

# Introduction à l'analyse de réseaux complexes

## 3 - Systèmes de recommandation

Lionel Tabourier

LIP6 – Université Pierre et Marie Curie / CNRS

Cours à l'Université Babeş-Bolyai  
Mai 2017

# Plan

- 1 Introduction – Principe de la recommandation
- 2 Les principales stratégies
- 3 Le problème de la diversité dans la recommandation

## Principe de la recommandation

expérience quotidienne: on consulte un produit  
→ la plateforme recommande une liste de produits

**selon quelle méthode?**

**Définition:** système de **filtrage de l'information**  
pour proposer les plus pertinentes pour les besoins de  
l'utilisateur

## Principe de la recommandation

expérience quotidienne: on consulte un produit  
→ la plateforme recommande une liste de produits

**selon quelle méthode?**

**Définition:** système de **filtrage de l'information**  
pour proposer les plus pertinentes pour les besoins de  
l'utilisateur

# Quelques exemples

## recommandation de vidéos

The screenshot shows the YouTube Music interface. At the top, there's a search bar and navigation links for 'Parcourir', 'Musique', and 'Ajouter une vidéo'. Below the navigation, there's a welcome message: 'Bienvenue dans l'univers musical YouTube | Learn more.' The main section is titled 'Musique' with a dropdown menu for 'Tous les genres'. A 'Tout regarder' button is present. The 'Recommandations musicales' section displays a grid of video thumbnails with titles and durations. The 'En direct sur YouTube' section features a live stream by 'Tiësto - Live Q&A'. The 'Top des musiques' section shows a list of top tracks for the week of August 23, 2011, including 'Je peux dead' by Kim, 'Super Bass' by Nicki Minaj, 'You and I' by Lady Gaga, 'Party Rock Anthem' by LMFAO, and 'Last Friday Night (T.G.I.F.)' by Katy Perry. At the bottom, there are category filters: 'Plus de musique', 'R&B', 'Jazz', 'Musique du monde', 'Théâtre et cinéma', and 'Pop/Rock'.

YouTube

Parcourir Musique Ajouter une vidéo

lioustone

Vidéos Musique En direct

Bienvenue dans l'univers musical YouTube | Learn more.

Musique Tous les genres

Tout regarder

Recommandations musicales

Magic Dance  
Parce que vous avez regardé Guem, concert Paris, 6 mai 2008

Guem Concert Elysée Montn  
Parce que vous avez regardé Guem, concert Paris, 6 mai 2008

Guem et Zaka Percussion  
Parce que vous avez regardé Guem, concert Paris, 6 mai 2008

Tajianska  
de Goran Bregovic  
Parce que vous avez regardé Tajianska -Goran Bregovic-

Titi Robin & Gulabi Sapera  
Parce que vous avez regardé Jivula, Petite mère sulfure

Zakir Hussain in Perth - Maa! Tabla and Bansuri dialogue. Zakir  
Parce que vous avez regardé Jivula, Petite mère sulfure

TITI ROBIN & ZE LUIS NASTI  
Parce que vous avez regardé Jivula, Petite mère sulfure

Jaco Pastorius  
Parce que vous avez regardé Jaco Pastorius-"Donna Lee"

Titi Robin Trio, Jaipur  
Parce que vous avez regardé Jivula, Petite mère sulfure

Titi ROBIN à AUCAMVILLE | Still  
de Anishai Cohen  
Parce que vous avez regardé Anishai Cohen - Shir Prielis -

Dub FX - Love Me Or Not  
Parce que vous avez regardé Dub FX 10/10/2008 Love

En direct sur YouTube

Tiësto - Live Q&A  
de officialtiësto  
15:00 (EDT)

