Partiel d'Algorithmique

Durée 2 heures

(Correction)

Important : Ce document ne fait que proposer des corrections aux exercices du partiel.

Exercice 1 : Algorithme randomizé et structures aléatoires

Dans cet exercice, on considère l'algorithme suivant qui va générer un graphe aléatoire avec n sommets (étiquettes) et M arêtes.

Algorithme 1 : RandomGraphGenerator

Data : n, M
Result : Un graphe aléatoire $\mathbb{G}(n, M)$
$V := \{1, \cdots, n\}$;
$E := \emptyset$;
(⋆ $G = (V, E) =$ graphe vide ⋆)
$r := 0$;
for $i := 1$ *to* M (⋆ *i est le compteur d'arêtes* ⋆)
do
  repeat
    Prendre x uniformément au hasard (u.a.r.) dans l'ensemble V.
    Prendre y u.a.r. dans l'ensemble $V - \{x\}$.

    $r := r + 1$; (⋆ r est un *compteur* de **répétitions** sur les (x, y) déjà liés ⋆)
  until $(x, y) \notin E$
  (⋆ *Répéter jusqu'à trouver* **deux sommets distincts non encore reliés** *par une arête* ⋆)

  $E := E \bigcup \{(x, y)\}$; (⋆ on rajoute l'arête (x, y) dans G ⋆)
  $i := i + 1$;
end
return (V, E) ; (⋆ on renvoie le graphe $G = (V, E)$ ⋆)

1. Dans cette question, on supposera que $n^{7/8}$ est un entier. Montrer que si $M = n^{7/8}$ alors le compteur r vérifie $r = M$ avec une probabilité tendant vers 1 avec n tendant vers l'infini ("vraisemblablement, il n'y aura pas de retirages/répétitions").

▷
- *Le compteur r compte le nombre de fois ou on a tiré une arête. On tire exactement M arêtes sans répéter un seul tirage ssi $r = M$.*
- *On peut numéroter les arêtes de 1 à $\frac{n(n-1)}{2}$ par $a_1, a_2, \cdots, a_{n(n-1)/2}$.*
- *L'algorithme choisit uniformément au hasard une arête (x, y) sur les sommets x et y donc chaque arête a une probabilité de*
$$\frac{2}{n(n-1)}$$
d'être choisie (avant le test "$(x, y) \in E$" dans le "repeat" ... "until").
- *Dans la boucle principale ("for i"), supposons que $i = j$, ça veut dire qu'on a déjà tiré j arêtes. Pour qu'il y ait répétition, dans le "repeat" ... "until", on a retiré une arête déjà présente. La probabilité que ça arrive est*
$$\frac{j}{n(n-1)/2 - j}.$$
On note X_j la variable aléatoire
$$X_j = \begin{cases} 1, & \text{on a du retirer à l'étape } i = j \text{ de la boucle principale} \\ 0, & \text{sinon} \end{cases}$$

On a

$$\mathbb{P}[X_j = 1] = \frac{j}{n(n-1)/2 - j}\,.$$

En espérance, le nombre de j tels que $X_j = 1$ vérifie

$$\mathbb{E}[\sum_{j=1}^{M} X_j] = \sum_{j=1}^{M} \mathbb{E}[X_j] = \sum_{j=1}^{M} \frac{j}{n(n-1)/2 - j} < \sum_{j=1}^{M} \frac{M}{n(n-1)/2 - M} < \frac{M^2}{n(n-1)/2 - M}$$

On instancie le calcul avec $M = n^{7/8}$

$$\mathbb{E}[\sum_{j=1}^{M} X_j] < O\left(\frac{n^{7/4}}{n^2}\right) = O(n^{-1/4})\,.$$

C'est une espérance qui tends vers 0 quand n est grand donc l'inégalité de Markov nous dit que aucun des compteurs X_j ne s'incrémentera avec très forte probabilité. Autrement dit dans le "repeat" ... "until", on ne retirera jamais, i.e. $r = M$.

`Autre observation.`

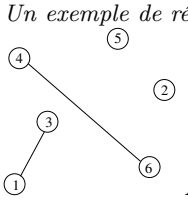
C'est comme lancer uniformément au hasard des boules dans des paniers. On peut voir que les arêtes sont les paniers. On a $n(n-1)/2$ paniers et M boules. Pour instancier, imaginons $n = 10^8$. On tire/lance $n^{7/8} = 10^7$ boules/balles dans presque 0.5×10^{16} urnes/paniers (toutes les arêtes possibles). Du coup, on n'aura presque jamais une urne avec plus qu'une boule.

2. Dans cette question, on supposera que $n^{1/4}$ est un entier. Montrer que si $M = n^{1/4}$ alors l'algorithme RandomGraphGenerator(n, M) ne produit qu'un graphe aléatoire avec des *sommets isolés* et des *arêtes isolées* avec une probabilité tendant vers 1 quand $n \to +\infty$.

