الجمهورية الجزائرية الديمقراطية الشعبية

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DR. YAHIA FARES DE MEDEA

FACULTE DES SCIENCES

**Département de Mathématiques et Informatique**



# Mémoire de Fin d'Études

**Présenté pour l'obtention du Diplôme**
**de Master en Informatique**

*Option* : Ingénierie des systèmes et technologie web

## THEME

# Système de Recommandation Sociale

**Proposé et dirigé par,**

Mme SADOUKI Fatiha

**Préparé et présenté par,**

Melle. Aicha GUERBA

Mme. Nassima BENMEHIDI

# Dédicace

*Je dédie ce travail*

*A mon fils Yassine,*

*A  ma famille,*

*Et à toutes mes collègues.*

**BENMEHIDI Nassima**

*Je dédie ce travail*

*A mon père et ma mère,*

*A mes frères Nasereddine et Mustapha, leurs femmes et leurs enfants,*

*A mon petit frère Abdennour,*

*A mes sœurs Fatma zohra , Khadidja et surtout ma petite sœur Meriem,*

*Et à toutes mes amies.*

**GUERBA Aicha**

# Remerciements

Nous tenons à remercier tout d'abord ***Madame SADOUKI Fatiha,*** promotrice de ce mémoire pour avoir initié ce travail de recherche, ainsi que pour son encadrement dans un atmosphère toujours conviviale au bon déroulement de ce travail.

Nous remercions tous les responsables de l'université Yahia FARES de Médéa, spécialement le département informatique, de nous avoir accepté et donné cette chance de passer le master.

Merci à tous les enseignants d'informatique de l'université de Médéa pour leurs efforts, leurs sérieux et leurs engagements.

Merci à tous les étudiants du master ISTW pour leurs partages d'informations, leurs esprits de groupe et leurs encouragements sur Facebook.

Enfin, nous ne saurions terminer sans remercier ceux qui ont toujours été au plus près de nous en veillant à notre réussite, nos familles.

# Résumé

Ces dernières années, le contenu disponible sur le Web a augmenté de manière considérable dans ce qu'on appelle communément le Web social. Pour l'utilisateur, il devient de plus en plus difficile de recevoir du contenu de qualité sans se voir rapidement submergé par le flot incessant d'information. L'utilisateur a besoin d'informations plus précises qui reflètent ses intérêts. Nous nous concentrons sur la détection du réseau social de l'utilisateur qui représente un élément clé pour améliorer la recommandation. L'originalité de notre approche est basée sur la proposition d'une technique de détection de liens sociaux sur la base du comportement d'annotation des utilisateurs afin de construire un réseau social, puis analyser ce réseau et faire enfin une recommandation sociale. L'approche proposée a été testée et évaluée sur le jeu de données des vidéos MovieLens.

Mots clés : recommandation, recommandation sociale, tag, réseau social.

# Abstract

In recent years, content available on the Web has increased significantly in what is commonly called the social Web. For the user, it becomes more and more difficult to receive quality content without being quickly overwhelmed by the incessant flow of information. The user needs more precise information that reflects his interests. We focus on detecting the user's social network which represents a key element to improve the recommendation. The originality of our approach is based on a technique of detection of social links based on user annotation behavior in order to build a social network, then analyze this network and finally make a social recommendation. The proposed approach has been tested and evaluated on the MovieLens video data set.

Keywords: recommendation, social recommendation, tag, social network.

**الملخص**

*في السنوات الأخيرة، زاد المحتوى المتاح على شبكة الإنترنت بشكل ملحوظ في ما يسمى عادة الشبكة الاجتماعية. بالنسبة للمستخدم، يصبح أكثر وأكثر صعوبة الحصول على محتوى عالي الجودة دون أن تطغى بسرعة تدفق مستمر من المعلومات. يحتاج المستخدم إلى معلومات أكثر دقة تعكس اهتماماته. ونحن نركز على الكشف عن الشبكة الاجتماعية للمستخدم الذي يمثل عنصرا أساسيا لتحسين التوصية. وتستند أصالة نهجنا على تقنية الكشف عن الروابط الاجتماعية القائمة على سلوك تعليقات المستخدم من أجل بناء شبكة اجتماعية، ثم تحليل هذه الشبكة، وأخيرا تقديم توصية* اجتماعية.
وقد تم اختبار النهج المقترح وتقييمه على مجموعة بيانات الفيديو Movielens

*كلمات البحث : توصية، توصية الاجتماعية، العلامة، الشبكة الاجتماعية.*

# SOMMAIRE

**Introduction générale** ..... **1**
**Partie 1 : Etat de l'art** ..... **3**
**1 Systèmes de recommandation** ..... **4**
**1.1 Introduction** ..... **4**
**1.2 Approches de recommandation** ..... **4**
1.2.1 Approches basées sur le contenu ..... 4
1.2.1.1 Principe ..... 4
1.2.1.2 Limitations des approches basées sur le contenu ..... 5
1.2.2 Approches collaboratives ..... 7
1.2.2.1 Filtrage collaboratif basé sur les utilisateurs (U2U) ..... 7
1.2.2.2 Filtrage collaboratif basé sur les items (I2I) ..... 8
1.2.2.3 Limitations du filtrage collaboratif ..... 9
1.2.3 Approches démographiques ..... 9
1.2.4 Approches sociales ..... 10
1.2.5 Approches basées sur la connaissance ..... 11
1.2.6 Approches hybrides ..... 12
1.2.6.1 Hybridation pondérée ..... 13
1.2.6.2 Hybridation à Commutation ..... 13
1.2.6.3 Hybridation mixée ..... 13
1.2.6.4 Hybridation par combinaison de caractéristiques ..... 13
1.2.6.5 Hybridation en cascade ..... 13
1.2.6.6 Hybridation par Augmentation de caractéristiques ..... 14
1.2.6.7 Hybridation méta-niveau ..... 14
**1.3 Caractéristiques du modèle utilisateur** ..... **14**
**1.4 Modélisation du profil utilisateur** ..... **16**
1.4.1 Sélection de données pour les profils utilisateurs ..... 16
1.4.1.1 Producteurs de données ..... 16
1.4.1.2 Sources de données ..... 16
1.4.2 Prétraitement de données ..... 17
1.4.3 Structuration des données pour les profils utilisateurs ..... 17
1.4.3.1 Données explicites (feedback explicites) ..... 17

1.4.3.2 Données implicites (feedback implicites) ..... 17
1.4.3.3 Données de contexte ..... 17
1.4.3.4 Données sémantiques ..... 18
1.4.3.5 Données de sécurité ..... 18
1.4.4 Fouille de données pour les profils utilisateurs ..... 18
1.4.4.1 Modélisation comportementale (behaviour modeling) ..... 19
1.4.4.2 Modélisation des centres d'intérêts (interest modeling) ..... 19
1.4.4.3 Modélisation des intentions (intention modeling) ..... 19
**1.5 Evaluation et de validation des performances du système ..... 19**
1.5.1 Evaluation en ligne ..... 19
1.5.2 Evaluation offline ..... 20
1.5.3 Précision ..... 21
1.5.3.1 Précision de la prédiction des votes ..... 21
1.5.3.2 Précision d'une Top-N liste ..... 22
**1.6 Conclusion ..... 23**
**2 Recommandation sociale ..... 24**
**2.1 Introduction ..... 24**
**2.2 Médias sociaux ..... 24**
**2.3 Analyse des réseaux sociaux ..... 27**
2.3.1 Définition de graphe ..... 27
2.3.2 Définition de Réseau social ..... 28
2.3.3 Mesures de centralité issues de la Recherche d'Information (RI) ..... 28
2.3.3.1 Centralite de degre (degree centrality) ..... 29
2.3.3.2 Centralité de proximité (closeness centrality) ..... 29
2.3.3.3 Centralité d'intermédiarité (betweenness centrality) ..... 29
2.3.3.4 PageRank ..... 30
2.3.3.5 HITS (Hypertext Induced Topic Search) ..... 31
**2.4 Recommandation sociale ..... 32**
2.4.1 Définition ..... 32
2.4.2 Modélisation d'un réseau d'information sociale ..... 34
2.4.2.1 Modélisation d'un réseau d'information sociale ..... 34
2.4.2.2 Processus de recherche d'information sociale ..... 35
**2.5 Conclusion ..... 35**

**Partie 2 : Approche proposée pour le système de recommandation sociale ..... 36**
**3 Modélisation ..... 37**
**3.1 Introduction ..... 37**
**3.2 Architecture générale du système ..... 37**
**3.3 Création du réseau social ..... 39**
3.3.1 Création du réseau ..... 39
3.3.2 Pondération du réseau ..... 41
**3.4 Analyse du réseau d'information sociale ..... 43**
3.4.1 Calcul du score social des utilisateurs (Page Rank) ..... 43
3.4.2 Calcul du score de l'importance sociale des ressources ..... 43
**3.5 Classement des ressources et recommandation ..... 44**
**3.6 Mise en œuvre de l'approche proposée ..... 44**
**3.7 Conclusion ..... 45**
**4. Evaluation et validation ..... 46**
**4.1 Introduction ..... 46**
**4.2 Cadre d'évaluation ..... 46**
4.2.1 Jeux de données ..... 46
4.2.2 Jugements de pertinence ..... 47
4.2.3 Métriques d'évaluation ..... 47
**4.3 Evaluation de l'efficacité de l'approche proposée ..... 48**
**Conclusion et perspective ..... 51**
**Références bibliographiques ..... 52**
**Annexe A - Movielens ..... 60**
**Annexe B - Programme Java ..... 63**

# Liste des figures

Figure 1.1 Recommandation démographique…….……10

Figure 1.2 Recommandation sociale……..…….....11

Figure 1.3 Recommandation à base de connaissance…….....12

Figure 2.1 Exemple d'utilisateurs connectés…….…….....32

Figure 2.2 Le réseau d'information sociale…….….……..35

Figure 3.1 Architecture du système de recommandation sociale……..…….....38

Figure 3.2  Exemple de réseau d'information sociale...…….……39

Figure 3.3 Exemple de réseau social…….…….…..….……41

Figure 3.4 Exemple de réseau social pondéré…….……..…….....43

Figure 4.1 Précision du système de recommandation sociale..……..48

Figure 4.2 Evaluation de l'efficacité de l'approche……..…….....48

# Liste des tableaux

Tableau 1.1 Exemple de matrice de notes……..……..8

Tableau 2.1 Comparaison entre filtrage collaboratif et social…….…….....33

Tableau 4.1 Caractéristiques du réseau social……..……47

Tableau 4.2 Apport du modèle ……..…….…….49

# Introduction générale

La surcharge cognitive d'un utilisateur due à une trop grande quantité d'information est un problème majeur des systèmes hypermédia. Les systèmes de recommandation ont été introduits pour résoudre ce problème et sont maintenant utilisés couramment pour améliorer la navigation de l'utilisateur sur le Web (gain en temps et facilité de recherche).

Un système de recommandation est une application qui prévoit les éléments d'information susceptibles d'intéresser l'utilisateur, en fonction du contenu des éléments à recommander, du profil de l'utilisateur, du comportement collectif des utilisateurs. Les éléments d'information peuvent comprendre des marchandises, film, musique, livres, du texte, actualités, blogs, page web et articles.

Il existe de nombreuses sources pour extraire les intérêts d'un utilisateur. Le profil de l'utilisateur contient des informations fournies explicitement par l'utilisateur lui-même, par exemple : nom, prénom, âge, profession, ses intérêts, ou implicitement à travers les intérêts qui peuvent être extraits de son comportement social par exemple : le comportement d'annotation, ou de son réseau social par exemple : ses amis.

Pour atteindre notre principal objectif qui consiste à améliorer la recommandation, il est nécessaire dans un premier temps de définir le contexte social qui revient à identifier, extraire et quantifier, à partir du réseau social, une propriété sociale telle que le co-marquage ; il faut ensuite estimer une pertinence sociale intégrant cette caractéristique sociale, telle que l'importance sociale d'autre part ; et enfin utiliser ce contexte dans la recommandation afin d'améliorer la qualité des ressources recommandées.

L'approche que nous allons proposer est une approche sociale, elle repose sur l'utilisation des informations du réseau social d'un individu afin d'améliorer la pertinence des recommandations qui lui sont faites. Nous pouvons ainsi recommander des objets à un utilisateur en privilégiant les objets qui ont plu à ses amis ou les objets qui ont plu à d'autres utilisateurs qui partagent les mêmes centres d'intérêt que lui. Pour cela, nous allons d'abord construire le réseau social à partir d'un ensemble de tags utilisés par les utilisateurs pour identifier une ressource, ensuite faire l'opération de la recommandation sociale et enfin

l'évaluation du système qui est une phase primordiale car elle sert à tester et valider la fiabilité du système qui sera mis en place. Cette validation sera faite sur une collection de données des préférences filmographiques de grande taille (MovieLens).