Top des musiques

Tout regarder

Semaine du 23 août 2011

1 Je peux dead de Kim

2 Super Bass de Nicki Minaj

3 You and I de Lady Gaga

4 Party Rock Anthem de LMFAO

5 Last Friday Night (T.G.I.F.) de Katy Perry

Plus de musique R&B Jazz Musique du monde Théâtre et cinéma Pop/Rock


## Quelques exemples

### recommandation de produits culturels

Close

### Other Movies You Might Enjoy

[Amélie](#)




Add

☆☆☆☆☆

Ⓢ Not Interested

[Y Tu Mama Tambien](#)




Add

☆☆☆☆☆

Ⓢ Not Interested

[Guys and Dolls](#)




Add

☆☆☆☆☆

Ⓢ Not Interested

[Mostly Martha](#)




Add

☆☆☆☆☆

Ⓢ Not Interested

[Only Human](#)




Add

☆☆☆☆☆

Ⓢ Not Interested

[Russian Dolls](#)




Add

☆☆☆☆☆

Ⓢ Not Interested

Close



**Eiken has been added to your Queue at position 2.**

This movie is available now.

Move To Top Of My Queue

---

[< Continue Browsing](#)      [Visit your Queue >](#)

## Quelques exemples

### recommandation de produits variés

#### Produits fréquemment achetés ensemble



Prix pour les trois: EUR 16,52

[Ajouter ces trois articles au panier](#)

[Afficher la disponibilité du produit et le mode de livraison](#)

- Cet article** : Fahrenheit 451 de Ray Bradbury Broché EUR 4,84
- Le Meilleur des mondes de Aldous Huxley Poche EUR 4,27
- 1984 de George Orwell Poche EUR 7,41

#### Les clients ayant acheté cet article ont également acheté

Page :



Le Meilleur des mondes  
de Aldous Huxley  
★★★★☆ (46)  
EUR 4,27



1984 de George Orwell  
★★★★★ (95)  
EUR 7,41



Chroniques martiennes  
de Ray Bradbury  
★★★★☆ (17)  
EUR 5,41



Fahrenheit 451 de Ray  
Bradbury de Guy Renotte  
EUR 5,70



Un bonheur  
insoutenable de Ira Levin  
★★★★☆ (13)  
EUR 5,70

# Quelques exemples

## recommandation d'informations

	<b>Josh Widdicombe</b>  @joshwiddicombe Comedian and Countryfile fan. DVD out now <a href="https://amazon.co.uk/Josh-Widdicombe...">amazon.co.uk/Josh-Widdicombe...</a>	 Suivre 
	<b>Pierre Boudes</b> @PierreBoudes Maître de conférences en informatique, université Paris 13 (@m2pls, VP SI @univ_paris13) GPG: 0x9b852dca86196fc3 <a href="https://mamot.fr/@pierreb">mamot.fr/@pierreb</a> Suivi par Raphaël FS et SéminaireCodeSources.	 Suivre 
	<b>Joe Wilkinson</b>  @gillinghamjoe I'm not aging brilliantly if i'm honest	 Suivre 
	<b>Olivier Berger</b> @olberger Research Engineer, FLOSS etc. Suivi par SéminaireCodeSources.	 Suivre 
	<b>Pénélope Bagieu</b>  @PenelopeB Now the party don't start til I walk in	 Suivre 
	<b>Le Cnam</b>  @LeCnam Conservatoire national des arts et métiers Grand établissement d'enseignement supérieur dédié à la	 Suivre 



## Quelques exemples

### recommandation dans les moteurs de recherche

**Google** Agora Paris Saint Germain

Environ 276 000 résultats (0,20 secondes) [Recherche avancée](#)

**Tout**  
[Images](#)  
[Vidéos](#)  
[Actualités](#)  
[Shopping](#)  
[Plus](#)

**Paris**  
[Changer le lieu](#)

**Le Web**  
[Pages en français](#)  
[Pays : France](#)  
[Pages en langue étrangère traduites](#)  
[Plus d'outils](#)

[Hotel Agora Paris Saint Germain \\*\\*\\*, Official Site - Near Notre-Dame](#) ☆ 🔍 - [ Traduire cette page ]  
Hotel **Agora Paris Saint-Germain** \*\*\*, a charming hotel in Paris near the Pantheon and Notre-Dame district.  
[www.hotelagorasaintgermain.com/](#) - [En cache](#)