▷

```
----------------------------------------
1ere solution: compter les réalisations
----------------------------------------
```

Un exemple de réalisation avec $n = 6$ et $M = 2$ est donné par



Au total, nous avons

$$\binom{n(n-1)/2}{M} \qquad \textit{graphes possibles.}$$

En effet, on doit choisir M arêtes parmi $n(n-1)/2$.
La question revient donc à quantifier la probabilité d'une réalisation avec
- *$n - 2M$ sommets totalement isolés (2 et 5 dans le dessin ci-dessus)*
- *M arêtes indépendantes (arête (1,3) et (4,6) dans la figure ci-dessus)*

On commence par choisir les sommets isolés

$$\binom{n}{n-2M} = \frac{n!}{(n-2M)!(2M)!} = \binom{n}{2M} \textit{ choix.}$$

Sur le reste, on construit M pairs de sommets

$$\frac{(2M)!}{M!} \times \frac{1}{2^M}\,.$$

On vérifie
- *pour $M = 1$, on a deux sommets s_1, s_2 :*
  l'unique pair $\{s_1, s_2\}$ et ça fait bien $2!/1!/2^1 = 1$.
- *pour $M = 2$, on a quatre sommets s_1, s_2, s_3, s_4 à apparier :*
  $\{s_1, s_2\}$ et $\{s_3, s_4\}$ OU
  $\{s_1, s_3\}$ et $\{s_2, s_4\}$ OU
  $\{s_1, s_4\}$ et $\{s_2, s_3\}$.
  La formule donne $4!/2!/2^2$, ca fait bien 3.

– *pour $M = 3$, on a 6 sommets $s_1, s_2, s_3, s_4, s_5, s_6$ à apparier :*
il y en a 15 (j'oublie le "s_" pour faire court !).

```
12,34,56 | 12, 35,46 | 12, 36, 45 |
13, 24, 56 | 13, 25,46 | 13, 26, 45 |
14, 23, 56 | 14, 25,36 | 14, 26,35 |
15, 23, 46 | 15, 24, 36 | 15, 26, 34 |
16, 23, 45 | 16, 24, 35 | 16, 25, 34
```

On vérifie pour $M = 3$, on a bien $(2M)!/M!/2^M = 6!/3!/2^3 = 720/6/8 = 15$.

Explications :
PREMIERE ARETE. On choisit un sommet parmi $2M$. On l'apparie avec un des $2M - 1$ restant
SECONDE ARETE. On choisit un sommet parmi $2M - 2$. On l'apparie avec un des $2M - 3$ restant
TROISIEME ARETE. On choisit un sommet parmi $2M - 4$. On l'apparie avec un des $2M - 5$ restant
...
M ième ARETE. On choisit un sommet parmi $2M - 2(M - 1)$. On l'apparie avec un des $2M - 2(M - 1) - 1$ restant
Ca nous fait

$$(2M).(2M-1) \times (2M-2)(2M-3) \times (2M-2(M-1)).(2M-2(M-1)-1) = (2M)!$$

Comme il n'y a pas d'ordre sur les arêtes dans le dessin du graphe (pas de première arête ni de seconde arête, ...) on a un facteur de permutations à compenser d'où

$$\frac{1}{M!}\,.$$

Comme dans une arête, il n'y a pas d'ordre on a un facteur

$$\frac{1}{2^M}\,.$$

Au total, la probabilité qu'on recherche est

$$\textit{proba via ratio} = \binom{n}{2M} \times (2M)!/M!/2^M \times \frac{1}{\binom{n(n-1)/2}{M}}$$

On calcule

$$\binom{n(n-1)/2}{M} \sim \frac{n^{2M}}{2^M M!}\,.$$

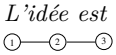
Quand $M = n^{1/4}$ (le calcul ci-dessous marche toujours vrai quand M est négligeable devant $\sqrt{n}$ et c'est le cas pour $n^{1/4}$), un calcul rapide donne que c'est approximativement

$$\textit{proba via ratio} \sim \frac{n^{2M}}{2M!} \times (2M)!/M!/2^M \times \frac{2^M M!}{n^{2M}} = 1\,.$$

CQFD

```
---------------------------------------------------------------------
2nde solution: faire du Markov pour montrer qu'on n'aura jamais deux
arêtes collées
---------------------------------------------------------------------
```

