

Matière : **ANALYSE DE GRAPHES ET RESEAUX SOCIAUX**

Contrôle Terminal

Date : 04/01/2023 Durée : 2H00
SUPPORTS DE COURS AUTORISES

Exercice 1 (10 points)

On considère une collection de documents scientifiques écrits par des auteurs, que l'on souhaite représenter sous forme de graphes. Chaque document D est écrit par un auteur ou plusieurs co-auteurs. On considère deux types de liens :

- **Lien de citation** entre documents $Cite(D_i, D_j)$: un document D_i cite un autre document D_j .
- **Lien de collaboration** entre auteurs $Coll(A_k, A_l)$: 2 auteurs A_k et A_l collaborent s'ils sont co-auteurs d'au moins une publication.

Soit le corpus de documents suivants où sont indiquées les informations sur les documents et (co-)auteurs associés ainsi que les liens de citation entre documents.

A1, A2 : D1 [D1->D2 ; D1->D4] (Lire : A1, A2 co-auteurs de D1 ; D1 cite D2 ; D1 cite D4)
A2, A3 : D2 [D2->D4] (Lire : A2, A3 co-auteurs de D2 ; D2 cite D4)
A3 : D3 [D3->D4] (Lire : A3 auteur de D3 ; D3 cite D4)
A1, A4 : D4 [D4->D1] (Lire : A1, A4 co-auteurs de D4 ; D4 cite D1)

Le tableau suivant présente la fréquence de chaque mot dans chaque document D1, D2, D3 et D4.

	covid	vaccin	virus	symptôme	respiration
D1	3	4	0	6	0
D2	4	0	4	0	3
D3	5	3	0	4	0
D4	0	0	0	2	0

On considère la représentation de ce corpus grâce à un graphe pondéré des documents $G-Doc$ et un graphe pondéré des auteurs $G-Aut$. On propose de pondérer le lien de citation et lien de collaboration sur la base des intuitions suivantes :

- Le poids d'un lien de citation $wCite(D_i, D_j)$, est d'autant plus important que les documents D_i et D_j ont des contenus similaires.
- Le poids d'un lien de collaboration $wColl(A_k, A_l)$, est d'autant plus important que les auteurs A_k et A_l sont co-auteurs d'un plus grand nombre de documents.

Questions

[Q1] Pondération des liens :

- Proposer et justifier des formules pour le calcul des poids $wCite(D_i, D_j)$ et $wColl(A_k, A_l)$, basées sur les intuitions présentées ci-avant. Préciser clairement vos notations et justifiez vos choix.

- ii) A partir du corpus présenté, illustrer les formules proposées pour calculer $wCite(D1, D2)$; $wColl(A2, A3)$.
- iii) Donnez les représentations graphiques ou représentations matricielles du graphe des documents $G-Doc$ et graphe des auteurs $G-Aut$ issu du corpus présenté.
- iv) Lequel de ces deux graphes : a) répond le mieux à l'hypothèse du « petit monde » (*small world*) ? b) a une structure plus communautaire ? Justifier votre réponse.

- [Q2] On souhaite répondre à la question générale : « *Quels sont les auteurs A_k , qui sont les plus susceptibles de collaborer avec un auteur A_l ?* »
- i) Préciser de quelle tâche d'analyse de graphes, vue en cours, permettrait de répondre à cette question ?
 - ii) Présenter les principales étapes de cette tâche, spécifique au cadre d'application présenté, en vue de répondre précisément à la question : « *Quels sont les auteurs A_k , qui sont les plus susceptibles de collaborer avec l'auteur A_3 ?* »
-

Exercice 2 (10 points)

On souhaite mettre en œuvre un système de recommandation (SR) de *films* sur une plateforme sociale (ex. Netflix). La plateforme a mis en place des partenariats avec des personnalités reconnues (ex. acteurs) pour influencer le comportement de consommateurs. Chaque influenceur lance sa campagne de publicité sur ou plusieurs films au fil du temps. Chaque cinéphile a la possibilité de s'abonner (suivre) un de ces influenceurs avec un degré de confiance (*trust*) donné.

On vous présente les données suivantes concernant les cinéphiles $C1..C5$ et films $f1..f4$:

- Tableau T1 : Matrice des notes des films des cinéphiles aux films sur une échelle 1..5 ;
- Tableau T2 : Pour chaque influenceur $I1..I3$, le(s) films dont il a assuré(s) une campagne publicitaire ;
- Tableau T3 : Matrice d'adjacence du graphe d'abonnement des cinéphiles aux influenceurs $I1..I3$ avec les degrés de confiance associés sur une échelle 0..1.

Questions

- [Q1] Vous souhaitez vérifier l'impact des influenceurs sur la consommation des films par les cinéphiles, en évaluant l'hypothèse selon laquelle « *les films ayant fait l'objet d'une campagne publicitaire par un influenceur I ont plus de chances d'être vus/notés par les cinéphiles C_l qui suivent I* ».

Proposer les étapes d'analyse qui permet de valider vs. invalider cette hypothèse.

- [Q2] Prédire le prochain film i par un cinéphile c en estimant le score de prédiction r_{ci}

- i) Sans considérer le graphe des interactions sociales entre cinéphiles et influenceurs :
 - i.1 Préciser clairement la méthode utilisée ainsi que les étapes associées ;
 - i.2 Illustrer sur un exemple de prédiction de votre choix avec des calculs basés sur les données présentées (tableau T1). La pertinence de votre choix est à justifier.
- ii) En considérant les interactions sociales entre cinéphiles et influenceurs :
 - ii.1 Préciser clairement la méthode utilisée ainsi que les étapes associées ;
 - ii.2 Illustrer sur un exemple de prédiction de votre choix avec des calculs basés sur les données présentées (tableaux T1, T2 et T3). La pertinence de votre choix est à justifier.

	f1	f2	f3	f4
C1	2	3		
C2		5		3
C3	1			1
C4		1	4	
C5	1		2	1

Tableau T1 : Matrice des notes des cinéphiles pour les films
Echelle 1..5

	<i>Publicité</i>
I1	f2 f4
I2	f3
I3	f1

Tableau T2 : Campagnes publicitaires des influenceurs
sur les films

	I1	I2	I3
C1	0.8	0	0.2
C2	0.7	0	0
C3	0	0.9	0
C4	0.1	0.8	0
C5	0	0	0

Tableau T3 : Matrice d'adjacence du graphe d'interactions abonnés-influenceurs avec
pondération des liens basée sur la confiance