[Agora Saint Germain \(Paris, France\) - Avis Hôtel - TripAdvisor](#) ☆ 🔍  
★★★★☆ 350 avis  
**Agora Saint Germain, Paris** : Consultez les 350 avis de voyageurs, 120 photos, et les meilleures offres pour **Agora Saint Germain, ...**  
[www.tripadvisor.fr/Hotel\\_Review-g187147-d222912-Reviews-Agora\\_Saint\\_Germain-Paris\\_Ile\\_de\\_France.html](#) - [En cache](#)

[Hôtel Agora Saint Germain Paris : Hotel Paris](#) ☆ 🔍  
Hôtel **Agora Saint Germain Paris** 3 étoiles : Hotel Paris.  
[Afficher le plan de 42 Rue des Bernardins, 75005 Paris](#)  
[fr.federal-hotel.com](#) > ... > **Hotel Paris 05** - [En cache](#) - [Pages similaires](#)

 **Hotel Agora Paris Saint-Germain** ☆ 🔍  
[Page Google Adresses](#)  
42, Rue des Bernardins 75005 Paris - 01 46 34 13 00  
★★★★☆ 158 avis - [Donner votre avis](#)  
 **Cedric Sahuc** a trouvé cette adresse **géniale**. "Very nice boutique hotel in **Paris**, friendly staff and free wifi :)"

## Quelques aspects du problème

- collecte d'information
- **calcul des recommandations pertinentes:**
  - constituer une banque d'informations produits
  - construire un "modèle utilisateur"
  - extraire une liste de recommandations du modèle
- présentation des résultats (non étudié ici)

## Quelques aspects du problème

- collecte d'information
- **calcul des recommandations pertinentes:**
  - constituer une banque d'informations produits
  - construire un "modèle utilisateur"
  - extraire une liste de recommandations du modèle
- présentation des résultats (non étudié ici)

milliers d'algorithmes différents, regroupés en stratégies

# Plan

- 1 Introduction – Principe de la recommandation
- 2 Les principales stratégies
- 3 Le problème de la diversité dans la recommandation

## Collecte d'information

### Filtrage actif, données explicites

- utilisateur donne explicitement son intérêt
- **avantage:** qualité de l'information
- **inconvénient:** biais de déclaration  
surestimation, sous-estimation, stratégie utilisateur ...

exemples: étoiles Amazon, *liking*, ajout favori ...

## Collecte d'information

### Filtrage passif, données implicites

- en général pas de demande à l'utilisateur
- **avantage:** pas de biais de déclaration
- **inconvénient:** biais d'attribution  
plusieurs utilisateurs, achat pour qqun d'autre ...  
exemples: fréquence de consultation, clics, achat ...

# Les différentes approches algorithmiques

Une typologie en trois approches:

1. Personnalisée
2. Orientée contenu (*content-based filtering*)
3. Sociale (*collaborative filtering*)

En pratique aujourd'hui: **hybride**, utilisant les 3 aspects

# Les différentes approches algorithmiques

Une typologie en trois approches:

1. Personnalisée
2. Orientée contenu (*content-based filtering*)
3. Sociale (*collaborative filtering*)

En pratique aujourd'hui: **hybride**, utilisant les 3 aspects



## Recommandation personnalisée

Information utilisée: **comportement passé de l'utilisateur**

- historique de navigation
- historique de recherche
- clics

Utilisée seule, on ne recommande des contenus déjà consultés  
⇒ toujours utilisée avec d'autres sources d'informations