Dans ce travail, nous allons étaler les différentes approches de recommandation existantes, nous parlerons aussi du profil utilisateur, de son contenu, des étapes de construction ainsi que les différents modèles. Nous parlerons aussi de la recommandation sociale, des algorithmes d'analyse d'un réseau social et des types de la recommandation sociale. Par la suite nous détailleront notre approche proposée pour la construction d'un système de recommandation sociale. Nous commencerons par extraire le réseau social à partir d'un réseau d'information sociale et ceci sur la base d'un ensemble de tags utilisés par les utilisateurs pour annoter des vidéos. Analyser ensuite le réseau social déduit par une des algorithmes, puis classer et recommander ces ressources. A la fin nous testerons notre solution sur la collection de données MovieLens afin de valider le système.

Ce mémoire est divisé en quatre chapitres. Dans le premier chapitre nous allons présenter le fonctionnement des systèmes de recommandations, les différentes approches de la littérature, et l'évaluation des recommandations. Dans le deuxième chapitre nous entamons le concept du web social, des réseaux sociaux ainsi que la recommandation sociale. Le troisième et le quatrième chapitre concernent respectivement la modélisation et l'évaluation du système de recommandation sociale proposé.

# Partie 1 : Etat de l’art

- Chapitre 1 : Systèmes de recommandation
- Chapitre 2 : Recommandation sociale

# 1 Systèmes de recommandation

## 1.1 Introduction

Un système de recommandations aide les utilisateurs à faire leurs choix dans un domaine où ils disposent de peu d'informations pour trier et évaluer les alternatives possibles. L'importance de tels systèmes apparaît clairement dans les environnements où de gigantesques quantités d'informations en ligne surpassent de loin les capacités de recherche de n'importe quel être humain [75].

L'essor du Web et sa popularité ont notamment contribué à la mise en place de tels systèmes comme dans le domaine du e-commerce. Citons par exemple les sites Web populaires Amazon[1] dans le e-commerce et également CiteSeer[2], outil référençant des articles de recherche. Les systèmes de recommandation peuvent être vus initialement comme une réponse donnée aux utilisateurs ayant des difficultés à prendre une décision dans le cadre d'utilisation d'un système de recherche d'information classique [1].

Dans ce chapitre nous allons présenter les différentes approches de recommandation ensuite nous aborderons le profil utilisateur à savoir ses caractéristiques et sa modélisation et enfin les techniques d'évaluation et de validation de performance d'un système de recommandation.

## 1.2 Approches de recommandation

Ils existent plusieurs approches de recommandation dans la littérature :

### 1.2.1 Approches basées sur le contenu

#### 1.2.1.1 Principe

Le filtrage basé sur le contenu (Content-based Filtering), qui est une évolution générale des études sur le filtrage d'information, s'appuie sur le contenu des documents (thèmes abordés) pour les comparer à un profil lui-même constitué de thèmes. Chaque utilisateur du système

[1] http://www.amazon.com
[2] http://citeseerx.ist.psu.edu

possède alors un profil qui décrit ses propres centres d'intérêt [3]. Deux fonctionnalités centrales ressortent, pour un système de filtrage :

- La sélection des documents pertinents vis-à-vis du profil.
- La mise à jour du profil en fonction du retour de pertinence fourni par l'utilisateur sur les documents qu'il a reçus, la mise à jour se fait par intégration des thèmes abordés dans les documents jugés pertinents [4].

Par exemple, avec cette approche, un système de recommandation de films essaie de détecter le maximum d'attributs en communs entre deux films, le premier bien apprécié par l'utilisateur actif dans le passé et le nouveau film (le même genre, même directeur, même acteur…) [2].

Le contenu du profil dépend de la méthode utilisée dans l'analyse du contenu des documents. Pour ce faire, plusieurs techniques d'apprentissage en ligne ont été utilisées, des techniques d'analyse textuelle empruntées à la recherche d'informations pour la recommandation de documents textuels (sites web, articles…). Le profil est souvent sous forme d'un vecteur de mots-clés avec des poids. Le poids associé à chaque mot reflète l'importance de ce terme pour l'utilisateur. Ces mots sont souvent extraits à l'aide de la mesure TF-IDF ( Term Frequency - lnverse Document Frequency). Ce vecteur est ensuite comparé à celui du document, pour ce faire plusieurs mesures peuvent être utilisées telles que la mesure des vecteurs cosinus [5].

$$\text{Profile_contenu (c)} = \{(t_j^c, w_j^c)\},\ j=1 \ldots k$$

$$\text{Contenu (s)} = \{(t_i^s, w_i^s)\},\ i=1 \ldots n$$

$$u(c, s) = \cos(\overrightarrow{w_c}, \overrightarrow{w_s}) \models \sum_{i=1,k} \frac{w_{i,c}}{\sqrt{\sum_{i=1,k} w_{i,c}^2}} \times \frac{w_{i,s}}{\sqrt{\sum_{i=1,k} w_{i,s}^2}}$$

Avec :

$w_{i,c}$ : Le poids du terme i dans le vecteur du profil utilisateur.

$w_{i,s}$ : Le poids du terme i dans le vecteur du document s,

(1.1)

### 1.2.1.2 Limitations des approches basées sur le contenu

Le filtrage des documents basé sur le contenu ne permet pas d'intégrer d'autres facteurs de pertinence que le facteur thématique. Pourtant il existe de nombreux autres facteurs de

pertinence comme par exemple l'adéquation entre le public visé par l'auteur et l'utilisateur, ou encore la qualité scientifique et le degré de précision des faits présentés, la fiabilité de la source d'information, ... [2].

- L'effet dit " entonnoir " restreint le champ de vision des utilisateurs. En effet, le profil évolue toujours dans le sens d'une expression du besoin de plus en plus spécifique, qui ne laisse pas de place à des documents pourtant proches mais dont la description thématique diffère fortement. Par exemple, lorsqu'un nouvel axe de recherche surgit dans un domaine, avec de nouveaux termes pour décrire les nouveaux concepts, ces termes n'apparaissent pas dans le profil, ce qui élimine automatiquement les documents par filtrage, l'utilisateur n'aura donc jamais l'occasion d'exprimer un retour de pertinence positif envers ce nouvel axe de recherche, à moins d'en avoir connaissance par ailleurs et de modifier son profil manuellement en ajoutant les termes pertinents [4].
- Tous les contenus ne peuvent pas être représentés avec des mots-clés (exemple : les images,…).
- Des items représentés par le même ensemble de mots-clés ne peuvent pas être distingués.
- Les utilisateurs avec des milliers d'achats / items posent un problème pour construire un profil.
- "over- specialization " : limitation aux items similaires.
- Pour produire des recommandations précises, l'utilisateur doit fournir un " feedback "sur les suggestions - les utilisateurs.
- Entièrement basé sur les scores d'articles et de sujets d'intérêt : moins il y a de scores, plus l'ensemble de recommandations possibles est limité.
- Il faut faire attention à l'évolution des intérêts de l'utilisateur et les prendre en compte [6].
- La difficulté d'indexation des documents multimédias : la croissance des documents multimédias (texte, image, vidéos, ...) pose le problème de la prise en compte de l'information structurelle des documents pour aider à identifier les contenus multimédias pertinents.
- L'effet de masse : l'utilisateur ne bénéficie pas des jugements que d'autres utilisateurs peuvent faire sur les documents qu'il reçoit, l'utilisateur doit procéder lui-même à l'analyse des documents reçus, analyse qui fait intervenir d'autres critères que celui de la thématique [7].
- Problème de démarrage à froid : Un nouvel utilisateur du système éprouve des difficultés à exprimer son profil en spécifiant des thèmes qui l'intéressent. Ceci malgré les techniques d'apprentissage ou l'utilisateur fournit des textes exemples [8].

## 1.2.2 Approches collaboratives

Les approches basées sur le contenu présentant des inconvénients, la recherche a concentré ses efforts sur l'élaboration d'une nouvelle approche dite collaborative. Il existe deux types de systèmes de recommandation collaboratif : le filtrage collaboratif basé sur les utilisateurs (U2U : User to User) qui exploite les relations entre utilisateurs et le filtrage collaboratif basé sur les items (I2I : Item to Item) qui se base sur les relations entre items. Ces deux types d'approches utilisent les notes (implicites ou explicites) des utilisateurs pour les items comme seule connaissance [9].

### 1.2.2.1 Filtrage collaboratif basé sur les utilisateurs (U2U)

Cette approche tend à assimiler un utilisateur à un ensemble d'utilisateurs ayant les mêmes intérêts et à s'inspirer des objets appréciés par ceux-ci afin d'établir les recommandations [10, 11, 12, 13, 14]. Le principe est que les utilisateurs ayant partagé les mêmes préférences dans le passé sont susceptibles de partager leurs préférences dans le futur.

Le coefficient de corrélation de Pearson [11] est souvent utilisé afin de calculer la similarité entre deux utilisateurs u1 et u2.

$$pearson(u_1, u_2) = \frac{\sum_{i \in \mathcal{I}} (note(u_1, i) - \overline{note(u_1)})\ (note(u_2, i) - \overline{note(u_2)})}{\sqrt{\sum_{i \in \mathcal{I}} (note(u_1, i) - \overline{note(u_1)})^2}\ \sqrt{\sum_{i \in \mathcal{I}} (note(u_2, i) - \overline{note(u_2)})^2}} \quad (1.2)$$

où $note(u, i)$ est la note de l'utilisateur $u$ pour l'item $i$
$\overline{note(u)}$ est la note moyenne de l'utilisateur $u$.

Une fois la similarité entre l'utilisateur cible et les autres utilisateurs calculée, les notes des items non consultées par l'utilisateur cible sont déduites à partir des notes des utilisateurs les plus similaires à cet utilisateur, en effectuant leur moyenne par exemple.

Prenons l'exemple de matrice de notes figurant sur la table 1, les quatre premiers utilisateurs ont évalué les items 1 à 5. Donna n'a évalué que les 4 premiers items. L'idée du filtrage collaboratif est de déterminer qui de Rose, Martha, River ou Jack, a les goûts les plus proches de Donna afin de pouvoir estimer la note qu'elle attribuerait à l'item 5.

Le calcul du coefficient de Pearson entre Donna et les autres utilisateurs aboutit aux résultats suivants :

| | $Item_1$ | $Item_2$ | $Item_3$ | $Item_4$ | $Item_5$ |
|---|---|---|---|---|---|
| Rose | 1 | 5 | 4 | 3 | 1 |
| Martha | 2 | 4 | 2 | 5 | 2 |
| River | 4 | 4 | 1 | 2 | 4 |
| Jack | 3 | 5 | 5 | 1 | 3 |
| Donna | 5 | 4 | 2 | 2 | ? |

***Tableau 1.1- Exemple de matrice de notes***

$$\begin{aligned} pearson(Donna, Rose) &= -0.40 \\ pearson(Donna, Martha) &= -0.33 \\ pearson(Donna, River) &= 0.91 \\ pearson(Donna, Jack) &= 0.17 \end{aligned}$$

Nous pouvons voir que les notes de River sont les plus corrélées avec celles de Donna. De ce fait, étant donné que River a apprécié l'item5, il est probable qu'il soit apprécié par Donna également. Par conséquent, l'item5 peut être recommandé à Donna [9].

### 1.2.2.2 Filtrage collaboratif basé sur les items (I2I)

Lorsque le nombre d'utilisateurs dépasse le nombre d'items de façon importante, des problèmes de passage à l'échelle se posent. C'est pourquoi, le filtrage collaboratif basé sur les items a été proposé [14, 15, 16, 17, 18]. Le principe est d'exploiter la similarité entre les items (et non plus entre les utilisateurs) afin de générer des recommandations.

A l'inverse de l'approche basée sur le contenu, les similarités entre items sont calculées grâce au contenu collaboratif, c'est-à-dire aux notes de ces items, plutôt qu'à leur description.

Les similarités entre items peuvent être calculées grâce à une mesure de similarités cosinus adaptée :

$$sim(i_1, i_2) = \frac{\sum_{u \in \mathcal{U}} (note(u, i_1) - \overline{note(u)})\,(note(u, i_2) - \overline{note(u)})}{\sqrt{\sum_{u \in \mathcal{U}} (note(u, i_1) - \overline{note(u)})^2}\ \sqrt{\sum_{u \in \mathcal{U}} (note(u, i_2) - \overline{note(u)})^2}} \quad (1.3)$$

où $note(u, i)$ est la note de l'utilisateur $u$ pour l'item $i$
$\overline{note(u)}$ est la note moyenne de l'utilisateur $u$.

Reprenons l'exemple de matrice de notes figurant sur la table 1. Afin de déterminer la note de l'item 5 attribuée par Donna, on calcule la similarité en termes de notes entre tous items :

$$\begin{aligned} sim(i_5, i_1) &= 1 \\ sim(i_5, i_2) &= -0.64 \\ sim(i_5, i_3) &= -0.54 \\ sim(i_5, i_4) &= -0.31 \end{aligned}$$

Il est clair que les notes des items i1 et i5 sont très similaires (identiques dans cet exemple), la note de Donna pour l'item i1 sera donc utilisée pour déterminer sa note pour l'item i5. L'item lui sera donc recommandé [9].

