



# TP1 : Filtrage collaboratif

Recommandations de films

SOUCCOUCHETTY Darlène | TOKO Lucy - Camille | ZHOU Marine M2 MIAGE





# Recommandations différentes/identiques pour chaque mesure

## Initialisation

10 personnes

1. Lucy
2. Darlène
3. Marine
4. Christian
5. Haru
6. Skander
7. William
8. Louisa
9. Cédric
10. Lisa

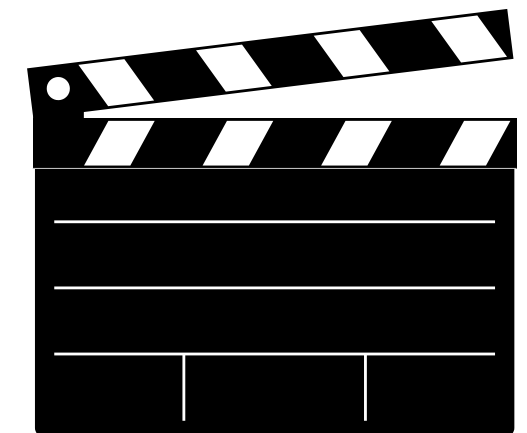
15 films

- Harry Potter
- Interstellar
- Barbie ★★★★★
- Spiderman
- Mon voisin Totoro
- Wall-E
- Mr. Bean
- Hunger games
- Labyrinthe
- Twilight ★★☆☆☆
- Titanic
- Cendrillon
- Madagascar
- Tortue Ninja
- Shrek ★★★★★☆

5 mesures de similarité

Distance de Manhattan\*  
Distance Euclidienne\*  
Coefficient de Pearson  
Fonction Cosinus  
Distance de Tchebychev\*

Film à recommander  
à Anne



*\*NB : Pour les recommandations basées sur des distances (Manhattan, Euclidienne, Tchebychev). On considère que la mesure de similarité correspond à l'inverse de la distance*



## Distance de Tchebychev

Définition : Soient 2 personnes A et B ayant noté les mêmes  $n$  films et soient  $\{A_1, \dots, A_n\}$  et  $\{B_1, \dots, B_n\}$  avec  $i \in \{1, \dots, n\}$  les ensembles tels que  $A_i$  est la note attribuée au  $i$ -ème film par A et  $B_i$  la note attribuée au  $i$ -ème de B.

$$d(A, B) = \max_{i \in [1, n]} (|A_i - B_i|).$$

→ Avantages principaux de la distance de Tchebychev:

- Mise en avant de la différence maximale entre les évaluations de deux utilisateurs
- Calcul simple et intuitif



# Recommandations différentes/identiques pour chaque mesure

## Génération de dictionnaires

	Lisa	Haru	Anne
Harry Potter		2	
Interstellar	3.5		4.5
Barbie	2		5.0
Spiderman	4.5	4.5	
Mon voisin Totoro			
Wall-E		5.0	2.0
Mr. Bean			
Hunger games	4.0	3.0	
Labyrinthe	1.0	2.5	
Twilight		4.0	
Titanic	3.5		
Cendrillon			2.5
Madagascar			
Tortue Ninja			
Shrek		3.5	

- Utilisation du module aléatoire *random*
- Génération aléatoire de notes comprises entre 0 et 5 (pas = 0.5) pour chaque film
- Anne n'a pas vu au moins 50% des films (random)
- Fonction récursive tant que le dictionnaire ne respecte pas le taux de cellules vides (entre 30% et 50%)

# Recommandations différentes/identiques pour chaque mesure

## Suggestions des films

### Itération n

	Lisa	Haru	Anne
Harry Potter		2	
Interstellar	3.5		4.5
Barbie	2		5.0
Spiderman	4.5	4.5	
Mon voisin Totoro			
Wall-E		5.0	2.0
Mr. Bean			
Hunger games	4.0	3.0	
Labyrinthe	1.0	2.5	
Twilight		4.0	
Titanic	3.5		
Cendrillon			2.5
Madagascar			
Tortue Ninja			
Shrek		3.5	

Ensemble = {"Harry Potter", "Barbie", "Interstellar"}

Manhattan → "Harry Potter"

Euclidienne → "Barbie"

Pearson → "Interstellar"

Cosinus → "Harry Potter"

Tchebychev → "Barbie"

Recommandations différentes : Tant que l'ensemble n'est pas composé de 5 films différents

Recommandations identiques : Tant que l'ensemble est composé de plus d'1 film



### Itération n+1

	Lisa	Haru	Anne
Harry Potter			4.0
Interstellar	2.0	3.5	
Barbie			
Spiderman		4.5	2.0
Mon voisin Totoro			
Wall-E	3.5	5.0	5.0
Mr. Bean	1.0		
Hunger games			
Labyrinthe	4.0		
Twilight			
Titanic	3.5		3.5
Cendrillon		1.5	
Madagascar			
Tortue Ninja	5.0	4.0	
Shrek	4.0		4.0