## Recommandation personnalisée

Information utilisée: **comportement passé de l'utilisateur**

- historique de navigation
- historique de recherche
- clics

Utilisée seule, on ne recommande des contenus déjà consultés  
⇒ toujours utilisée avec d'autres sources d'informations

# Recommandation orientée contenu

## Basée sur les propriétés de l'objet

Principe usuel:

- extraire caractéristiques d'un contenu  
ex: mots d'une page web, catégories du système . . .
- comparaisons aux préférences de l'utilisateur

Dépend de la qualité de profilage des contenus

# Recommandation orientée contenu

## Basée sur les propriétés de l'objet

Principe usuel:

- extraire caractéristiques d'un contenu  
ex: mots d'une page web, catégories du système . . .
- comparaisons aux préférences de l'utilisateur

Dépend de la qualité de profilage des contenus

## Profilage des contenus

Trouver des **caractéristiques significatives**:

- mots d'une page web
- catégories pour un film
- auteur d'un livre . . .

## Profilage des contenus

Trouver des **caractéristiques significatives**:

- mots d'une page web
- catégories pour un film
- auteur d'un livre ...

### Un cas pratique: recommandation de documents

- profilage par les termes (ou leur racine)
- comment trouver les termes significatifs?  
traitement du langage naturel
- simple comptage? **pas adapté**
- *term frequency–inverse document frequency* (TF.IDF)
  - fréquence d'un mot: son importance
  - inverse de la fraction des documents qui le contiennent:  
à quel point il est informatif

## Profilage des contenus

Trouver des **caractéristiques significatives**:

- mots d'une page web
- catégories pour un film
- auteur d'un livre ...

### Evaluer la similarité

- Exemple de l'indice de Jaccard

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

- Exemple de la similarité cosinus
- ...

## Profilage des contenus

**Problème de recommandation orientée contenu**  
des infos sont faciles à intégrer (ex: genre)  
d'autres non: tout ce qui relève de la **subjectivité**



# Recommandation sociale

## **Recommandation basée sur les similarités entre utilisateurs**

# Recommandation sociale

## Recommandation basée sur les similarités entre utilisateurs

### Représentation matricielle des données

Filtrage actif:

Movie ratings

	Amy	Jef	Mike	Chris	Ken
The Piano	-	-	+		+
Pulp Fiction	-	+	+	-	+
Clueless	+		-	+	-
Cliffhanger	-	-	+	-	+
Fargo	-	+	+	-	?

## Recommandation sociale

### Recommandation basée sur les similarités entre utilisateurs

#### Représentation matricielle des données

Avec filtrage passif:

	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$
$a_1$		1		1	1		
$a_2$	1				1		
$a_3$	1			1			1
$a_4$				1		1	
$a_5$					1	1	
$a_6$			1			1	

# Recommandation sociale

## Recommandation basée sur les similarités entre utilisateurs

### Représentation matricielle des données

#### Recherche de similarité entre produits, entre utilisateurs

#### Remarques:

- la taille des matrices peut être énorme
- matrices souvent très creuses
- le profilage prend souvent le temps en compte

## Centrée utilisateur (*user-based*)

Movie ratings

	Amy	Jef	Mike	Chris	Ken
The Piano	-	-	+		+
Pulp Fiction	-	+	+	-	+
Clueless	+		-	+	-
Cliffhanger	-	-	+	-	+
Fargo	-	+	+	-	?

### Principe:

- sélectionner un panel d'utilisateurs selon similarité
- prédire l'appréciation des produits avec panel

## Centrée utilisateur (*user-based*)

Movie ratings

	Amy	Jef	Mike	Chris	Ken
The Piano	-	-	+		+
Pulp Fiction	-	+	+	-	+
Clueless	+		-	+	-
Cliffhanger	-	-	+	-	+
Fargo	-	+	+	-	?