#### 1.2.2.3 Limitations du filtrage collaboratif

Le filtrage collaboratif a également des limitations. Les utilisateurs étant généralement réticents à l'idée de noter les items, la matrice de notes, associant les notes des utilisateurs aux items, est très éparse. Le phénomène est appelé la matrice creuse. On peut donc se retrouver dans une situation où un item est très peu évalué. Ou encore, qu'il ne soit évalué que par des personnes aux centres d'intérêt différents. En conséquence, il sera difficile de recommander un tel item, car il ne fera pas partie des éléments appréciés par un certain groupe d'utilisateurs [19].

De plus, le filtrage collaboratif souffre du démarrage à froid du point de vue utilisateur, mais également du point de vue item. Lorsqu'un nouvel item est introduit dans la base, aucune information n'existe quant à l'appréciation des utilisateurs à son égard. Par conséquent, un tel article ne pourra être recommandé. Pour pallier ce problème, les méthodes hybrides ont été proposées [9].

### 1.2.3 Approches démographiques

C'est une recommandation simple qui propose des items par rapport au profil démographique d'usager (Figure 1.1). Elle comporte à partager les usagers en plusieurs classes ou groupes par rapport aux informations démographiques telles que le sexe, l'âge, la profession, la localisation, la langue, le pays, ...

Le principe de cette approche est que deux utilisateurs ayant évolués dans un environnement similaire partagent des gouts communs que deux utilisateurs ayant évolués dans des environnements différents et ne partageant donc pas les mêmes codes [21].

De nombreux sites utilisent cette solution simple qui propose une offre de contenu "personnalisé". Par exemple, les utilisateurs sont redirigés vers un site Web particulier en fonction de leur langue ou pays.

Ces approches ont été très populaires dans la littérature de marketing, mais ont reçu peu d'attention dans le domaine des algorithmes de recommandation [20].

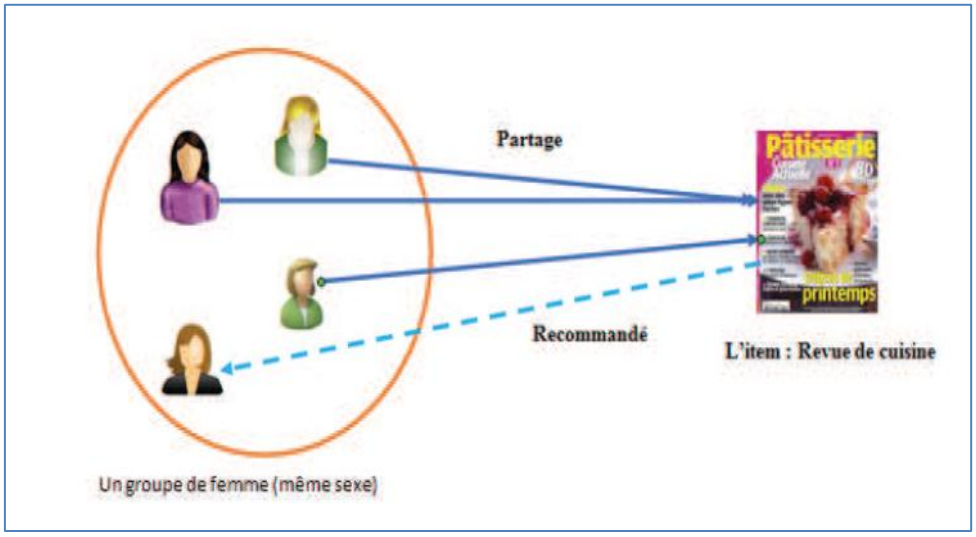



***Figure 1.1- Recommandation démographique***

## 1.2.4 Approches sociales

La plupart des réseaux sociaux (Facebook, twitter, …) s'appuient sur la recommandation sociale (Figure 1.2).

L'idée de base est donc de dire que si des utilisateurs liés dans le réseau social et qui ont partagé des mêmes intérêts dans le passé, il y a de fortes chances qu'ils partagent aussi les mêmes gouts dans le futur.

Le système propose des recommandations à partir des relations de l'usager avec ces amis dans le réseau social, et parfois cette recommandation dépend aussi de la valeur de confiance d'usager dans chacun de ses amis, l'exemple le plus connu de cette recommandation est la

section des pages et des groupes qui apparait dans la partie droite d'une page Facebook. L'importance décisionnelle du bouton ≪ I Like≫ de Facebook a donné un succès croissant dont 55% des utilisateurs sont influences par leurs amis [20].




***Figure 1.2- Recommandation sociale***

## 1.2.5 Approches basées sur la connaissance

Les recommandations sont générées en utilisant des connaissances spécifiques dont certaines caractéristiques d'items répondent aux préférences de l'usager (Figure 1.3).

Généralement les systèmes à base de connaissances sont plus fiables par rapport à d'autres types de recommandation si les données limitées sont disponibles, à savoir, si le système ne peut pas compter sur l'existence d'un historique de l'utilisateur [20].

### a- Raisonnement à base des cas

Le raisonnement à base des cas tire parti de la régularité du monde réel afin de résoudre des problèmes en recherchant la solution d'un cas semblable rencontré et résolu dans le passé [22].

**Piamrat, Viho, Bonnin, et Ksentini** [22] ont utilisé cette approche dans les systèmes de recommandation, ils estiment combien les besoins ou les préférences (description de problème) de l'usager correspondent aux recommandations possibles (solutions du problème) en se basant sur le comportement de consommation précédente (cas précédents).

**b- Raisonnement à base de contraintes**

La recommandation à base de contraintes exploite des bases de connaissances prédéfinies qui contiennent des règles explicites sur la façon de relier les exigences des clients avec des fonctionnalités d'item [20].

Par exemple, un utilisateur peut être intéressé à acheter des produits avec un certain ensemble de caractéristiques et dans une gamme de prix spécifique [20].

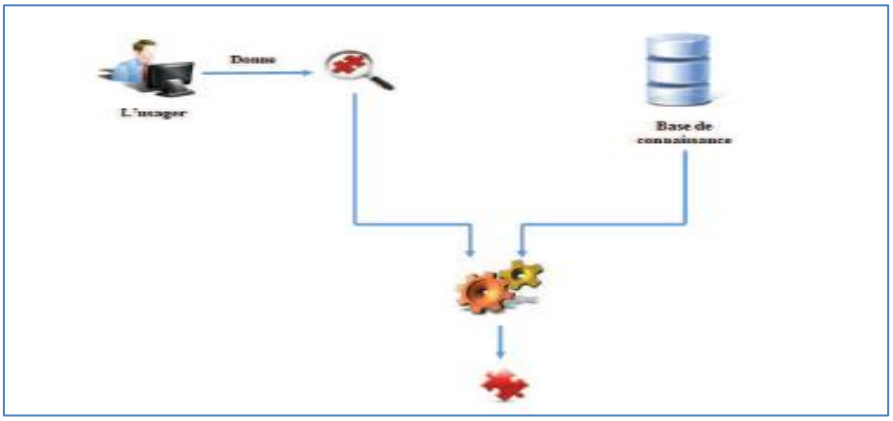



***Figure 1.3- Recommandation à base de connaissance***

## 1.2.6 Approches hybrides

Les approches hybrides sont des approches qui combinent deux ou plusieurs approches de recommandation. Ces combinaisons permettent de bénéficier des avantages des approches utilisées, de pallier leurs inconvénients et de proposer des recommandations plus pertinentes. Par exemple, le démarrage à froid au niveau des items des approches collaboratives, peut bénéficier des avantages des approches basées sur le contenu. En effet, dans une approche basée sur le contenu, seule la description de l'item est utilisée dans la recommandation, il n'y a donc pas besoin que l'item soit noté par un certain nombre d'utilisateurs avant de pouvoir être recommandé, comme c'est le cas dans les approches collaboratives [20].

La meilleure description des méthodes hybrides a été faite par **Burke** [23]. Alors, selon **Burke** on peut distinguer sept façons de combiner les méthodes traditionnelles :

#### 1.2.6.1 Hybridation pondérée

Une méthode hybride qui combine la sortie d'approches distinctes, utilisant par exemple, une combinaison linéaire des scores de chaque technique de recommandation [20].

#### 1.2.6.2 Hybridation à Commutation

C'est une technique qui permet de faire le choix d'un modèle de recommandation parmi plusieurs, en se basant sur plusieurs critères. La détermination de la technique appropriée dépend de la situation. Le système se doit alors de définir les critères de commutation, ou les cas ou l'utilisation d'une autre technique est recommandée. Ceci permet au système de connaitre les points forts et les points faibles des techniques de recommandation qui le constituent [20].

#### 1.2.6.3 Hybridation mixée

Dans cette approche, le recommandeur ne combine pas, mais augmente la description des ensembles de données, en prenant en considération les estimations des utilisateurs et la description des items. La nouvelle fonction de prédiction doit faire face aux deux types de descriptions et permet d'éviter les problèmes posés par le filtrage collaboratif, à savoir, le démarrage à froid [20].

#### 1.2.6.4 Hybridation par combinaison de caractéristiques

Dans un hybride basé sur la combinaison de caractéristiques, les données provenant de techniques collaboratives sont traitées comme une caractéristique, et une approche basée sur le contenu est utilisé sur ces données [20].

#### 1.2.6.5 Hybridation en cascade

La cascade implique un processus étape par étape. Dans ce cas, une technique de recommandation est appliquée en premier, produisant un ensemble de candidats potentiels. Puis, une deuxième technique raffine les résultats obtenus dans la première étape. Cette méthode a pour avantage que si la première technique génère peu de recommandations, ou si ces recommandations sont ordonnées afin de permettre une sélection rapide, la deuxième technique ne sera plus utilisée [20].

#### 1.2.6.6 Hybridation par Augmentation de caractéristiques

L'augmentation de caractéristiques est semblable à la cascade, mais dans ce cas-là les résultats obtenus (le classement ou la classification) de la première technique sont utilisées par le deuxième comme une caractéristique ajoutée [20].

#### 1.2.6.7 Hybridation méta-niveau

Dans un hybride basé sur méta niveau, une première technique est utilisée, mais différemment que la précédente méthode (augmentation de caractéristiques), non pas pour produire de nouvelles caractéristiques, mais pour produire un modèle. Et dans la deuxième étape, c'est le modèle entier qui servira d'entrée pour la deuxième technique [24].

Après avoir cité les différentes approches de recommandation passant aux caractéristiques du modèle de l'utilisateur qui présente l'élément clé autour duquel se fait la recommandation.

## 1.3 Caractéristiques du modèle utilisateur

Un certain nombre de caractéristiques doivent être analysées lors de la définition d'un modèle utilisateur et aident à déterminer sa conception. Selon [25], les modèles utilisateurs peuvent être classés selon plusieurs facteurs :

- **Adaptatif Vs. Adaptable** : Un modèle est dit adaptatif lorsque celui-ci est mis à jour automatiquement par le système. A l'inverse, il est dit adaptable si l'utilisateur intervient directement dans la construction du modèle.

- **Modélisation implicite Vs. explicite** : Dans une collecte explicite d'informations, l'utilisateur est interrogé sur des caractéristiques statiques (âge, localisation, centres d'intérêt, ...). Certaines applications demandent également de noter les items (plateformes de films, e-commerce ...). Les utilisateurs étant souvent réticents à la communication, une découverte implicite de leurs centres d'intérêt s'avère souvent nécessaire.

  Une collecte implicite consiste à analyser le comportement de l'utilisateur. Elle repose sur l'utilisation des techniques de Data et WEB mining. De ce fait, les données d'usage de

l'utilisateur sont analysées et ses préférences sont déduites en fonction de ses recherches, des items consultés, de la durée des consultations d'items ...

Le plus souvent, les deux méthodes sont combinées pour obtenir une meilleure modélisation de l'utilisateur.

- **Modélisation dynamique Vs. statique** : La modélisation dynamique consiste à modifier continuellement le profil de l'utilisateur selon les informations acquises au fil du temps.

La modélisation statique, quant à elle, interdit les modifications du profil après qu'il ait été établi.

- **Individu Vs. groupe d'individus** : Les modèles de groupes représentent des caractéristiques basées sur des stéréotypes, tandis que les modèles individuels représentent les caractéristiques de chaque utilisateur. La classification d'un individu dans un groupe s'effectué le plus souvent à l'aide d'algorithmes de fouille de données, de classification ou de clustering.

- **Modélisation interne Vs. externe** : Cette modélisation concerne le niveau d'intégration du modèle dans l'application. En effet, le modèle est interne lorsqu'il fait partie intégrante de l'application. Inversement, si le modèle est une entité dissociée de l'application, le modèle est dit externe.

