Math ECG 2. 2022-2023

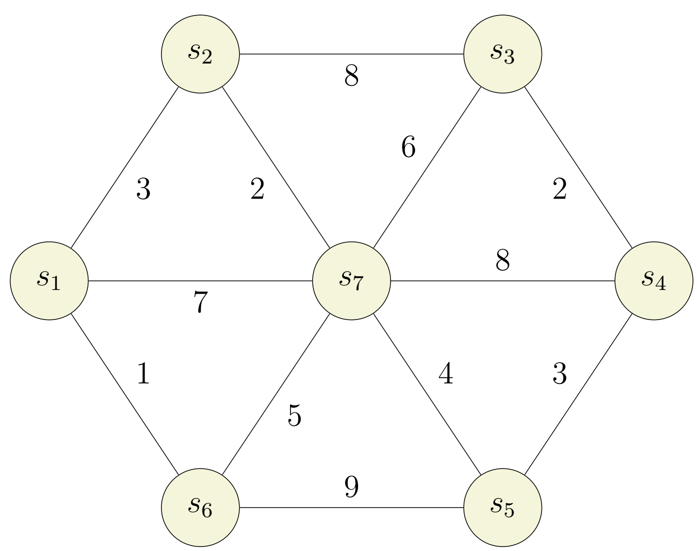
Mathématiques Appliquées - F. Gaunard
http://frederic.gaunard.com
ENC Bessières, Paris 17e.

Devoir Maison n°5

Solution

Exercice 1

(1) (a) Le coefficient $a_{i,j}$ de la matrice des poids donne le poids de l'arrête entre les sommets i et j. On remarque d'ailleurs que, le graphe n'étant pas orienté, cette matrice est symétrique. La présence d'un $+\infty$ signifie que les sommets correspondants ne sont pas reliés par une arrête. On reporte alors facilement les coefficients sur le graphe :



(b) Un graphe est complet si tous les sommets sont deux à deux reliés. Ce n'est pas le cas ici. Par exemple, le sommet s_5 n'est pas relié au sommet s_1.
Un graphe est connexe si, pour toute paire de sommet, il existe une chaîne reliant ces sommets. C'est le cas ici.

(2) On dresse ci-dessous le tableau correspondant au calcul des plus courts chemins de s_5à chacun des sommets du graphe par l'algorithme de Dijkstra. A chaque étape, on fait le test de *relâchement*; si la distance cumulée en passant pas le nouveau sommet est inférieure à celle déjà calculée, on remplace et on passe alors par ce sommet. On trouve que pour relier s_5 à s_1, le plus court chemin est $s_5 - s_7 - s_2 - s_1$ de longueur 9.

s_1	s_2	s_3	s_4	s_5	s_6	s_7	étape i	choix sommet étape $i+1$
$+\infty$	$+\infty$	$+\infty$	$+\infty$	0	$+\infty$	$+\infty$	0	s_5
$+\infty$	$+\infty$	$+\infty$	3^{s_5}	0	9^{s_5}	4^{s_5}	1	s_4
$+\infty$	$+\infty$	5^{s_4}	3^{s_5}	0	9^{s_5}	4^{s_5}	2	s_7
11^{s_7}	6^{s_7}	5^{s_4}	3^{s_5}	0	9^{s_5}	4^{s_7}	3	s_3
11^{s_7}	6^{s_7}	5^{s_4}	3^{s_5}	0	9^{s_5}	4^{s_7}	4	s_2
9^{s_2}	6^{s_7}	5^{s_4}	3^{s_5}	0	9^{s_5}	4^{s_7}	5	s_1
9^{s_2}	6^{s_7}	5^{s_4}	3^{s_5}	0	9^{s_5}	4^{s_7}	6	s_6
9^{s_2}	6^{s_7}	5^{s_4}	3^{s_5}	0	9^{s_5}	4^{s_7}	7	$-$