L'idée est de démontrer qu'on n'aura jamais un motif comme dessiné ci-dessous



c'est-à-dire aucun chemin de longueur 2 avec forte probabilité dans tout le dessin du graphe. Pour deux sommets x et y quelconques, la probabilité que l'arête (x, y) soit dessinée dans le graphe est

$$\frac{M}{\binom{n(n-1)/2}{M}}\,.$$

On fixe 3 sommets qu'on nommera x, y et z. Ces 3 sommets peuvent former un chemin de longueur 2 de trois manières via $x - -y - -z$, $y - -x - -z$ ou $x - -z - -y$. Pour ces 3 sommets, on définit la variable aléatoire

$$X_{x,y,z} = \begin{cases} 1, & \textit{s'il y a un chemin de longueur 2 entre les sommets } x, y \textit{ et } z \\ 0, & \textit{sinon} \end{cases}$$

On calcule

$$\mathbb{P}[X_{x,y,z} = 1] = \mathbb{P}[\textit{il y a (au moins) un chemin de longueur 2 entre } x,y \textit{ et } z] = 3 \times \left(\frac{M}{\binom{n(n-1)/2}{M}}\right)^2\,.$$

Pour $M = n^{1/4}$, nous avons

$$\binom{n(n-1)/2}{M} \sim \frac{n^{2M}}{2^M M!}\,.$$

et

$$\mathbb{P}[X_{x,y,z} = 1] \sim 3\frac{n^{1/2}2^{2M}(M!)^2}{n^{4M}}$$

Soit X le compteur du nombre de chemin(s) de longueur 2.

$$X = \sum_{x,y,z} X_{x,y,z}$$

et (linéarité de l'espérance)

$$\mathbb{E}[X] = \sum_{x,y,z} \mathbb{E}[X_{x,y,z}] = \binom{n}{3}\mathbb{E}[X_{1,2,3}] \sim \binom{n}{3} \times 3\frac{n^{1/2}2^{2M}(M!)^2}{n^{4M}}$$

On a $M! < M^M$ et $\binom{n}{3} < n^3$. Donc pour $M = n^{1/4}$,

$$\mathbb{E}[X] < 3n^{3+1/2}2^{2n^{1/4}}n^{2n^{1/4}}\frac{1}{n^{4n^{1/4}}} = \textit{Polynôme}(n) \times \left(\frac{2^2n^2}{n^4}\right)^{n^{1/4}} \longrightarrow 0\,.$$

On conclut avec l'inégalité de Markov : avec une forte probabilité, le compteur X ne s'incrémentera jamais, donc on n'aura pas de chemin de longueur 2.

3. Pour quelles valeurs du paramètre M, r atteint $r = n(n-1)/2$ avec probabilité presque 1 ?

▷ *Si $r = n(n-1)/2$, cela voudrait dire qu'on aura déjà collecté toutes les arêtes. Et qu'on ne pourra donc plus rajouter une seule arête au graphe (car elle existe déjà). Dans un graphe le nombre d'arêtes vaut au plus $n(n-1)/2$ et pour collecter toutes les arêtes (voir cours sur collecteur de coupons ou pokemon), il faut $M = O(n(n-1)/2\log(n(n-1)/2))$. Donc, $r = n(n-1)/2$ n'arrivera pas car $M \leq n(n-1)/2$.*

Exercice 2 : Algorithmique randomizée

Dans cet exercice on s'intéresse aux problèmes d'élection (vus en cours) dans les modèles synchrones quand le graphe support du réseau est complet et anonyme (les processeurs sont non identifiés).

1. Dans cette question, on considère le modèle de "radio networks" avec exactement 2 processeurs communiquant via un canal radio (avec ou sans détecteur de collision). A un instant discret t, chacun des deux processeurs tente d'accaparer le canal avec probabilité p. Calculer la probabilité qu'un des deux processeurs réussisse à obtenir seul le canal.

▷

```
**************** Correction ******************
```

$$\textit{Binomial}(2,p) = 2.p.(1-p)$$

2. On se fixe $\varepsilon > 0$. On dira dans cette partie qu'un algorithme d'élection réussit avec très forte probabilité s'il succède à élire un leader avec une probabilité d'au moins $1 - O(\varepsilon)$. Dans les mêmes hypothèses que la question précédente, montrer qu'on ne peut pas élire un leader avec très forte probabilité en moins de $\Omega(\log\frac{1}{\varepsilon})$ * .

▷

```
**************** Correction ******************
```

$$\mathbb{P}[\textit{Succès au temps } t] \leq 2.p.(1-p) = c$$

Le temps d'attente pour un succès est donc une loi géométrique. On ne réussit après k essais avec probabilité $(1-c)^k$. Ce taux d'échec est moins que $(1-c)^k \leq \varepsilon$ pour $k \geq O(\log\varepsilon)$.

*. On rappele que $g(n) = \Omega(f(n))$ se lit "g grandit au moins aussi vite que f" ou encore "l'ordre de grandeur de g est au moins celui de f".

3. Généraliser les deux questions précédentes dans le cas d'un réseau complet à k nœuds, i.e. (i) calculer la probabilité qu'une élection ait lieu à un temps quelconque dans un réseau à k processeurs et (ii) montrer qu'il faut un temps logarithmique pour avoir un tel succès avec très forte probabilité.