- **Modélisation opaque Vs. lisible** : Un modèle est opaque si l'utilisateur n'a pas le droit de le consulter. Autrement, il est dit lisible.

- **Modèle à court terme Vs. à long terme** : Les modèles à court terme sont construits et utilisés sur une courte période (une session). Quant aux modèles à long terme, ils sont construits au fur et à mesure des sessions et évoluent selon l'évolution de l'utilisateur.

- **Granularité** : Il concerne le niveau de détail utilisé pour la représentation du modèle.

Tous ces facteurs doivent être pris en considération avant la mise en place du modèle [2].

## 1.4 Modélisation du profil utilisateur

La modélisation de l'utilisateur consiste à étudier le processus d'extraction et d'analyse des interactions entre l'utilisateur et le système d'information (SI) pour construire des profils utilisateurs. Les profils utilisateurs construits par ce processus peuvent être perçus comme des données, des informations ou des connaissances en fonction des usages qui en sont faits. Les différentes étapes de processus de modélisation sont structurées en 4 phases : la sélection des données (selection), le prétraitement des données (preprocessing), la transformation des données (transformation), et la fouille de données (data mining) [26].

### 1.4.1 Sélection de données pour les profils utilisateurs

Dans le processus de développement des profils utilisateurs, la phase de sélection de données fait intervenir deux concepts : les producteurs de données (individus qui produisent les données à utiliser pour construire le profil de l'utilisateur) et les sources de données (provenance et accessibilité des données) [26].

#### 1.4.1.1 Producteurs de données

Les données sélectionnées pour construire les profils utilisateurs peuvent être produites soit [26] :

- ***par l'utilisateur lui-même*** : c'est le cas le plus courant.
- ***par les individus qui lui sont similaires.***
- ***par les individus de son réseau social***.

#### 1.4.1.2 Sources de données

Par sources de données, nous entendons les systèmes (systèmes d'exploitation, logs de bases de données, logs de serveurs Web, ...) et les applications (social bookmarking[1], ...) par lesquels les données utilisateurs peuvent être collectées [26].

[1]Au lieu de mettre simplement les pages qu'on apprécie ou qu'on juge comme des références dans les favoris de son navigateur, on les inscrit sur un site de social bookmarking pour les partager avec les autres utilisateurs.

## 1.4.2 Prétraitement de données

Une fois les sources de données définies à l'étape de sélection des données, l'étape du prétraitement traite les contenus de ces sources de données afin qu'ils soient conformes aux modèles ou aux entrées des algorithmes des étapes ultérieures du processus [26].

## 1.4.3 Structuration des données pour les profils utilisateurs

Plusieurs modèles de données sont proposés dans ce sens, en fonction des objectifs de chaque auteur [27, 28,29]. Une synthèse des travaux de la littérature qui visent à structurer les différents types de données à prendre en compte lors de la construction des profils utilisateurs permet de distinguer cinq grandes catégories de données : les données explicites, les données implicites, les données de sécurité, les données de contexte et les données de sémantique.

### 1.4.3.1 Données explicites (feedback explicites)

Les données explicites sont les données fournies de manière explicite par l'utilisateur (via des formulaires par exemple).

### 1.4.3.2 Données implicites (feedback implicites)

Les données implicites sont les données collectées en observant le comportement des utilisateurs ou en scrutant leurs activités (et qui sont généralement utilisées pour déterminer leurs centres d'intérêts).

### 1.4.3.3 Données de contexte

Au-delà des données implicites issues du comportement de l'utilisateur, les profils utilisateurs doivent prendre en compte les données de contexte qui influencent très souvent ce comportement. **Abbar, Bouzeghoub, Kostadinov et Lopes** [30] définissent un modèle multidimensionnel de contexte utilisateur intégrant plusieurs dimensions plus ou moins présentes dans les travaux de la littérature :

- ***Le contexte temporel*** : qui est lié au paramètre temps. Ce contexte permet par exemple de définir des profils à court à terme ou à long terme de l'utilisateur [31,32].

- ***Le contexte spatial*** : qui est lié au paramètre lieu. Le profil d'un utilisateur au travail n'est pas par exemple le même que son profil à la maison [33].

- ***Le contexte matériel*** : qui est lié au dispositif matériel utilisé par l'utilisateur (taille de l'écran, vitesse du processeur, ...) et qui pourra servir à définir des contraintes spécifiques de présentation de documents multimédias par exemple [34,35].

- ***Le contexte environnement*** : qui est lié aux caractéristiques ambiantes de l'environnement d'interaction de l'utilisateur (température par exemple).

- ***Le contexte émotionnel*** : qui est lié à l'état psychique de l'utilisateur à un moment donné.

#### 1.4.3.4 Données sémantiques

Dans le cas de la construction des profils utilisateurs par traitement de données textuelles, les données sémantiques permettent d'enrichir la sémantique des profils construits. Ceci est nécessaire pour lever d'éventuelles ambigüités liées à la polysémie de termes. Par exemple, « Java » peut désigner un langage de programmation en informatique ou île indonésienne en géographie [26].

#### 1.4.3.5 Données de sécurité

Il s'agit des paramètres de sécurité que l'utilisateur définit pour limiter ou autoriser le traitement de ses données afin de construire son profil.

### 1.4.4 Fouille de données pour les profils utilisateurs

Une fois les données structurées, elles sont utilisées par les algorithmes de fouille de données pour construire le profil de l'utilisateur. Plusieurs techniques de fouille de données peuvent être utilisées. **Gao, Liu et Wu** [37] classifient les techniques utilisées suivant les modèles de profils utilisateurs à construire (modélisation du comportement, modélisation des centres d'intérêts, modélisation des intentions).

#### 1.4.4.1 Modélisation comportementale (behaviour modeling)

Ce type de modélisation consiste à analyser les comportements des internautes via les historiques de navigation ou les transactions qu'ils effectuent sur des serveurs Web [38], dans le but de déterminer des parcours de navigation récurrents, de valider des stratégies marketing ou de vérifier la pertinence des campagnes marketing.

#### 1.4.4.2 Modélisation des centres d'intérêts (interest modeling)

Les centres d'intérêts sont définis par une fonction c(i) qui donne **le degré d'intérêt** ou de désintérêt d'un utilisateur pour **un item i** en analysant son comportement antérieur [43].

#### 1.4.4.3 Modélisation des intentions (intention modeling)

La modélisation des intentions des utilisateurs consiste donc à construire un modèle qui permettra d'identifier le but de chaque utilisateur du système d'information. Par exemple, des clients d'un site de e-commerce peuvent être divisés en deux groupes : ceux qui ont réellement pour but d'acheter et ceux qui n'ont pas pour but d'acheter.

## 1.5 Evaluation et de validation des performances du système

Ils existent plusieurs techniques pour évaluer un système de recommandation nous citons parmi eux : le test en ligne et le test hors ligne.

### 1.5.1 Evaluation en ligne

Ce test permet d'évaluer l'application sur de vrais utilisateurs. Dans ce cas de figure, les utilisateurs sont orientés vers plusieurs alternatives du système de recommandation afin de pouvoir comparer leurs performances. L'intérêt présenté par ce genre d'évaluation est l'obtention de résultats réels, dans la mesure où les utilisateurs sont les utilisateurs finaux qui ne sont pas mis au courant de l'exécution du test [9]. L'évaluation en ligne peut se faire également en situation virtuelle en faisant appel à un groupe d'utilisateurs pour tester le système et récupérer leurs impressions [57].

Le test en ligne peut être très coûteux et parfois même impossible, une alternative plus simple, moins coûteuse et plus répandue dans l'évaluation des systèmes de recommandation est l'évaluation offline [9].

## 1.5.2 Evaluation offline

Le test hors ligne est une simulation qui requiert la disposition d'un jeu de données pouvant servir à reproduire le comportement des utilisateurs. Ce jeu de données contiendra par exemple les notes attribuées par des utilisateurs pour certains items. L'hypothèse de départ implique que le comportement de l'utilisateur lors de la collecte initiale des données sera similaire à son comportement lors de l'exécution de l'évaluation du système de recommandation. Afin de procéder à une simulation, le jeu de données doit être échantillonné en deux ensembles : un ensemble d'apprentissage et un ensemble de test selon la proportion 80/20, c'est-à-dire que 80% du jeu de données est utilisé pour l'apprentissage et 20% pour le test [9]. Pour s'assurer que le résultat de l'évaluation est indépendant du partage apprentissage-test, il est possible de répéter le processus n fois, le résultat final est la moyenne des n résultats, ce mode est appelé ***validation croisée (cross validation)***. Pour un rapport 80/20 le processus est répété 5 fois, ainsi, le jeu initial est trié aléatoirement et divisé en 5 parties représentant chacune 20% du jeu initial, à chaque itération on choisit une partie comme jeu de test et le reste est utilisé pour le jeu apprentissage [57].

L'échantillon d'apprentissage sert à établir le modèle alors que l'échantillon de test sert à valider. La validation consiste à appliquer le système de recommandation sur l'échantillon d'apprentissage et à comparer les résultats obtenus avec l'échantillon de test. Ce type d'évaluation ne pose pas de problème de fuite d'utilisateurs. Elle permet d'intégrer facilement une grande masse d'utilisateurs. Elle n'est pas très sensible aux potentiels changements dans le comportement de l'utilisateur [9].

## 1.5.3 Précision

La mesure de la précision d'un système de recommandation reste le critère dominant lors de la comparaison des performances de plusieurs systèmes de recommandation [57].

La précision mesure la capacité d'un système de recommandation à prédire des recommandations pertinentes pour ses utilisateurs. Il peut s'agir de mesurer le potentiel du système à prédire *la valeur d'un vote*, ou à prédire *une liste de recommandations pertinentes*.

Plusieurs mesures ont été utilisées dans la littérature pour évaluer la précision [57, 66, 67,68, 69].

### 1.5.3.1 Précision de la prédiction des votes

La précision de la prédiction des votes ne peut être calculée que pour les algorithmes de recommandation qui prédisent la valeur du vote d'un item. On note par R l'ensemble des votes $r_{ui}$ de tous les utilisateurs u de U sur les items de i de I. Pour évaluer un système de recommandation, on divise l'ensemble R en deux sous-ensembles $R_{apprentissage}$ et $R_{test}$ distincts, le premier est utilisé pour l'apprentissage de la fonction de prédiction pred(u, i), le second pour évaluer la précision de la prédiction. On définit par T l'ensemble des couples (u, i) de $R_{test}$ pour lesquels le système de recommandation a prédit la valeur du vote.

a. **Erreur Absolue Moyenne**

(Mean Absolute Error MAE) mesure la moyenne des écarts entre un vote prédit et son correspondant dans $R_{test}$. L'erreur absolue moyenne (Formule 1.4) a été utilisée pour évaluer plusieurs systèmes de recommandation [57, 67, 70,71] et est encore populaire de nos jours.

$$MAE = \frac{\sum_{(u,i)\in T} |pred(u,i) - v_{ui}|}{|T|} \quad (1,4)$$

b. **Racine de l'Erreur Quadratique Moyenne**

Root Mean Squared Error (RMSE) (Formule 1.5) est devenue la mesure de référence surtout après son utilisation pour mesurer la précision dans le challenge Netflix [57,71] pour la recommandation de films. Comparée à la MAE, la RMSE pénalise les écarts importants, ainsi un écart de 2 points aura pour effet d'augmenter la somme des erreurs par 2 au niveau de la MAE, alors qu'il augmentera la somme par 4 au niveau de la RMSE.

$$RMSE = \sqrt{\frac{\sum_{(u,i)\in T}(pred(u,i) - v_{ui})^2}{|T|}} \quad (1.5)$$

Plus la MAE ou la RMSE est faible, meilleure est la précision du système. En plus de ces deux mesures, on trouve la MAE normalisée (NMAE) et la RMSE normalisée (NRMSE) deux versions normalisées par l'intervalle des valeurs des votes ($v_{max} - v_{min}$) permettant ainsi de comparer des systèmes s'exécutant sur des jeux de données différents.

### 1.5.3.2 Précision d'une Top-N liste

Le principe consiste plutôt à recommander à l'utilisateur courant une liste $L(u_a)$ contenant N items susceptibles de l'intéresser. On parle alors de recommandation Top-N. La mesure de la précision revient alors à mesurer la capacité du système à décider si un item est pertinent ou pas pour l'utilisateur courant. Si les votes sont numériques, il est primordial de définir un seuil de vote permettant de classer les items de la base de test $R_{test}$ en items pertinents et items non pertinents. Ainsi, pour des votes définis sur une échelle de 1 à 5, on peut considérer que les items ayant un vote $\geq 4$ sont pertinents pour l'utilisateur u et ceux ayant un vote $< 4$ sont non pertinents. La base $R_{apprentissage}$ est utilisée pour l'apprentissage de L(u).