(3) Au vu de la matrice implémentée, il s'agit du graphe ci-dessus. La liste renvoyée est celle correspondant aux distances de la dernière ligne du tableau de l'algorithme de Dijkstra en partant de s_1 (qui est donc le sommet listé 0). Ainsi, le sommet listé 4 est en fait le $5-$ième sommet, c'est à dire s_5 et la commande

```
print(Dijkstra_dist(G, 4))
```

va donc renvoyer la dernière ligne du tableau qu'on vient de former ci-avant, à savoir

Affichage Python

```
> > >
[9, 6, 5, 3, 0, 9, 4]
```

Exercice 2* - Graphes aléatoires

Dans tout l'exercice n désigne un entier supérieur ou égal à 2 et p un réel de l'intervalle $]0;1[$.
Soit $S = \{s_1, ..., s_n\}$ un ensemble fini de n sommets. On s'intéresse dans cet exercice à des graphes aléatoires construits à partir de l'ensemble S.
Plus précisément, pour tout couple $(i,j) \in [\![1,n]\!]^2$ avec $i < j$, on introduit les variables aléatoires $T_{i,j}$ mutuellement indépendantes de même loi $\mathcal{B}(p)$. Les arêtes du graphe sont les paires de sommets $\{s_i, s_j\}$ pour lesquelles $T_{i,j} = 1$.
On dit qu'un sommet est isolé s'il n'y a aucune arête incidente à ce sommet.
On introduit N_n la variable aléatoire égale au nombre d'arêtes du graphe. Pour $k \in [\![1,n]\!]$, on note D_k la variable aléatoire qui prend pour valeur le degré du sommet s_k (c'est à dire le nombre d'arêtes incidentes à ce somment).
Enfin, on note X_k la variable aléatoire qui vaut 1 si le sommet s_k est isolé et 0 sinon puis Z_n celle égale au nombre de sommets isolés du graphe.

(1) Il y a deux boucles imbriquées. Ainsi, pour i variant entre 0 et $n-2$ et j entre $i+1$ et $n-1$, si `rd.rand()<p` (ce qui représente la bernoulli de paramètre p, à savoir $T_{i,j}$, on relie les sommets i et j. Donc la liste d'adjacence du sommet i prend en plus le sommet j et inversement.

```
import numpy as np
import numpy.random as rd

def list_adj(S, p):
    n=len(S)
    l=[ [ ] for k in range(n)]
    for i in range(n-1):
        for j in range(i+1, n):
            if rd.random()<p :
                l[i].append(S[j])
                l[j].append(S[i])
    return l
```

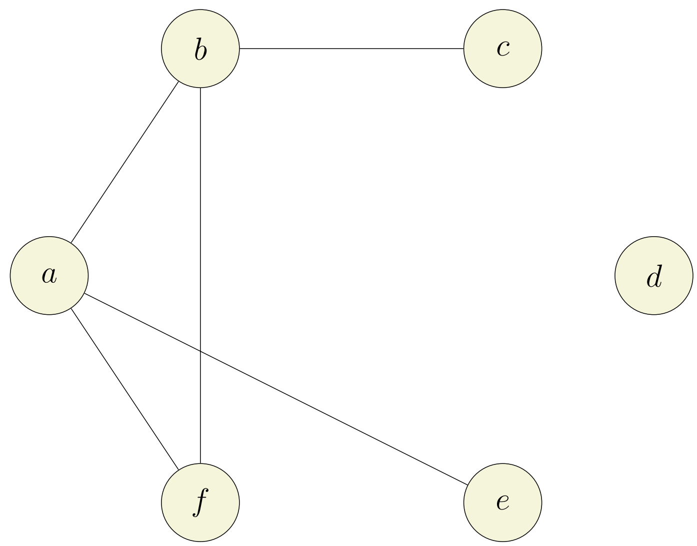
(2) On exécute alors la commande suivante qui génère l'affichage suivant

```
list_adj('abcdef', 1/3)
```

Affichage Python

```
>>>
[['b', 'e', 'f'], ['a', 'c', 'f'], ['b'], [], ['a'], ['a', 'b']]
```

(a) La liste d'adjacence ci-dessus peut se représenter comme suit:



(b) Ce graphe ne contient qu'un seul sommet isolé, le sommet d.