▷

`********** Correction **********`

$$(i)\mathbb{P}[\textit{Succès au temps } t] \leq \binom{k}{1} p^1 (1-p)^{k-1}$$

$$(ii)\textit{idem, il faut un temps d'au moins } O(\log \varepsilon)$$

4. Déduire des questions précédentes un algorithme Monte Carlo pour élire un leader parmi n processeurs dans le cas d'un réseau avec detecteur de collision. Votre algorithme doit s'arrêter en un temps connu d'avance avec un taux d'erreur de moins de $1/n^3$.

▷

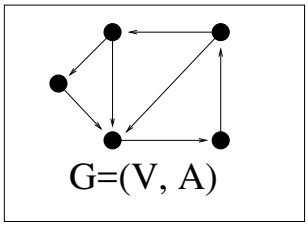
`********** Correction **********`

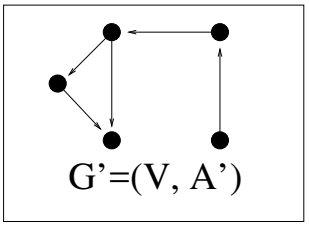
Pour i de 1 *à* $\lceil 3 \log n \rceil$ *faire*
Avec probabilité $1/n$ *tenter de prendre la main.*
Si un seul proc. prend la main alors il est élu. ***BREAK/FIN****. FinSi*
FinPour

Cet algorithme se finit en un temps moyen $O(1)$ *avec un élu mais si on veut un taux d'erreur de* $1/n^3$ *alors il faut d'après les questions précédentes* $-O(\log 1/n^3)$ *(*$\varepsilon = 1/n^3$*).*

Exercice 3 : Problèmes d'approximation dans les graphes dirigés

Dans cet exercice on s'intéresse aux graphes dirigés (ou digraphes) et en particulier aux sous-graphes acycliques. Le problème algorithmique est le suivant.
Input : Un graphe dirigé $G = (V, A)$.
Output : $A' \subseteq A$ tel que le digraphe $G' = (V, A')$ ne contienne pas de cycle orienté.
Optimisation : *maximiser* $|A'|$, i.e. le nombre d'arcs dans G'.





1. En observant qu'un sommet s appartenant à un cycle orienté possède toujours un arc rentrant et un arc sortant proposer un algorithme glouton offrant une solution au problème ci-dessus (sans l'optimisation).

▷

1. *Etiqueter arbitrairement les sommets de* 1 *à* n.
2. *Enlever pour chaque sommet dans l'ordre soit l'ensemble des arcs rentrant, soit l'ensemble des arcs sortant. Pour maximiser, on enlève "au max".*

2. Quel est le facteur d'approximation de votre algorithme ?

▷ 2. *L'algorithme tourne en temps* $O(n^2)$ *car "pour chaque sommet, tester les arcs". On a* $|OPT| \leq |A|$ *et la solution fournie est d'au moins* $|A|/2$.

Exercice 4 : Formules CNF (forme normale conjonctive ou Conjunctive Normal Form)

Une formule CNF $F_{n,m}$ est donnée par $C_1 \wedge C_2 \wedge \cdots \wedge C_m$ avec chaque clause C_i est de la forme $y_1 \vee y_2 \vee \cdots y_k$ avec les $y_i \in \{x_1, x_2, \cdots, x_n, \bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n\}$ et k est un entier positif quelconque. On dit qu'une instance $F_{n,m}$ est SAT s'il existe une affectation des variables booléennes x_i qui va satisfaire toutes les clauses C_i de $F_{n,m}$.

1. On note par $|C_i|$ la longueur de la clause C_i. Montrer que si $\sum_i^m 2^{-|C_i|} < 1$ alors $F_{n,m}$ est SAT.

▷ *On applique la règle de la méthode probabiliste.* $1/2^{|C_i|}$ *représente la probabilité qu'une clause de longueur* C_i *ne soit pas SAT suivant une affectation aléatoire uniforme des variables. La somme* $\sum_{i=1}^m 2^{-|C_i|}$ *est une espérance. En effet, on définit*

$$X_i = \begin{cases} 0 & \textit{si la clause } i \textit{ de longueur } |C_i| \textit{ est SAT} \\ 1 & \textit{sinon} \end{cases}$$

et $X = \sum_{i=1}^m X_i$ *(*X *est le "compteur" des clauses non-SAT). On a*

$$\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^m X_i] = \sum_{i=1}^m \mathbb{E}[X_i] = \sum_{i=1}^m 2^{-|C_i|}.$$

La moyenne d'un compteur (variable positive entière) est strictement plus petite que 1*, ce qui veut dire qu'il y a des réalisations avec* O *clauses non-SAT.*