### Hypothèses:

- préférences d'un utilisateur restent stables dans le temps
- les utilisateurs qui ont eu des goûts similaires dans le passé auront des goûts similaires dans le futur

## Centrée utilisateur (*user-based*)

Movie ratings

	Amy	Jef	Mike	Chris	Ken
The Piano	-	-	+		+
Pulp Fiction	-	+	+	-	+
Clueless	+		-	+	-
Cliffhanger	-	-	+	-	+
Fargo	-	+	+	-	?

### Remarques et pistes de réflexions:

- biais vers certains individus (ceux qui notent beaucoup, positivement ...) ⇒ à corriger
- comment sélectionner les plus similaires? et combien?
- millions utilisateurs ⇒ **recherche en temps réel impossible**

## Centrée objet (*item-based*)

### Principe:

- même principe, mais similarité entre objets (contenus)
- puis évaluation avec utilisateurs voisins



## Avantages et inconvénients de la reco. sociale

- pas besoin d'utiliser la sémantique du contenu
- si beaucoup d'utilisateurs, nécessité d'un traitement hors-ligne
- problèmes du **départ à froid** (*cold start*) et de la **rareté** (*sparsity*): le système a besoin de beaucoup d'informations  
ex: mêmes les utilisateurs les plus proches auront peu d'articles recommandés en commun

Comment contourner ces problèmes?

Attribuer un score par défaut en utilisant une autre méthode

**hybridation**

## Avantages et inconvénients de la reco. sociale

- pas besoin d'utiliser la sémantique du contenu
- si beaucoup d'utilisateurs, nécessité d'un traitement hors-ligne
- problèmes du **départ à froid** (*cold start*) et de la **rareté** (*sparsity*): le système a besoin de beaucoup d'informations  
ex: mêmes les utilisateurs les plus proches auront peu d'articles recommandés en commun

Comment contourner ces problèmes?

Attribuer un score par défaut en utilisant une autre méthode

**hybridation**

## Lien avec les réseaux complexes

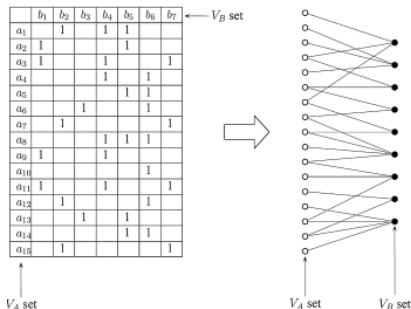
### Des graphes bipartis

	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$
$a_1$		1		1	1		
$a_2$	1				1		
$a_3$	1			1			1
$a_4$				1		1	
$a_5$					1	1	
$a_6$			1			1	

**Définition:** graphes dans lesquels les nœuds peuvent être regroupés en deux familles  $A$  et  $B$ , et il n'existe des liens qu'entre ces deux familles.

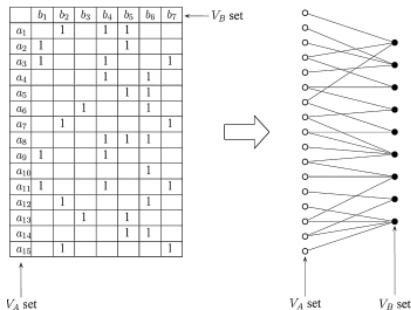
# Lien avec les réseaux complexes

## Des graphes bipartis



**Définition:** graphes dans lesquels les nœuds peuvent être regroupés en deux familles  $A$  et  $B$ , et il n'existe des liens qu'entre ces deux familles.