(c) La matrice d'adjacence du graphe est la matrice $A = (a_{i,j})$ où $a_{i,j} = 1$ ou 0 selon si le sommet i et relié au sommet j ou non. Le graphe n'étant pas orienté, la matrice est bien sûr symétrique et il s'agit de la matrice

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(3) Il suffit, pour chaque sommet du graphe, de voir si la liste des sommets adjacents est vide (c'est à dire de longueur 0) et le compter.

```
def nb_som_is(l):
    n=len(l)
    c=0
    for k in range(0, n):
        if len(l[k])==0 :
            c=c+1
    return c
```

(4) (a) Cette fonction renvoie la valeur moyenne du nombre de sommets isolés dur 1000 graphes aléatoires simulés avec le protocole précédent à partir d'une liste de sommets S (de cardinal n), il s'agit donc d'une *estimation* de l'espérance de la variable aléatoire Z_n.

(b) Au vu de la figure et ce que représente chaque courbe, on peut conjecturer que

$$E(Z_n) = n(1-p)^{n-1}.$$

(5) (a) Il est possible que tous les sommets soient isolés (si toutes les bernoulli renvoient 0), auquel cal N_n prend la valeur minimale 0. Au maximum, toute paire de sommets distincts est reliés par une arrête. Il y a $\binom{n}{2}$ paires de sommets donc un nombre maximum d'arêtes de $\binom{n}{2}$. On a bien $N_n(\Omega) \subset [\![0, \binom{n}{2}]\!]$.

(b) Il est clair que $[N_n = 0]$ si et seulement si aucune paire de sommet n'est reliée par une arête, c'est à dire si et seulement si, pour tout $i \in [\![1, n-1]\!]$, pour tout $j \in i+1, n]\!]$, $[T_{i,j} = 0]$. Ainsi, on a bien

$$P(N_n = 0) = P\left(\bigcap_{i=1}^{n-1} \bigcap_{j=i+1}^{n} [T_{i,j} = 0]\right).$$

Par indépendance mutuelle des variables $T_{i,j}$, on a donc

$$P(N_n = 0) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} (1-p) = (1-p)^{n(n-1)/2}.$$

(c) De la même manière, il faut relier tous les sommets donc il faut que chaque v.a $T_{i,j}$ prenne la valeur 1:

$$P(N_n = 0) = P\left(\bigcap_{i=1}^{n-1} \bigcap_{j=i+1}^{n} [T_{i,j} = 1]\right) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} p = p^{n(n-1)/2}.$$

(6) (a) Soit $k \in [\![1, n]\!]$. D_k compte le nombre d'arêtes qui partent (ou arrivent, le graphe n'est pas orienté) du sommet k. Il s'agit de compter combien de sommets lui sont reliés. Il est clair que D_k correspond donc à la somme des $T_{i,k}$ pour $i < k$ (c'est à dire que pour les sommets numérotés de 1à $k-1$ on compte ceux qui sont reliés à k) et des $T_{k,i}$ pour $i > k$. On peut donc écrire

$$D_k = \sum_{i=1}^{k-1} T_{i,k} + \sum_{i=k+1}^{n} T_{k,i}.$$

Il apparait que D_k est la somme de $n-1$ variables aléatoires indépendante de même loi de Bernoulli de paramètre p. D'après le cours, D_k suit donc une loi binomiale

$$D_k \hookrightarrow \mathcal{B}(n-1, p).$$

(b) Commençons par observer que, par (mutuelle) indépendance des variables aléatoires $T_{i,j}$,

