} \cap E_{w,v}.$$

Les événements $E_{u,w}$ et $E_{w,v}$ sont indépendants : $\mathbf{P}(E_{u,w} \cap E_{w,v}) = p^2$. Donc par union bound, $\mathbf{P}(T_{u,v}) \leq p + (n-2)p^2$. Si $p \ll n^{-1/2}$, alors $p + (n-2)p^2 \overset{n\to\infty}{\longrightarrow} 0$, donc $\mathbf{P}(T_{u,v}) \overset{n\to\infty}{\longrightarrow} 0$.

4. Montrer que si $p \gg n^{-1/2}$, alors $\mathbf{P}(T_{u,v}) \overset{n\to\infty}{\longrightarrow} 1$.

Correction : En reprenant les notations de la question précédente, on a :

$$\mathbf{P}(T_{u,v}^c) = \mathbf{P}(E_{u,v}^c \cap \bigcap_{w\in S-\{u,v\}} (E_{u,w} \cap E_{w,v})^c = (1-p)(1-p^2)^{n-2} \leq (1-p)e^{-(n-2)p^2} \overset{n\to\infty}{\longrightarrow} 0,$$

car $p^2 n \overset{n\to\infty}{\longrightarrow} \infty$. Donc $\mathbf{P}(T_{u,v}) \overset{n\to\infty}{\longrightarrow} 1$,

Soit D_u le degré du sommet u.

5. Montrer qu'il existe une constante c telle que pour tout n suffisamment grand $\mathbf{P}(D_u \leq c \ln n) \leq n^{-4}$ et $\mathbf{P}(D_u \geq c\sqrt{n}) \leq n^{-4}$. En déduire une borne inférieure de la probabilité que chaque sommet a un degré au moins $c \ln n$ et au plus $c\sqrt{n}$. *Indication :* d'abord borner p par $a\frac{\ln n}{n-1}$ pour a suffisamment grand, et borner D_u, le degré de u par une variable aléatoire binomiale.

Correction : On pose $p(n) = a(n)\ln(n)/(n-1)$, donc $a(n) \to +\infty$. Fixons c une constante. À partir d'un certain rang, $a(n) > c$. Dans un graphe $\mathcal{G}(n,p)$, on a $D_u \sim \mathcal{B}in(n-1,p) \geq \mathcal{B}in(n-1, c\ln n)$ à partir d'un certain rang. Donc l'inégalité de Chernoff,

$$\mathbf{P}(D_u \leq c\ln n) \leq \mathbf{P}(D_u \leq a(n)\ln n(1-\delta)) \leq e^{-\delta^2 a(n)\ln n/2} = n^{-\delta^2 a(n)/2},$$

avec $\delta = 1 - c/a(n)$. On aura $\mathbf{P}(D_u \leq c\ln n) \leq n^{-4}$ dès que $a(n) > 8 + 2c$, ce qui arrive pour n assez grand.

On pose $p(n) = a(n)\sqrt{n}/(n-1)$. Cette fois, $a(n) \to 0$, Pour tout $c > 0$, dès que $c = 2a > 2a(n)$, $D_u \sim \mathcal{B}in(n-1,p) \leq \mathcal{B}in(n-1, a\sqrt{n}/(n-1))$. Pour tout $c > 0$, dès que $c = 2a > 2a(n)$, on peut écrire d'après l'inégalité de Chernoff (avec $\delta = 1$),

$$\mathbf{P}(D_u \geq c\sqrt{n}) \leq \mathbf{P}(D_u \geq c\sqrt{n}) \leq \mathbf{P}(\mathcal{B}in(n-1, a\sqrt{n}/n-1) \geq c\sqrt{n}) \leq e^{-a\sqrt{n}/3} \leq n^{-4}$$

pour n assez grand.

Par union bound, $\mathbf{P}(\forall u,\ c\ln n \leq D_u \leq c\sqrt{n}) \geq 1 - 2/n^3$.

On dit que la propriété A est satisfaite *avec forte probabilité* si $\lim_{n\to\infty} \mathbf{P}(A) = 1$ (la propriété A dépend de n). Par exemple, « $\max_u D_u \leq c\sqrt{n}$ » est satisfaite avec forte probabilité.