## Lien avec les réseaux complexes



Recommandation sociale  $\sim$  “compléter” un graphe biparti

## Recommandation dans quelques cas pratiques

### Amazon

cœur du système: filtrage collaboratif basé contenus

*“les clients ayant acheté cet article ont également acheté”*

de 20 à 30% du Chiffre d’Affaires directement lié à la recommandation

## Recommandation dans quelques cas pratiques

### Google search engine

PageRank est une forme de recommandation sociale  
modifiée par l'historique et analyse sémantique des contenus

⇒ **un système de recommandation hybride**

## Recommandation dans quelques cas pratiques

### Pandora vs LastFM (années 2000-2010)

- 2 radios en ligne, recommandations différentes
- **Pandora** est à dominante contenu:  
profil “ADN” des morceaux musicaux
- **LastFM** est à dominante sociale:  
basée sur du filtrage collaboratif classique  
collecte de données exhaustive
- analyse comparée:
  - détection de nouveaux artistes:  
LastFM démarre à froid  
mais Pandora plus lent car demande analyse manuelle
  - qualité des recommandations:  
tendance des systèmes sociaux à peu renouveler?



# Plan

- 1 Introduction – Principe de la recommandation
- 2 Les principales stratégies
- 3 Le problème de la diversité dans la recommandation

## Bulles de filtre et chambre d'écho

*Filter-bubbles and echo chambers*

### Bulle à filtre

*Eli Pariser, TED Talk, 2011*

La tendance du système à **enfermer l'utilisateur** dans une  
“bulle culturelle”

Pariser décrit comment le système tend à conforter l'utilisateur  
plutôt que stimuler l'esprit critique.

Aspect important: caractère implicite de l'enfermement.

## Bulles de filtre et chambre d'écho

### *Filter-bubbles and echo chambers*

#### **Chambre d'écho**

La tendance du système à ne proposer à l'utilisateur que des avis en accord avec ses propres opinions.

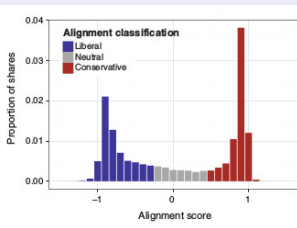
B.Gates: *“(Social media) let you go off with like-minded people, so you are not mixing and sharing and understanding other points of view [ . . . ] It's turned out to be more of a problem than I, or many others, would have expected.”*

## Le cas Facebook

Exposure to ideologically diverse news and opinion on Facebook, Bakshy *et al.*, *Science*, 2015

### Mesurer “filter bubble” à grande échelle

- 10 millions d'utilisateurs
- alignement politique déclaré (de *liberal* à *conservative*)
- 226.000 URLs partagées, alignement politique évalué

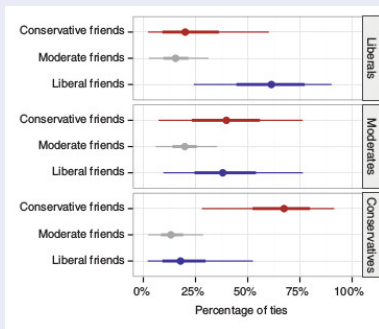


## Le cas Facebook

Exposure to ideologically diverse news and opinion on Facebook, Bakshy *et al.*, *Science*, 2015

### Mesurer “filter bubble” à grande échelle

- Mesure les “*cut-across ties*”

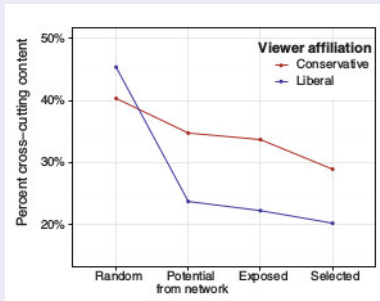


## Le cas Facebook

Exposure to ideologically diverse news and opinion on Facebook, Bakshy *et al.*, *Science*, 2015

### Mesurer “filter bubble” à grande échelle

- Mesure les “*cut-across clicks*”



## Le cas Facebook

**Conclusion principale de l'étude:**  
**L'homophilie** provient avant tout du réseau d'amis

Janvier 2017: polémique des "*Fake news*"  
Facebook retire la personnalisation de ses *Trending Topics*

## Le cas Facebook

**Conclusion principale de l'étude:**  
L'**homophilie** provient avant tout du réseau d'amis

Janvier 2017: polémique des "*Fake news*"  
Facebook retire la personnalisation de ses *Trending Topics*