Ensemble = {"Harry Potter", "Wall-E", "Titanic", "Interstellar"}

Manhattan → "Harry Potter"

Euclidienne → "Wall-E"

Pearson → "Interstellar"

Cosinus → "Harry Potter"

Tchebychev → "Barbie"

Recommandations différentes : Tant que l'ensemble n'est pas composé de 5 films différents

Recommandations identiques : Tant que l'ensemble est composé de plus d'1 film

Génération d'un nouveau dictionnaire aléatoire





# Démonstration interactive des recommandations obtenues

## Exemple de tableaux (recommandations identiques)

Nombre d'itérations effectuées = 35

Taux de cellules vides de la solution = 48,48%

	Titanic	Interstellar	Hunger games	Spiderman	Labyrinthe	Wall-E	Mr. Bean	Shrek	Madagascar	Tortue Ninja	Mon voisin Totoro	Harry Potter	Cendrillon	Twilight	Barbie
Darlène	3.5	2.0	2.5	1.5	2.0	3.0	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Christian	NaN	NaN	NaN	NaN	NaN	2.5	NaN	0.0	3.0	NaN	NaN	NaN	NaN	NaN	NaN
Skander	3.5	1.5	1.0	2.5	0.0	4.5	1.5	0.0	3.0	0.0	3.0	4.0	2.5	2.0	1.5
Lucy	5.0	NaN	NaN	NaN	NaN	NaN	3.5	NaN	NaN	NaN	4.0	NaN	NaN	4.5	NaN
William	3.0	0.0	4.0	NaN	5.0	NaN	2.5	2.0	4.5	3.5	2.0	1.0	4.5	4.5	2.0
Lisa	0.5	5.0	3.5	2.5	3.0	4.0	2.0	NaN	NaN	0.0	NaN	2.5	3.0	NaN	2.5
Haru	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Louisa	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	NaN
Cédric	5.0	0.0	1.0	3.5	3.5	1.5	1.0	4.0	3.0	2.5	5.0	0.0	3.5	5.0	1.5
Marine	4.5	NaN	0.5	NaN	3.5	3.0	NaN	3.0	2.5	4.5	NaN	NaN	3.5	4.5	NaN
Anne	3.0	NaN	1.0	NaN	NaN	2.0	NaN	1.5	NaN	4.0	NaN	NaN	5.0	NaN	3.0

Résultat = {'Twilight'}

Moyenne d'itérations à effectuer (sur 100 itérations) = ~11 itérations



# Démonstration interactive des recommandations obtenues

## Exemple de tableaux (recommandation différentes)

Nombre d'itérations effectuées = 2118

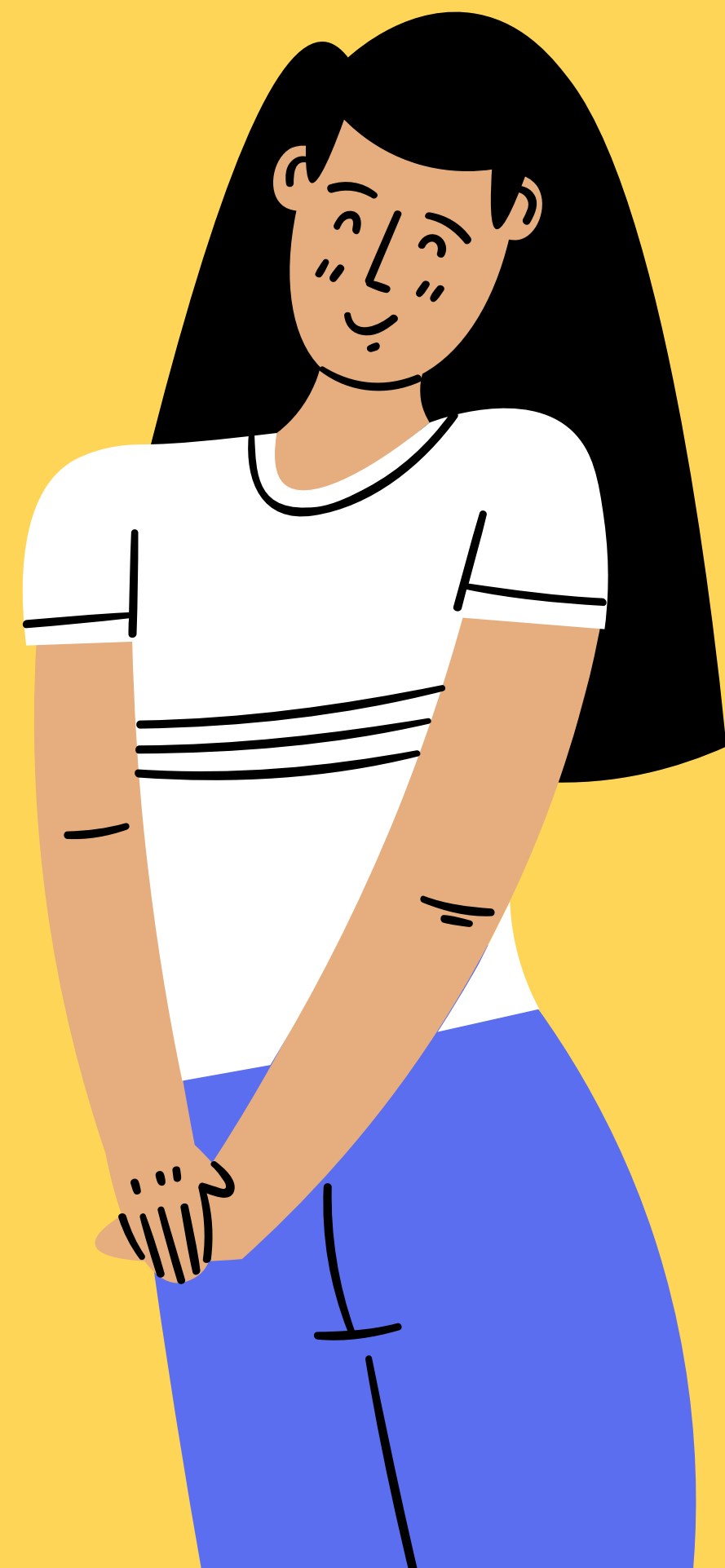
Taux de cellules vides de la solution = 43,01%