$$\operatorname{cov}(T_{i,m}, T_{j,r}) = 0$$

dès que $(i, m) \neq (j, r)$.
Soient $1 \leq k < \ell \leq n$, par bilinéarité de la covariance, en observant que $\ell \in [\![k+1, n]\!]$ et $k \in [\![1, \ell-1]\!]$, on a

$$\begin{aligned}
\operatorname{cov}(D_k, D_\ell) &= \operatorname{cov}\left(\sum_{i=1}^{k-1} T_{i,k} + \sum_{i=k+1}^{n} T_{k,i}, \sum_{j=1}^{\ell-1} T_{j,\ell} + \sum_{j=\ell+1}^{n} T_{\ell,j}\right) \\
&= \cancel{\sum_{i=1}^{k-1} \sum_{j=1}^{\ell-1} \operatorname{cov}(T_{i,k}, T_{j,\ell})} + \cancel{\sum_{i=1}^{k-1} \sum_{j=\ell+1}^{n} \operatorname{cov}(T_{i,k}, T_{\ell,j})} + \\
&\qquad + \sum_{i=k+1}^{n} \sum_{j=1}^{\ell-1} \operatorname{cov}(T_{k,i}, T_{j,\ell}) + \cancel{\sum_{i=k+1}^{n} \sum_{j=\ell+1}^{n} \operatorname{cov}(T_{k,i}, T_{\ell,j})} \\
&= \operatorname{cov}(T_{\ell,k}, T_{\ell,k}) = V(T_{\ell,k}) \\
&= p(1-p)
\end{aligned}$$

Leur covariance n'est pas nulle: les variables D_k et D_ℓ ne sont donc pas indépendantes.

(c) On peut écrire $Z_n = \sum_{k=1}^n X_k$, où X_k est une Bernoulli de paramètre (et donc d'espérance) $P(D_k = 0) = (1-p)^{n-1}$. Par linéarité de l'espérance

$$E(Z_n) = \sum_{k=1}^{n} E(X_k) = n(1-p)^{n-1},$$

ce qui est bien la formule conjecturée ci-avant.

(7) (a) Soient $i < j$. Les sommets i et j sont tous deux isolés si et seulement si toutes les variables $T_{i,k}, T_{k,i}, T_{j,k}$ et $T_{k,j}$ valent 0. Il y a $n-1$ sommets qu'on ne veut pas relier à i et donc $n-2$ autres qu'on ne veut pas relier à j, ainsi il y a $n-1+n-2 = 2n-3$ évènements indépendants de probabilités $1-p$ qui correspondent à l'évènement cherché. Il suit que

$$P([X_i = 1] \cap [X_j = 1]) = (1-p)^{2n-3}.$$

(b) On a écrit plus haut que $Z_n = \sum_{k=1}^{n} X_k$. Par identité remarquable, on a alors

$$Z_n^2 = \sum_{i=1}^{n} X_i^2 + 2 \sum_{1 \le i < j \le n} X_i X_j.$$

Or, X_i est une variable de Bernoulli, donc $X_i^2 = X_i$ et on a bien

$$Z_n^2 = \sum_{i=1}^{n} X_i + 2 \sum_{1 \le i < j \le n} X_i X_j.$$

Par linéarité de l'espérance, il suit que

$$E(Z_n^2) = \sum_{i=1}^{n} E(X_i) + 2 \sum_{1 \le i < j \le n} E(X_i X_j).$$

Or, $E(X_i) = (1-p)^{n-1}$ et $X_i X_j$ est encore une loi de Bernoulli de paramètre (et donc d'espérance) $P([X_i = 1] \cap [X_j = 1]) = (1-p)^{2n-3}$ par ce qui précède. Il suit que

$$E(Z_n^2) = n(1-p)^{n-1} + 2(1-p)^{2n-3} \sum_{1 \le i < j \le n} 1 = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}.$$