6. Soit B un événement. Montrer que si A est satisfaite avec forte probabilité et $\lim_{n\to\infty} \mathbf{P}(B \mid A) = q$, alors $\lim_{n\to\infty} \mathbf{P}(B) = q$.

Correction : $\mathbf{P}(B) = \mathbf{P}(B|A)\mathbf{P}(A) + \mathbf{P}(B|A^c)\mathbf{P}(A^c)$. Or $\mathbf{P}(A) \to 1$, donc $\mathbf{P}(A^c \to 0)$, et $\mathbf{P}(B|A^c) \leq 1$. Ainsi, $\mathbf{P}(B) = \mathbf{P}(B|A)\mathbf{P}(A) + \mathbf{P}(B|A^c)\mathbf{P}(A^c) \to \mathbf{P}(B|A) \times 1 + 0 = \mathbf{P}(B|A)$.

Soit $R = \lceil\sqrt{n/\ln n}\rceil$ et considérons l'algorithme 1, où $S \longrightarrow v$ (resp. $v \longrightarrow T$) est le plus court chemin de A vers v (resp. de v vers T) d'après les sondages effectués par l'algorithme). On veut montrer que la probabilité que l'algorithme retourne un chemin valide est au moins $1 - \alpha(n\ln n)^{-1/2}$ pour une constant α (on rappelle qu'à chaque étape (=sondage), le graphe évolue).

Algorithme 1 : Path finding

begin
- $B_S \leftarrow \{S\}$; $V_S \leftarrow \emptyset$;
- **repeat**
  - choisir une sommet u dans $B_s \setminus V_S$; $N(u) \leftarrow$ `sonder` u;
  - $B_S \leftarrow B_S \cup N(u)$; $V_S \leftarrow V_S \cup \{u\}$;
- **until** *pendant R étapes*;
- $B_T \leftarrow \{T\}$; $V_T \leftarrow \emptyset$;
- **repeat**
  - choisir un sommet u dans $B_T \setminus V_T$; $N(u) \leftarrow$ `sonder` u;
  - $B_T \leftarrow B_T \cup N(u)$; $V_T \leftarrow V_T \cup \{u\}$;
- **until** *pendant R étapes*;
- **if** $B_S \cap B_T \neq \emptyset$ **then**
  - soit $v \in B_S \cap B_T$ et retourner $S \longrightarrow v \longrightarrow T$
- **else**
  - retourner « pas de chemin S vers T »

L'algorithme construit deux boules autour de S et T en utilisant les sondages. Si ces deux boules s'intersectent, alors l'algorithme retourne un chemin possible de S vers T. Les ensembles V_S et V_T sont les ensembles de sommets *visités* : ils ont été trouvés pas le processus de sondages, puis sondés.

7. Montrer qu'il existe $c' > 0$ tel que à chaque étape et à la fin de l'algorithme, avec forte probabilité, $|B_S| \leq c'n/\sqrt{\ln n}$.

Correction : Soit A l'événement « à chaque date $t \leq R$, chaque sommet a un degré compris entre $c \ln n$ et $c\sqrt{n}$ ». Par union bound, $\mathbf{P}(A^c) \leq 2/n^3 \times R \to 0$. Donc $\mathbf{P}(A) \to 1$.

Soit B l'événement « $|B_S| \leq c'n/\sqrt{\ln n}$ à chaque étape et à la fin de l'algorithme ». On a $\mathbf{P}(B|A) = 1$. En effet, si tous les sommets ont degré au plus $c\sqrt{n}$, alors à chaque étape, on ajoute au plus $c\sqrt{n}$ sommets à B_S, et donc à la fin (et par conséquent à chaque étape), on a au plus $Rc\sqrt{n} \sim cn/\sqrt{\ln n}$ sommets dans B_S.

8. Montrer qu'il existe une constante c'' telle qu'à chaque étape, le nombre de nouveaux sommets découverts ($N(u) \setminus B_S$) est au moins $c'' \ln n$ avec forte probabilité. En déduire qu'à la fin de l'algorithme, $|B_S| \geq c'' R \ln n$ avec forte probabilité.