- **La précision, le rappel**

La précision et le rappel sont deux mesures très populaires pour l'évaluation des systèmes de recherche d'information depuis leur proposition par Cleverdon en 1968 [57,73]. Ils sont également largement utilisés pour l'évaluation des systèmes de recommandation proposant une Top-N liste [57,58]. On définit par $Iu_{test}$ le sous ensemble d'items tests de Iu présents dans la base de test $R_{test}$. On note par $P(u) \subset Iu_{test}$ la liste des items de la base de test pertinents

pour l'utilisateur u. Le rappel (Resp. la précision) est calculé(e) pour chaque utilisateur u puis agrégé(e) sur tous les utilisateurs de U tel(le) que le définit la Formule 1.6 (Resp. la Formule 1.7).

$$Rappel = \frac{1}{|U|} \sum_{u \in U} \frac{|L(u) \cap P(u)|}{|P(u)|} \quad (1.6)$$

$$Précision = \frac{1}{|U|} \sum_{u \in U} \frac{|L(u) \cap P(u)|}{|L(u)|} \quad (1.7)$$

**La précision** mesure la proportion de recommandations réellement pertinentes pour l'utilisateur u parmi les items recommandés (dans L(u)).

**La précision @X** : précision calculée des top X items les plus élevés dans le classement.

**Le rappel** mesure le taux de recommandations pertinentes parmi l'ensemble des items réellement pertinents pour u (dans P(u)). Plus le taux de précision et de rappel sont élevés, meilleure est la performance de l'algorithme de recommandation.

# 1.6 Conclusion

Après avoir détaillé les différents concepts sur les systèmes de recommandation c'est-à-dire les approches de recommandation, les profils utilisateurs et les méthodes d'évaluation et de validation d'un système nous allons passer au deuxième chapitre qui concerne la recommandation sociale et qui représente le thème de ce travail.

# 2 Recommandation sociale

## 2.1 Introduction

La Recommandation sociale est une recommandation collaborative qui prend en considération les relations sociales : confiance sociale, amitié, groupe d'amis, …

Dans ce chapitre nous allons parler des médias sociaux et de leurs émergences, nous allons aussi présenter les réseaux sociaux et leurs méthodes d'analyse et enfin nous détaillerons la recommandation sociale et les différentes étapes pour construire un système de recommandation sociale.

## 2.2 Médias sociaux

L'expression « **médias sociaux** » est de plus en plus utilisée et tend à remplacer le terme de Web 2.0 et recouvre les différentes activités qui intègrent la technologie, l'interaction sociale et la création de contenu. Les médias sociaux utilisent l'intelligence collective dans un esprit de collaboration en ligne.[1]

Par le biais de ces moyens de communication sociale, des individus ou des groupes d'individus forment un réseau social, collaborent, créent ensemble du contenu Web, organisent le contenu, l'indexent, le modifient ou font des commentaires, le combinent avec des créations personnelles. Les deux termes de Web 2.0 et de médias sociaux demeurent assez proches et concernent une grande variété de sites différents : les blogs, les wikis, et les réseaux sociaux numériques de tout type.[1]

Les médias sociaux utilisent beaucoup de techniques, telles que les flux RSS et autres flux de syndication Web, les blogues, les wikis, le partage de photos (Flickr), le vidéo-partage (YouTube), les podcasts, les réseaux sociaux (Facebook), le bookmarking collaboratif, les mashups, les mondes virtuels ou les microblogues (Twitter).[1]

[1] https://fr.wikipedia.org /wiki/Médias_sociaux.

Parmi les nombreux types de sites sociaux nous pouvons distinguer :

- **Blogs**

Un blog, anglicisme pouvant être francisé en blogue et parfois appelé cyber carnet ou bloc-notes, est un type de site web ou une partie d'un site web utilisé pour la publication périodique et régulière d'articles, généralement succincts, et rendant compte d'une actualité autour d'un sujet donné ou d'une profession [1].

- **Flux RSS**

Sigle venant de l'anglais « Rich Site Summary » ou encore « Really Simple Syndication ») est une famille de formats de données utilisés pour la syndication de contenu Web.

Un produit RSS est une ressource du World Wide Web dont le contenu est produit automatiquement (sauf cas exceptionnels) en fonction des mises à jour d'un site Web. Les flux RSS sont des fichiers XML qui sont souvent utilisés par les sites d'actualité et les blogs pour présenter les titres des dernières informations consultables [1].

- **Wikis**

Un wiki est une application web qui permet la création, la modification et l'illustration collaboratives de pages à l'intérieur d'un site web. Il utilise un langage de balisage et son contenu est modifiable au moyen d'un navigateur web. C'est un outil de gestion de contenu, dont la structure implicite est minimale, tandis que la structure explicite émerge en fonction des besoins des usagers [1].

- **Flickr**

Flickr, de l'anglais to flickthrough, feuilleter, est un site web de partage de photographies et de vidéos gratuit, avec certaines fonctionnalités payantes. En plus d'être un site web populaire auprès des utilisateurs pour partager leurs photos personnelles, il est aussi souvent utilisé par des photographes professionnels [1].

[1] https://fr.wikipedia.org /wiki/Médias_sociaux.

- **YouTube**

YouTube est un site web d'hébergement de vidéos sur lequel les utilisateurs peuvent envoyer, évaluer, regarder, commenter et partager des vidéos [1].

- **Forums**

Sites qui facilitent la discussion au sein des communautés en ligne [1].

- **Réseau Social**

Site internet qui permet aux internautes de se créer une page personnelle afin de partager et d'échanger des informations, de photos ou des vidéos avec leur communauté d'amies et leur réseau de connaissance [1].

- **Sites d'indexation (bookmarking)**

Le social bookmarking en français « marque-page social », « navigation sociale » ou bien « partage de signets ») est un site pour les internautes de stocker, de classer, de chercher et de partager leurs liens favoris.

Dans un système de partage de signets, les utilisateurs enregistrent des liens vers les pages web qu'ils trouvent intéressantes. Ces listes de liens sont ensuite accessibles au public ou à un réseau particulier. Ainsi des gens ayant les mêmes centres d'intérêt peuvent les consulter, que ce soit par catégorie, par nom ou même de façon aléatoire. Certains s'appuient sur un répertoire permettant ainsi une certaine confidentialité [1].

[1] https://fr.wikipedia.org /wiki/Médias_sociaux.

## 2.3 Analyse des réseaux sociaux

L'analyse de réseaux se rapporte aux théories relationnelles qui permettent de formaliser les interactions sociales en termes de nœuds et de liens (notions issues de la Théorie des graphes)[1].

Les nœuds sont habituellement les acteurs sociaux interagissant mais ils peuvent aussi représenter des institutions. Les liens sont les relations entre ces nœuds. Il peut exister plusieurs sortes de liens entre les nœuds. Dans sa forme la plus simple un réseau social se modélise pour former une structure analysable où des liens effectifs entre les nœuds sont étudiés[1].

### 2.3.1 Définition de graphe

D'après [49], Un graphe G, selon sa définition mathématique, est une paire d'ensembles (V, E), où V est l'ensemble des nœuds (représentant les individus ou entités du réseau) du graphe G et E l'ensemble des arêtes entre les nœuds (représentant les liens ou relations entre les individus du réseau) du graphe G.

Un graphe G = (V,E) est un ensemble de sommets $V=\{v_1,v_2, ..., v_n\}$, un ensemble d'arêtes $E=\{e_1,e_2, ... ,e_m\}$ et une fonction d'incidence qui associe deux sommets par une arête $F : E \rightarrow V \times V$, où chaque arête est notée par e=uv.

L'ordre du graphe G désigne le nombre de sommets de G et est noté par |V|.

Le nombre d'arêtes du graphe G est appelé la Taille de G.

Un sous-graphe de G est noté $G'= (V',E')$ où $V' \subseteq V$, $E' \subseteq E$.

Un nœud est l'unité de base d'un réseau. Dans un réseau social on parle d'individu. Une arête est une connexion entre deux nœuds.

Une arête est orientée si elle est utilisée dans un seul sens. On parle d'arc lorsque l'arête est orientée.

[1] https://fr.wikipedia.org/wiki/Analyse_des_reseaux_sociaux

Une arête est pondérée lorsqu'on lui attribue un poids.

Un graphe orienté désigne un graphe avec des arêtes orientées (des arcs).

Un graphe pondéré désigne un graphe avec des arêtes pondérées.

Les étiquettes ou des labels peuvent être associés aux nœuds et/ou des arêtes d'un graphe. On parlera alors d'un graphe étiqueté.

### 2.3.2 Définition de Réseau social

D'après[1], Le réseau social est un média social qui se définit comme un ensemble d'identités sociales telles que des individus ou encore des organisations reliées entre elles par des liens créés via des interactions sociales. Il se représente par une structure ou une forme dynamique d'un groupement social dans laquelle l'utilisateur se crée des « amis » ou des « contacts » selon des communautés d'intérêt qui lui sont propres. Les réseaux sociaux peuvent être de type professionnel (Ex : LinkedIn) ou d'amie (Ex : Facebook).

Un réseau social peut être représenté par un graphe G = (V,E) où l'ensemble des nœuds V représente les entités sociales (les acteurs) et l'ensemble des arcs E = (V × V) représente les relations sociales entre eux.

### 2.3.3 Mesures de centralité issues de la Recherche d'Information (RI)

A la fin des années 90, plusieurs chercheurs ont envisagé d'utiliser les liens entre documents (en plus des contenus textuels) afin d'améliorer les performances des moteurs de recherche sur le web. La pertinence d'un document par rapport à une requête est alors calculée en combinant la similarité du document par rapport à la requête avec la centralité du document. Nous présentons ci-dessous quelques-uns des algorithmes les plus connus pour le calcul de centralité dans les graphes de documents. Ces algorithmes sont tous basés sur l'idée que l'importance d'un document est relative à l'importance des documents auxquels il est connecté [50].

[1] http://blog.akova.ca/2011/08/web-2-0-medias-et-reseaux-sociaux-quelques-definitions

### 2.3.3.1 Centralite de degre (degree centrality)

La centralité de degré mesure l'importance relative d'un nœud dans un graphe en fonction du nombre de ses liens. Elle peut être définie en considérant le nombre de liens sortants (out-degree) ou entrants (in-degree) dans un graphe orienté, ou simplement le nombre de liens deg($v_i$), normalisé par le nombre maximal de liens, dans un graphe non orienté [59] :

$$Degree(v_i) = \frac{deg(v_i)}{|V| - 1} \tag{2.1}$$

### 2.3.3.2 Centralité de proximité (closeness centrality)

D'après la centralité de proximité, un nœud est considéré comme important s'il peut rapidement atteindre les autres sommets [60 ,61]. La mesure usuelle est égale à l'inverse de la somme des distances du nœud $v_i$ à tous les autres nœuds du graphe. La distance entre deux nœuds correspond à la longueur du plus court chemin (distance géodésique) les reliant :

$$CCentr(v_i) = \frac{1}{\sum_{\substack{v_j \in V \\ j \neq i}} |ShortPath(v_i, v_j)|} \tag{2.2}$$

où Short Path($v_i$ , $v_j$) représente un des chemins les plus courts entre deux sommets $v_i$ et $v_j$ et Short Path($v_i$ , $v_j$) la longueur, en nombre de liens, du chemin.

Par convention, Short Path($v_i$ , $v_j$) est égal à |V| ou à $\infty$ s'il n'existe pas un tel chemin.

### 2.3.3.3 Centralité d'intermédiarité (betweenness centrality)

D'après la centralité d'intermédiarité, un nœud est considéré comme important s'il est placé sur un nombre important de chemins (géodésiques) reliant les autres nœuds du graphe [59,60]. Cet indicateur est défini par :

$$BCentr(v_i) = \sum_{\substack{(v_k, v_j) \in V \times V \\ i \neq k \neq j}} |g_{kj}(v_i)| / |g_{kj}| \tag{2. 3}$$

où $g_{kj}(v_i)$ (resp. $|g_{kj}(v_i)|$) est l'ensemble des plus cours chemins entre les sommets $v_k$ et $v_j$ qui passent par $v_i$ (resp. son cardinal) et $g_{kj}$ (resp. $|g_{kj}|$) l'ensemble des chemins les plus courts de $v_k$ à $v_j$ (resp. son cardinal).

Ainsi, plus il y a de chemins qui passent par un nœud, plus ce dernier est important.

Ces nœuds intermédiaires jouent un rôle important dans les processus de diffusion ou de propagation puisqu'ils ont un certain contrôle sur les interactions indirectes qui peuvent exister entre les nœuds non adjacents du réseau.

#### 2.3.3.4 PageRank

Le PageRank est un indice utilisé par le moteur de recherche Google pour connaitre la popularité des pages internet qu'il indexe. Cet indice est noté entre 0 et 10, la note maximale étant une page considérée comme très populaire.

Google considère qu'une page est populaire lorsqu'un nombre important de liens pointent vers elle. Ainsi, pour schématiser grossièrement, si une page A fait un lien vers une page B, cette dernière recevra un poids potentiel de "popularité". Ce poids dépend de plusieurs facteurs dont la popularité de la page A[1].