	Barbie	Shrek	Labyrinthe	Spiderman	Mr. Bean	Interstellar	Mon voisin Totoro	Twilight	Titanic	Harry Potter	Madagascar	Hunger games	Wall-E	Cendrillon	Tortue Ninja
Darlène	0.5	1.0	0.5	4.5	1.5	4.0	3.0	4.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Christian	3.0	4.5	4.5	1.0	3.0	2.0	4.0	0.0	3.5	1.5	0.0	5.0	3.0	NaN	NaN
Skander	NaN	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Lucy	3.5	3.5	2.0	0.5	2.0	3.5	1.0	2.5	3.5	4.5	1.5	1.5	0.5	0.0	1.5
William	NaN	1.5	2.5	3.0	5.0	3.5	NaN	NaN	NaN	0.5	3.5	3.5	3.0	0.0	NaN
Lisa	NaN	NaN	NaN	NaN	NaN	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Haru	4.5	4.0	2.0	4.5	2.5	0.5	5.0	3.5	0.5	3.5	1.0	4.0	NaN	0.5	4.5
Louisa	NaN	3.0	3.5	3.5	4.0	4.0	2.0	2.5	NaN	3.0	1.5	5.0	3.0	2.5	2.0
Cédric	4.0	NaN	4.0	1.0	1.0	5.0	1.0	0.0	4.0	0.5	NaN	4.0	5.0	3.5	NaN
Marine	3.0	NaN	NaN	1.5	NaN	NaN	3.0	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN
Anne	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.5	NaN	2.0	NaN	0.0	NaN	NaN	NaN

Résultat = {'Spiderman', 'Interstellar', 'Wall-E', 'Madagascar', 'Titanic'}

Moyenne d'itérations à effectuer (sur 100 itérations) = ~2254 itérations

**Merci de votre  
attention !**







# Avantages et limites des mesures de similarité choisies

/	Avantages	Limites
Distance de Manhattan inverse	Calcul simple et intuitif. Utile si les évaluations sont proches les unes des autres pour beaucoup de films.	Forte sensibilité aux valeurs extrêmes, peut donner trop d'importance à une grande différence entre des évaluations (lorsqu'une personne n'a pas vu beaucoup de films par exemple).
Distance Euclidienne inverse	Calcul simple et intuitif, tiens bien compte des différences entre les notes et de l'amplitude des différences.	Forte sensibilité aux valeurs extrêmes.
Coefficient de Pearson	Vérifie s'il y a une relation linéaire entre les évaluations (quand une personne donne une note élevée, l'autre personne a tendance à faire de même si le coefficient est proche de 1)	Sensibilité aux valeurs extrêmes.
Fonction Cosinus	Se base sur une tendance en termes de goûts.	Ne tiens pas assez compte de la valeur numériques des notes.
Distance de Tchebychev inverse	Calcul simple et intuitif. La différence maximale est la plus significative.	Forte sensibilité aux valeurs extrêmes. Donne une importance trop élevée aux valeurs extrêmes.