Correction : Considérons le sommet u. Si son degré est au moins $c \ln n$ et $|B_S| \leq c'n/\sqrt{\ln n}$, alors on peut borner inférieurement son degré hors de B_S par une loi binomiale $\mathcal{B}in(c \ln n, 1 - c'/\sqrt{\ln n})$: parmi les $c \ln n$ voisins de u, chaque a une probabilité au plus $c'/\sqrt{\ln n}$ d'être dans $B(S)$. En appliquant les bornes de Chernoff : $\mathbf{P}(\mathcal{B}in(c \ln n, 1 - c'/\sqrt{\ln n}) \leq c'' \ln n) \leq e^{-\mu\delta^2/2}$, avec $\mu = c \ln n(1 - c'/\sqrt{\ln n})$ et $\delta = 1 - \frac{c''}{c(1-c'/\sqrt{\ln n})}$. On peut choisir c'' tel que $c'' \leq c(1 - \frac{c'}{\sqrt{\ln n}}) - 4$, pour avoir $\mathbf{P}(\mathcal{B}in(c \ln n, 1 - c'/\sqrt{\ln n}) \leq c'' \ln n) \leq n^{-2}$.

Soit C l'événement « à chaque étape $t \leq R$, le nombre de sommets découverts est au moins $c'' \ln n$ ». On a $P(C^c \mid B) \leq Rn^{-2} \to 0$, et B est satisfait avec forte probabilité, donc C aussi. Si C est satisfait, alors à chaque étape on découvre au moins $c'' \ln n$ sommets, et donc à la fin $|B_S| \geq c'' R \ln n$. Si C est satisfait avec forte probabilité alors $\mathbf{P}(|B_S| \geq c'' R \ln n) \to 1$.

On analyse maintenant la probabilité que l'algorithme échoue. Il y a deux types d'échec :

— soit l'algorithme renvoie « pas de chemin de S vers T », alors qu'il y a un chemin de S à T ;

— soit l'algorithme renvoie un chemin qui n'est plus valide : une arête sur le chemin renvoyé a disparu pendant le processus de sondage.

9. Montrer que la probablité qu'un sommet visité v de V_T n'ait pas de voisin dans B_S est au plus $(1 - \frac{c'' R \ln n}{n})^{c \ln n}$ avec forte probabilité. En déduire que la probabilité qu'il n'y ait pas d'arête entre V_T et B_S est au plus $e^{-cc'' \ln n}$ pour n assez grand. Conclure.

Correction : Soit v un sommet de V_T. On borne la probabilité qu'il n'ait pas de voisin dans B_s comme suit : il a au moins $c \ln n$ voisins. Chaque voisin est choisi indépendamment dans les autres sommets. Un voisin est dans B_s avec probabilité $p' \geq c'' R \ln n/n$. Donc il n'y a pas de voisin dans B_s avec probabilité au plus $(1 - p')^{D_u} \leq (1 - c'' R \ln n/n)^{c \ln n} \leq n^{-cc'' R \ln n/n} = p''$ (les événements en jeu ne sont pas indépendants. En toute rigueur, il faudrait utliser les probabilité conditionnelles pour obtenir le résultat).

Il n'y a pas d'arête entre V_T et B_S avec probabilité au plus $p''^R \leq n^{-cc''}$.

On admet que le diamètre de G_t est en $O(\ln n)$ avec forte probabilité, ainsi que les diamètres de B_S et B_T. (Le diamètre d'un graphe est la longueur maximum des plus courts chemins entre chaque paire de sommets)

10. Quelle est la probablité que « l'arête e est retirée à l'étape t » ? Conclure quant à la probabilité que le chemin renvoyé est invalide.

Correction : Le chemin renvoyé reste invalide si une arête sur le plus court chemin renvoyé par l'algorithme est retirée du graphe à une étape précedente. Le nombre d'arête du chemin est au plus $b \ln n$ avec forte probabilité, avec b une constante. La probabilité qu'une de ces arêtes soit retirée est au plus $R(b \ln n)/n^2(1 - p) \leq b\sqrt{\ln n} n^{-3/2} \to 0$.

11. Conclure quant à la probabilité d'échec de l'algorithme.

Correction : La probabilité d'échec est soit un chemin invalide (probabilité qui tend vers 0 par la question 10), soit un chemin n'a pas été trouvé, alors qu'il y en avait un, et cette probabilité tend aussi vers 0 d'après la question 9.

Source : Aris Anagnostopoulos, Ravi Kumar, Mohammad Mahdian, Eli Upfal, and Fabio Vandin. 2012. Algorithms on evolving graphs. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (ITCS '12). p.149-160.