Le PageRank simplifié d'une page pi est donné par :

$$PR_s(p_i) = \sum_{p_j \in in(p_i)} \frac{PR_s(p_j)}{d^{out}(p_j)} \quad (2.4)$$

où *in(pi)* représente l'ensemble des pages qui pointent vers la page *p(i)* et $d^{out}(pj)$ représente le degré sortant de la page *pj*.

1 http://glossaire.infowebmaster.fr/pagerank

### 2.3.3.5 HITS (Hypertext Induced Topic Search)

Dans le cadre du projet Clever [62] développé par IBM en 1998, Kleinberg[63][64] a proposé l'algorithme HITS dont l'idée est d'exploiter la structure du web afin d'améliorer la qualité de la recherche d'information. Cependant, à la différence de PageRank qui assigne à chaque page un seul degré d'importance, l'algorithme HITS caractérise chaque page par deux degrés d'importance. Ces deux degrés, que Kleinberg appelle degrés d'autorité et d'hubité, sont respectivement des mesures de centralité par rapport aux liens entrants et aux liens sortants.

L'algorithme HITS considère que le degré d'autorité d'une page est égal à la somme des degrés d'hubité des pages qui la pointent (ou la citent). En d'autres termes, une page est une bonne autorité si elle est pointée par de bons hubs. Plus précisément, le degré d'autorité d'une page $p_i$ est défini par :

$$A(p_i) = \sum_{p_j \in in(p_i)} H(p_j) \qquad (2.5)$$

où $in(p_i)$ représente l'ensemble des pages qui pointent vers la page pi et $H(p_j)$ représente le degré d'hubité de la page $p_j$.

De manière similaire, HITS considère que le degré d'hubité d'une page est égal à la somme des degrés d'autorité des pages qu'elle pointe. Cela sous-entend qu'une page est un bon hub si elle pointe vers de bonnes autorités. Ainsi, le degré d'hubité d'une page $p_i$ est défini par :

$$H(p_i) = \sum_{p_j \in out(p_i)} A(p_j) \qquad (2.6)$$

où $out(p_i)$ représente l'ensemble des pages pointées par la page $p_i$ et $A(p_j)$ représente le degré d'autorité de la page $p_j$ [50].

# 2.4 Recommandation sociale

## 2.4.1 Définition

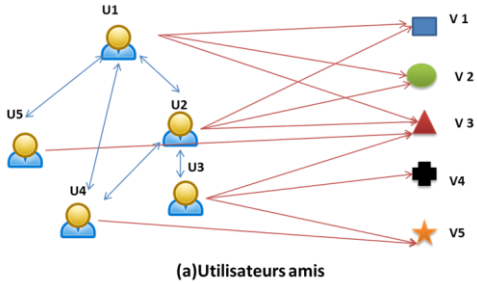
Les réseaux sociaux fournissent à la tâche de recommandation un nouveau paramètre qui est le cercle social de l'utilisateur. De ce fait, un système de recommandation social peut proposer à l'utilisateur des items qui sont populaires dans son cercle social [9].

Les systèmes de recommandation classique considèrent que les utilisateurs sont indépendants alors que les systèmes de recommandation sociale prennent en considération que l'utilisateur est connecté de plusieurs manières : amitié ou relation de confiance.

***Exemple :***

Soit T appartenant à n*n (n : nombre d'un ensemble des utilisateurs) pour noter la relation sociale utilisateur-utilisateur où T (i,j) =1 si u(i) se connecte à u(j), 0 sinon.

La figure 2.1 montre un simple exemple des utilisateurs connectés dans les media sociaux, et la figure 2.1(b) et 2.1(c) démontre leurs représentations des données pour le système de recommandation classique et le système de recommandation sociale. En plus de la matrice d'évaluation dans le système classique, les utilisateurs dans le système de recommandation sociale sont connectés en produisant des données sociales[1].




(a)Utilisateurs amis

| | V1 | V2 | V3 | V4 | V5 |
|---|---|---|---|---|---|
| U1 | 5 | ? | 2 | ? | ? |
| U2 | 4 | 4 | 5 | ? | ? |
| U3 | ? | ? | 4 | 4 | 1 |
| U4 | ? | ? | ? | ? | 3 |
| U5 | ? | ? | 1 | ? | ? |

(b) Système de recommandation classique

| | V1 | V2 | V3 | V4 | V5 |
|---|---|---|---|---|---|
| U1 | 5 | ? | 2 | ? | ? |
| U2 | 4 | 4 | 5 | ? | ? |
| U3 | ? | ? | 4 | 4 | 1 |
| U4 | ? | ? | ? | ? | 3 |
| U5 | ? | ? | 1 | ? | ? |

| | U1 | U2 | U3 | U4 | U5 |
|---|---|---|---|---|---|
| U1 | 0 | 1 | 0 | 1 | 1 |
| U2 | 1 | 0 | 1 | 1 | 0 |
| U3 | 0 | 1 | 0 | 0 | 0 |
| U4 | 1 | 1 | 0 | 0 | 0 |
| U5 | 1 | 0 | 0 | 0 | 0 |

( c) système de recommandation sociale

[1] http://www.podcastscience.fm/dossiers/2012/04/25/les-algorithmes-de-recommandation/

***Figure 2.1- Exemple d'utilisateurs connectés***

L'approche de filtrage collaboratif (CF) est généralement utilisée pour construire un système de recommandation sociale.

**Modèle CF pour recommandation sociale = modèle CF de base + modèle d'information sociale.**

Le tableau 2.1 présente une synthèse sur le filtrage collaboratif et le filtrage social :

| | Profil utilisateur | Source de données | Technique utilisée | Avantages/ inconvénient |
|---|---|---|---|---|
| Filtrage collaboratif basé sur les utilisateurs | Evaluation | Evaluation explicite ou implicite | Similarité vectorielle | + indépendant du format des documents<br>+ diversité<br>+ qualité prise en compte<br>+ Evolution dynamique du profil<br>_ Démarrage à froid<br>_ rareté des votes<br>_ complexité |
| Filtrage collaboratif basé sur les items | | | -Classification<br>- Analyse par la sémantique latente | + compense le manque des votes<br>+ rapidité du calcul<br>_ complexité de mise à jour |
| Filtrage social | Evaluation + entourage social | -Evaluation de l'utilisateur<br>- Evaluation de ses amis | score de confiance entre l'utilisateur et ses amis | + disponibilité de large communauté sociale sur le web<br>+ compense les votes manquants<br>+ n'exige aucun effort cognitif<br>_ non adapté aux domaines rationnels |

***Tableau 2.1 Comparaison entre filtrage collaboratif et social*** [1]

Quand nous parlons d'information sociale nous parlons généralement de réseaux sociaux. Plusieurs types de réseaux sociaux sont considérés en fonction des contextes d'études [65] :

- **réseaux de collègues ou personnes familières,**
- **réseaux d'amis,**
- **réseaux de confiance.**

[1] https://www.cairn.info/revue-document-numerique-2008-1-page-13.htm

## 2.4.2 Modélisation d'un réseau d'information sociale

Avant de passer à la modélisation d'un réseau d'information sociale quelques notions doivent être présentées à savoir les tags, l'opération d'étiquetage et les folksonomies.

- ***Un tag (étiquette)*** **:** est un mot clé librement choisi par un internaute pour décrire un contenu partagé sur le Web. Ce mot clé peut aussi être une combinaison de plusieurs mots. Un tag peut être associé à tout type d'information : des photos, des blogs, des vidéos, et même des pages présentant des produits ou des magasins. Comme les tags sont librement choisis, les contenus peuvent être décrits à partir de tout mot qui définit une relation entre la ressource en ligne et un concept activé dans l'esprit de l'internaute [51].

- ***Opération d'étiquetage*** **:** À l'utilisation de tags est liée la pratique d'étiquetage ou de tagging, c'est-à-dire l'association par un utilisateur d'un tag à une ressource donnée.

- ***Une folksonomie*** **:** est définie comme un modèle triparti où les ressources Web sont associées à un utilisateur par une liste de tags. Formellement, une folksonomie est un tuple $F = < U, T, R, A>$ où U, T et R représentent respectivement un ensemble d'utilisateurs, un ensemble de tags et un ensemble de ressources et A représente les relations entre les trois éléments précédents, c'est-à-dire $A \subseteq U \times T \times R$ [52].

### 2.4.2.1 Modélisation d'un réseau d'information sociale

Le réseau d'information sociale peut être formellement représenté par un graphe $G = (V, E)$ où l'ensemble des nouds $V = A \cup U \cup D \cup T$ représente les entités sociales avec A, U, D et T correspondant respectivement aux auteurs, aux utilisateurs, aux documents et aux annotations sociales. L'ensemble des arcs $E = (V \times V)$ représente les relations sociales reliant les différents types de nœuds (publier, co-auteur, amitié, citation, annotation, ...) [52].

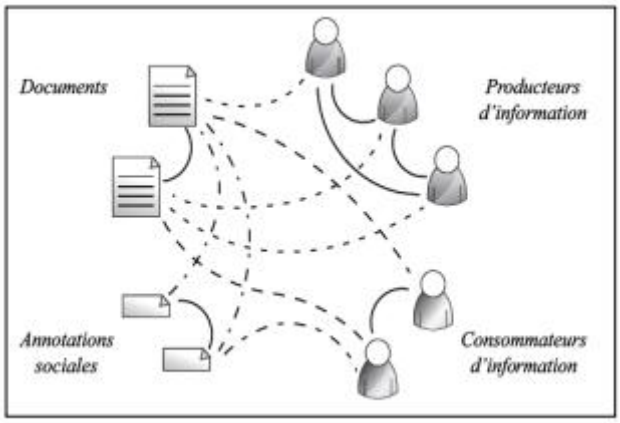



***Figure 2.2 : Le réseau d'information sociale [54]***

### 2.4.2.2 Processus de recherche d'information sociale

La recherche d'information sociale diffère d'autres approches par l'intégration de la structure du réseau social dans le processus de recherche. Ce dernier comprend trois étapes élémentaires : l'extraction de réseau social, l'analyse de réseau social et le classement par pertinence des ressources [53].

**Etape 1** : **Extraction de réseau social**. Dans cette étape, la structure du réseau social est extraite à partir du réseau d'information sociale.

**Etape2** : **Analyse des réseaux sociaux**. Dans cette étape, les méthodes d'analyse de réseaux sociaux sont appliquées sur le réseau social afin d'identifier les principales entités. Un score de pertinence sociale, basé sur les mesures de centralité, est attribué à chaque acteur du réseau.

**Etape3** : **Classement par pertinence de ressource**. Dans cette étape, un score de pertinence de ressources est calculé à base du score de pertinence sociale afin de produire un classement final des ressources [54].

## 2.5 Conclusion

Ce chapitre nous a permis de voir des concepts clés sur la recommandation sociale ainsi que les réseaux sociaux. Dans le chapitre qui suit nous nous intéresseront à l'approche que nous avons proposée pour construire notre système de recommandation sociale.

# Partie 2 : Approche proposée pour le système de recommandation sociale

- Chapitre 3 : Modélisation de l’approche.
- Chapitre 4 : Evaluation et validation.

# 3 Modélisation

## 3.1 Introduction

Le nombre croissant d'utilisateurs fournissant des informations sur eux-mêmes à travers leurs activités d'étiquetage sociales a induit l'émergence d'approches de profilage fondées sur les tags, qui supposent que les utilisateurs exposent leurs préférences sur des contenus au travers de tags [51]. De ce fait, les tags peuvent être utilisés pour construire des recommandations.

Dans ce chapitre nous allons présenter l'architecture générale du système proposé, les étapes de création du réseau social, l'analyse du réseau social à savoir le calcul du score de pertinence sociale des utilisateurs et des ressources, le classement de ces ressources selon le score et enfin la recommandation de ces ressources selon leurs score, le profil utilisateur ainsi que son réseau social.

## 3.2 Architecture générale du système

Dans notre approche nous proposons une recommandation des ressources à des utilisateurs à base de leur réseau social. Pour ce faire, nous devons créer en premier lieu le réseau social des utilisateurs à partir du réseau d'information sociale ensuite faire émerger les utilisateurs les plus importants afin de connaitre le score des ressources et après chercher les ressources des amis de l'utilisateur en cause, les filtrer et enfin les classer selon le score calculé.

Et donc, nous devons passer par les étapes suivantes : la création du réseau social, l'analyse du réseau social, le calcul d'importance des ressources et enfin la recherche et le classement de ces ressources (Figure 3.1).