## Pourquoi?

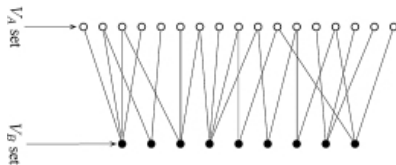
Schématiquement, l'algorithme souhaite  
"compléter un graphe biparti"



Divergence entre qualité pour l'algorithme  
et satisfaction (besoin) pour l'utilisateur

## Pourquoi?

Schématiquement, l'algorithme souhaite  
"compléter un graphe biparti"



Divergence entre qualité pour l'algorithme  
et satisfaction (besoin) pour l'utilisateur

## Mesures de qualité

**Cas binaire** (événement se produit ou non)  
ex: est-ce que l'utilisateur achète le produit? oui/non

	prévision: +	prévision -
réalité: +	vrai positif	faux négatif
réalité: -	faux positif	vrai négatif

**remarque:** notions très générales en prévision binaire

## Mesures de qualité

- **precision** (précision)

$$\mathbf{Pr} = \frac{\#vp}{\#vp + \#fp}$$

- **recall** (rappel)

$$\mathbf{Rc} = \frac{\#vp}{\#vp + \#fn}$$

- **F1-score**, compromis précision/rappel

$$\mathbf{F1} = \frac{2 \cdot \mathbf{Pr} \cdot \mathbf{Rc}}{\mathbf{Pr} + \mathbf{Rc}}$$

Pour l'algorithme, **bonne prévision**  $\equiv$  F1 élevé  
Qu'est-ce que la **satisfaction** de l'utilisateur?  
Qu'est-ce que le **besoin** de l'utilisateur?

## Mesures de qualité

- **precision** (précision)

$$\mathbf{Pr} = \frac{\#vp}{\#vp + \#fp}$$

- **recall** (rappel)

$$\mathbf{Rc} = \frac{\#vp}{\#vp + \#fn}$$

- **F1-score**, compromis précision/rappel

$$\mathbf{F1} = \frac{2 \cdot \mathbf{Pr} \cdot \mathbf{Rc}}{\mathbf{Pr} + \mathbf{Rc}}$$

Pour l'algorithme, **bonne prévision**  $\equiv$  **F1 élevé**

Qu'est-ce que la **satisfaction** de l'utilisateur?

Qu'est-ce que le **besoin** de l'utilisateur?

## Comment introduire de la diversité?

Question existe depuis plus de 10 ans ...

Improving Recommendation Lists Through Topic Diversification  
*WWW 2005, Ziegler et al.*

### Un filtrage collaboratif modifié

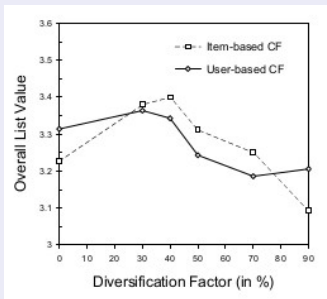
- filtrage collaboratif  $\Rightarrow$  liste de recommandations
- définit mesure de diversité de la liste (*Intra-list similarity*)
- augmenter la diversité en ajoutant des éléments moins bien classés

## Comment introduire de la diversité?

Question existe depuis plus de 10 ans ...

Improving Recommendation Lists Through Topic Diversification  
*WWW 2005, Ziegler et al.*

### Un filtrage collaboratif modifié



## Maintenant

### Projet Algodiv

Projet de recherche (ANR):

- collaboration sociologues (Institut de Sciences Politiques, Orange Labs) et d'informaticiens (LIP6, CAMS)
- **pour redéfinir et favoriser la diversité des systèmes de recommandation**

### Quelques idées à développer

- structure communautaire du graphe de recommandation
- *gatekeepers* du réseau
- maximiser des “flots” de diversité
- donner un retour à l'utilisateur