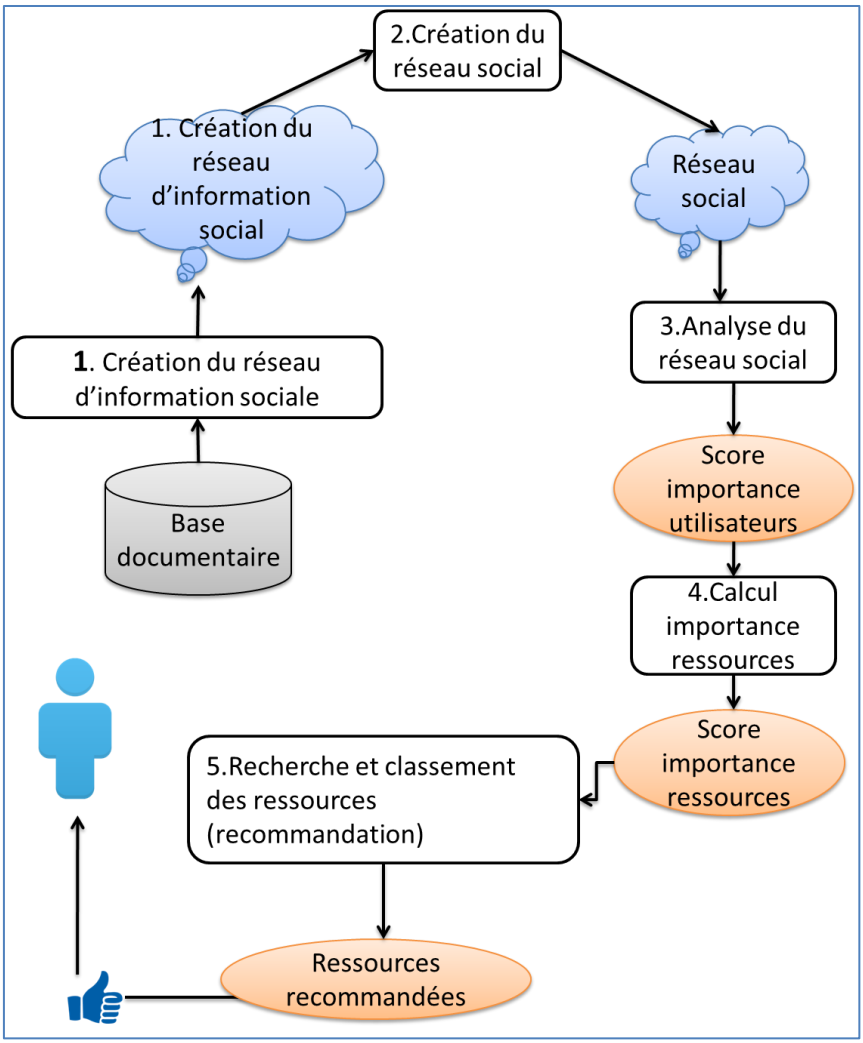



***Figure 3.1- Architecture du système de recommandation sociale du système proposé***

### 1. Création du réseau d'information sociale

Le réseau d'information sociale est extrait à partir de la base documentaire, il permet de créer et d'interpréter les liaisons entre utilisateurs, items et d'autres éléments de la base documentaires.

### 2. Création du réseau social

Dans cette étape nous devons construire les liens et relations entre les utilisateurs qui sont dans le réseau d'information sociale afin d'avoir un ensemble d'amis sous forme d'un réseau social.

### 3. Analyse du réseau social

Dans cette étape il faut calculer le score de centralité (l'importance) de chaque utilisateur dans le réseau social.

**4. Calcul de l'importance des ressources**

Dans cette étape le score de l'importance sociale de chaque ressource est calculé et ceci sur la base des scores de l'importance des utilisateurs.

**5. Recherche et classement des ressources**

Maintenant que nous avons les amis de l'utilisateur en cause, le score des ressources dans tous le réseau social, nous recherchons les items vus par ses amis puis recommander les ressources trouvées à l'utilisateur après classement de leurs scores.

# 3.3 Création du réseau social

## 3.3.1 Création du réseau

Notre réseau d'information sociale est composé de trois ensembles : un ensemble d'utilisateurs, un ensemble de tags et un ensemble de ressources (films). Nous utilisons aussi la notion de pertinence du tag (score entre ressource et tag).

**Pertinence du tag**

La pertinence du tag est un score d'un tag dans un film sur une échelle de 0 à 1. Il a été calculé en utilisant un algorithme de machine learning basé sur la contribution de l'utilisateur sur le contenu qui comprend des étiquettes, des notes et des critiques textuelles [74].

| Utilisateurs | Tag | items | Pertinence |
|---|---|---|---|
| User1 | Tag1 | V1 | |
| | | V2 | |
| | Tag2 | V 3 | 0,14 |
| | Tag3 | V3 | |
| | | V4 | |
| User2 | Tag1 | V3 | |
| | Tag2 | V3 | |
| | Tag3 | V6 | |
| User3 | Tag3 | V3 | 0,56 |
| | | V5 | |
| | Tag4 | V6 | 0,40 |
| User4 | Tag4 | V6 | |

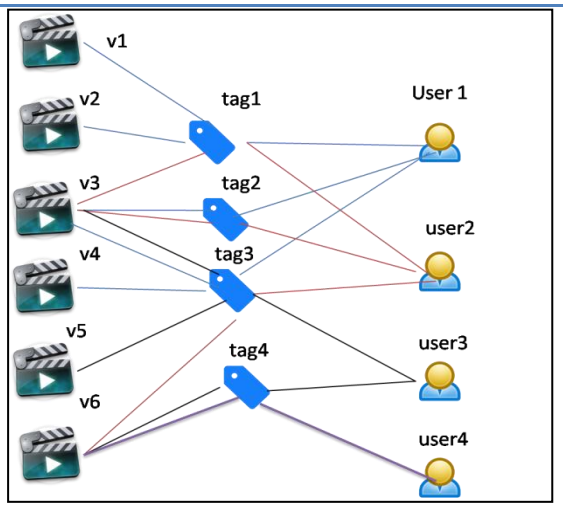



***Figure 3.2- Exemple de réseau d'information sociale***

A partir de ce réseau d'information sociale nous devons créer le réseau social des utilisateurs. Pour ce faire nous nous basons sur l'ensemble des tags exprimés par les utilisateurs pour annoter une ressource.

Nous avons décidé que notre réseau social connecte deux utilisateurs ***s'ils utilisent au moins un tag en commun pour marquer au moins une ressource en commun***, et donc un lien d'amitié est créé entre ces utilisateurs selon **l'algorithme Amitié** suivant :

## Algorithme amitié

**DEBUT**

u ∈ U= {u1, u2, u3, u4,……} --utilisateurs

t ∈ T= {t1, t2, t3, t4,……} --tags

r∈ R= {r1, r2, r3, r4,……} --ressources

n=0 --nombre de ressources annotées

Pour chaque (ui, uj) ∈ U*U :

```
Début
    Pour chaque (tk, tl) ∈ T*T
        Début
            Si tk(ui)=tl(uj) alors :
                Pour chaque (rm, rn) ∈ R*R
                Début
                    Si rm(tk)=rn(tl) alors :
                        n←n+1
                    finSi
                fin
            finSi
        fin
     Si  n>1 alors        /*critère : au moins un tag en commun*/
          Ajouter (ui.amis, uj)
          Ajouter (uj.amis, ui)
    fin
```

**FIN**

Le réseau social est représenté par un **graphe non orienté** où l'ensemble des nœuds représente les **utilisateurs** et l'ensemble des arcs représente les **relations d'amitié** entre ces utilisateurs.

Et donc le réseau social créé à partir du réseau d'information sociale de la figure 3.2 est le suivant :

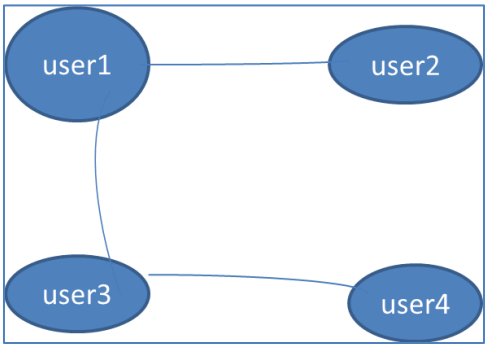



***Figure 3.3- Exemple de réseau social***

### 3.3.2 Pondération du réseau

Nous considérons que plus les utilisateurs co-marquent avec des tags communs plus leur relation est de poids élevé et par conséquent, ils ont une forte liaison (intérêt commun).

Nous proposons d'assigner des poids aux relations de co-marquage comme suit (indice de Jaccard) :

$$Co(i,j) = \frac{w_t(i,j)}{w_t(i)+w_t(j)-w_t(i,j)} \quad (3.1)$$

Avec :

- **wt(i,j) :** la somme des pertinences des items co-marqués par les utilisateurs ui et uj.

- **wt(i)** et **wt(j) :** la somme des pertinences des items marqués par chaque utilisateur ui et uj.

**L'algorithme poids** explique le calcul des paramètres wt(i) , wt(j) et wt(i,j) pour le calcul des poids des différentes relations de co-marquage.

**Algorithme poids**

**DEBUT**

u ∈ U= {u1, u2, u3, u4,……}--utilisateurs

t ∈ T= {t1, t2, t3, t4,……}--tags

r∈ R= {r1, r2, r3, r4,……} --ressources

wt(i,j)← 0, : la somme des pertinences des items co-marqués par les utilisateurs ui et uj.

wt(i)←0, la somme des pertinences des items marqués par l'utilisateur ui pour chaque utilisateur $u_i \in U$

Début

$w_t(i) \leftarrow 0$;

Pour chaque utilisateur $u_j$ $(u_j \neq u_i) \in U$

Début

$w_t(i,j) \leftarrow 0$;

Pour chaque item $r_k \in R$ marqué par $u_i$

Début

Si $r_k$ est marqué par $u_j$

Pour chaque tag $t_z \in T$ utilisé par $u_i$

début

Si $t_z$ est utilisé par $u_j$

$w_t(i) \leftarrow w_t(i)+pertinence(k,z)$; $w_t(i,j) \leftarrow w_t(i,j)+ pertinence(k,z)$;

fin

Finsi

fin

fin

fin

**FIN.**

Le réseau de la figure 3.3 devient alors :

| | | |
|---|---|---|
| **User1** | **User2** | **0,14/ (0,14+(0,14+0,56)-0,14)=0,2** |
| **User1** | User3 | 0,56/ ((0,14+0,56)+(0,40+0,56)-0,56)=0,5 |
| **User3** | User4 | 0,40/(0,40+(0,40+0,56)-0,40)=0,42 |

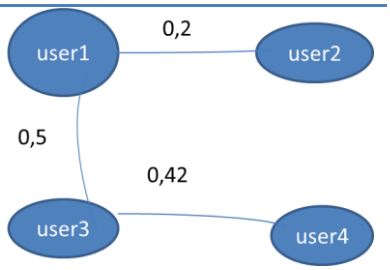



***Figure 3.4. Exemple de réseau social pondéré***

# 3.4 Analyse du réseau d'information sociale

## 3.4.1 Calcul du score social des utilisateurs (Page Rank)

Nous proposons d'appliquer cet algorithme pour analyser les nœuds de notre réseau social afin calculer **l'importance sociale** de chaque utilisateur dans ce réseau. On note ce score par : **CG(ui).**

## 3.4.2 Calcul du score de l'importance sociale des ressources

L'importance sociale d'une ressource est estimée à travers l'importance de ses utilisateurs. Après avoir calculé pour chaque utilisateur ui un score d'importance sociale CG (ui) à partir du réseau social des utilisateurs, le score d'importance sociale attribué à chaque ressource est calculé par la somme des scores sociaux de ses utilisateurs.

$$\mathrm{Imp_G}(d) = \sum_{i=1}^{m} C_G\,(u_i) \qquad (3.2)$$

Avec m : le nombre d'utilisateurs qui ont marqué la ressource d.

## 3.5 Classement des ressources et recommandation

Après avoir extrait notre réseau social, calculé le score social des utilisateurs (Page Rank), et calculé le score de pertinence sociale des ressources, nous allons utiliser toutes ces informations pour faire un classement des ressources à recommander aux utilisateurs.

Donc nous allons extraire les amis de chaque utilisateur, rechercher les ressources de ces amis et enfin classer ces ressources par priorité à l'aide de score d'importance des ressources et les recommander à l'utilisateur.

## 3.6 Mise en œuvre de l'approche proposée

Nous avons choisi **VISONE**[1] (VIsual SOcial NEtworks) pour le calcul de l'importance sociale. VISONE est un logiciel d'analyse et de représentation graphique des réseaux sociaux (SNA) développé par une équipe pluridisciplinaire de l'université de Konstanz, en constante évolution. Cet outil permet de récupérer des données, les analyser et enfin les visualiser sous forme de graphe.

Le logiciel possède de nombreuses qualités :

- Une interface utilisateur assez intuitive.
- Une bibliothèque de fonctions d'analyse du réseau qui s'est enrichie au fil des versions.
- Un véritable talent en matière graphique qui permet à l'utilisateur de réaliser des représentations à la fois claires et esthétiques (sauf pour les boucles).
- La possibilité, via un système d'onglets, d'ouvrir plusieurs réseaux à la fois, pour représenter, par exemple, l'état du réseau à différentes périodes ou encore des vues différentes d'un même réseau.

[1] http://visone.info/wiki/images/a/ad/Visone-tutorial-french.pdf

## 3.7 Conclusion

Dans ce chapitre nous avons vu comment construire le réseau social à base de tags décrits par les utilisateurs sur les ressources afin de pouvoir utiliser par la suite ce réseau dans le système de recommandation sociale. Le dernier chapitre concerne l'évaluation et la validation du système de recommandation sociale.

# 4. Evaluation et validation

## 4.1 Introduction

Dans ce chapitre, nous allons évaluer la performance de notre système de recommandation sociale pour s'assurer de sa capacité à satisfaire les besoins des utilisateurs.

Comme le domaine de la recommandation dérive du domaine de la recherche d'information, il est donc souvent normal d'utiliser des mesures d'évaluation de la recherche d'information. Certaines de ces mesures ont été ajustées aux besoins du domaine de la recommandation.

Dans ce qui suit nous présentons les principaux jeux de données utilisés pour l'évaluation *offline* des systèmes de recommandation, puis le principe de fonctionnement de ce mode d'évaluation, suivi de la description des principales mesures pour évaluer la précision des recommandations et enfin la comparaison de notre modèle à un autre système de recommandation.

## 4.2 Cadre d'évaluation

### 4.2.1 Jeux de données

Nous avons utilisé deux jeux de données ***Movielens 10m*** et ***Movielens 20m***. Du premier jeu de données nous avons utilisé les données : tag, movie et note et du deuxième jeu nous avons utilisé la pertinence des tags.

- **Movielens 20m :**

Cet ensemble de donnée décrit l'activité de marquage de vidéos à partir de movielens[1] qui est un service de recommandation de film. Il contient 20000263 notes et 465564 tags appliqués à 27278 films. Ces données ont été créées par 138493 utilisateurs entre le 9 janvier 1995 et le 31 mars 2015. Cet ensemble de données a été généré le 31 mars 2015 et mis à jour le 17 octobre 2016.

[1] (http://movielens.org)

- **Movielens 10m :**

Cet ensemble de données contient 10000054 notes et 95580 tags appliqués à 10681 films par 71567 utilisateurs du service de recommandation de films en ligne MovieLens.

- **Jeux d'apprentissage et de test**

Dans le jeu de données de MovieLens 10m, le partage entre apprentissage et test a été fait à l'aide du script (***split_ratings.sh***). Ainsi, pour chaque utilisateur, 80% de ses votes font partie du jeu apprentissage et 20% de ses votes font partie du jeu test. Une fois le modèle de recommandation construit à partir du jeu apprentissage, le modèle est sollicité pour prédire pour chaque utilisateur, l'utilité (vote ou pertinence) des items présents dans le jeu test. L'évaluation consiste alors, à comparer les recommandations proposées par l'algorithme aux votes réels se trouvant dans le jeu de test.

- **Caractéristique du réseau social**

Le tableau 4.1 contient les caractéristiques du réseau social des utilisateurs construit à partir du réseau d'information sociale.

| **Catégorie** | **Nombre** |
|---|---|
| Nombre d'utilisateurs | 1009 |
| Nombre de relations | 11484 |

***Tableau 4.1- Caractéristiques du réseau***

## 4.2.2 Jugements de pertinence

Pour constituer la liste des items pertinents, nous avons supposé que l'item recommandé est pertinent s'il existe dans le jeu de test.

## 4.2.3 Métriques d'évaluation

Dans ce cadre d'évaluation, nous utilisons la mesure classique de la précision pour les X premiers items recommandés (précision@X). C'est la proportion d'items pertinents dans les X premiers items retrouvés pour chaque utilisateur. Elle permet d'exprimer la satisfaction de l'utilisateur vis-à-vis des X premiers résultats pertinents. Elle constitue ainsi une mesure importante pour l'évaluation de la haute précision. Nous retenons les précisions pour les 1, 5, 10, 15 et 20 premiers movies notés respectivement P@1, P@5, P@10, P@15 et P@20. A la fin, des moyennes de toutes ces précisions (pour chaque X=1, 5, 10, 15 et 20) sont calculées sur toutes les recommandations selon la formule (1.7).

Nous obtenons les résultats suivants :

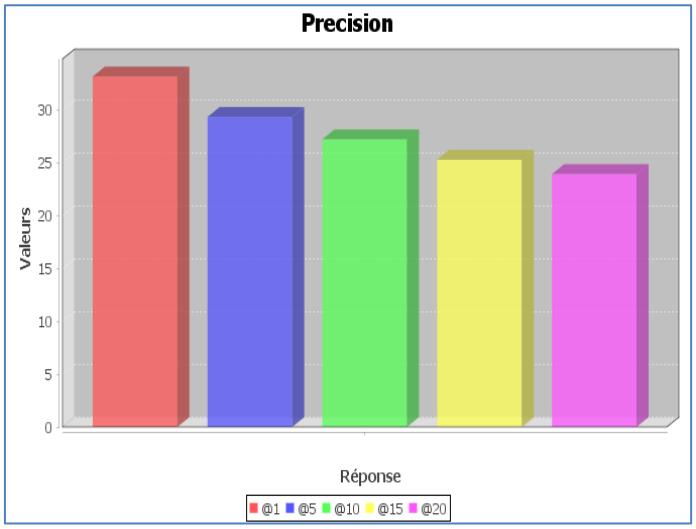



***Figure 4.1- Précision du système de recommandation sociale***

- **Analyse des résultats**

Nous remarquons que nous avons obtenu une précision de 33% pour le premier résultat (P@1), un résultat qui n'a pas atteint la moitié du total proposé. La précision diminue progressivement à 29% pour la P@5, 27% pour la P@10 et à 25% pour la P@15 jusqu'à arriver à 23% pour la P@20.

## 4.3 Evaluation de l'efficacité de l'approche proposée

Nous comparons notre système de recommandation sociale proposé au système classique collaboratif. Nous obtenons les résultats suivants :

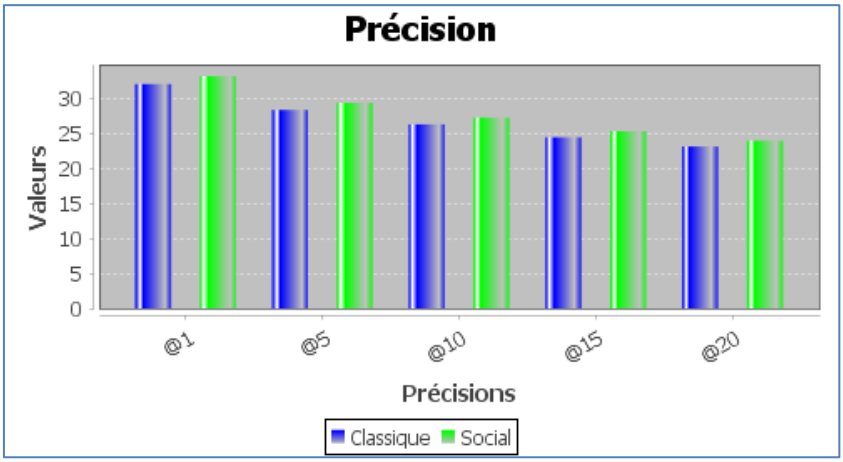



***Figure 4.2- Evaluation de l'efficacité de l'approche***

| | Classique | Notre modèle | Apport de notre modèle % au modèle classique |
|---|---|---|---|
| **P@1** | 32,05% | 33,20% | **1,15%** |
| **P@5** | 28,34% | 29,36% | **1,02%** |
| **P@10** | 26,30% | 27,24% | **0,94%** |
| **P@15** | 24,44% | 25,32% | **0,88%** |
| **P@20** | 23,13% | 23,96% | **0,83%** |

***Tableau 4.2- Apport du modèle***

- **Analyse des résultats**

Nous remarquons que les résultats du système de recommandation sociale ont amélioré les résultats du système classique. Comme décrit dans le tableau 4.2, notre modèle permet d'aboutir à une amélioration de la P@1 jusqu'à 1,15% par rapport au modèle classique.

De même, les résultats du système de recommandation sociale améliorent la P@5 de 1 ,02%, la P@10 de 0,94%, la P@15 de 0,88% et la P@20 de 0,83%, par rapport au modèle collaboratif classique.

Nous déduisons donc, que notre solution a amélioré les résultats du système classique ce qui satisfait les utilisateurs du système de recommandation.

- **Critiques des résultats**

Pour justifier les résultats faibles du système de recommandation sociale, plusieurs raisons peuvent être la cause :

- On peut se retrouve dans une situation où un item ne soit tagué que par des personnes aux centres d'intérêt différents. En conséquence, il sera difficile de recommander un tel item, car il ne fera pas partie des éléments du réseau d'information sociale.
- Plusieurs vidéos sont notées mais non taguées (66%) alors que notre solution démarre principalement d'un réseau social à base de tag ce qui implique qu'un utilisateur n'aura pas suffisamment d'amis importants.
- La construction du réseau se fait à base de tags, ces tags peuvent avoir le même sens mais pas la même syntaxe.

- **Amélioration des résultats**

Nous pouvons Améliorer les résultats de notre approche de plusieurs manières à savoir :

- Choisir un jeu de données très tagué, par exemple un jeu de données extrait d'un réseau social de marquage (exemple : réseau publique de marquage *Delicious*).
- Créer une sémantique entre les tags.
- …

# 4.4 Conclusion

Ce chapitre nous a permis d'évaluer notre approche qui intègre les relations sociales des utilisateurs à la recommandation classique à travers les tags utilisés pour marquer les vidéos et la pondération des relations sociales.

Les résultats expérimentaux de notre approche de construction de système de recommandation sociale ont été faibles et n'ont pas vraiment amélioré les résultats du système classique. De ce fait plusieurs propositions ont été mises en valeur afin d'augmenter la satisfaction de l'utilisateur final, à savoir : le changement du jeu de données et la création d'une sémantique entre les tags.

# Conclusion et perspective

Dans ce mémoire nous avons présenté une approche de recommandation de ressources vidéos basée sur les relations sociales. Nous avons détecté les liens sociaux en utilisant une technique basée sur les annotations des utilisateurs. Pour cela nous avons proposé une approche reposant sur l'analyse de graphes, en espérant notamment s'affranchir les limitations traditionnelles.

Nous avons également présenté une illustration et une évaluation que nous avons effectuée pour tester et valider notre approche et ceci en utilisant le jeu de données Movielens.

Comme perspectives nous pouvons améliorer notre système de recommandation sociale en renfonçant les liens d'amitiés et ceci en augmentant le nombre de tags en communs par exemple en considèrent que deux utilisateurs sont amis s'ils partagent au moins 3 tags en commun pour annoter les mêmes ressources et non pas juste un seul tag. De même nous proposons de rajouter d'autres attributs (par exemple : genre et réalisateur) comme critère de construction du lien d'amitié. Aussi nous pouvons construire au sein de notre réseau social des communautés où nous affectons un utilisateur à une communauté selon certains critères.

# Références bibliographiques

[1] **NICOLAS BECHET** « Etat de l'art sur les Systèmes de Recommandation » Projet AxIS de l'INRIA, dans le cadre du projet Addictrip.

[2] **CHETTI SARA** «Approche mixte entre la méthode de filtrage basé sur le contenu et la méthode de filtrage collaboratif pour la recherche d'informations», mémoire Master Informatique Université D'EL-OUED, Octobre 2014.

[3] **An Te NGUYEN**, thèse Doctorat : «COCoFil2 : Un nouveau système de filtrage collaboratif basé sur le modèle des espaces de communautés», Université Joseph Fourier Grenoblei, 23/11/2006.

[4] **CATHERINE B ERRUT ET NATHALIE D ENOS**, chapitre 8 «Filtrage collaboratif», p 242-268, E. GAUSSIER, M.H. STEFANINI, « Assistance intelligente à la RI », HermesLavoisier, 2003.

[5] O**. NOUALI, S. BENMEZUINE, M. DJAID, S. SIDI-BOUMEDIENE,** Chapitre rédigé «Filtrage de l'Information», p 27-35, Laboratoire Intelligence Artificielle, CE.R.I.S. T, Rue des 3 frères Aïssou, Ben Aknoun, Alger, Algérie, RIST Vol.7 N°02 Année 1997.

[6] **NEGRE ELSA**, projet de recherche: «Les systèmes de recommandation», Université Paris-Dauphine, 2008.

[**7] IMANE BOUSSEBOUGH EP.BOUGHERRA**, thèse Doctorat : «Les systèmes multi-agents dynamiquement adaptables », département informatique, Université Mentouri Constantine, 06/07/2011.

[8] **M. AMOKRANE BELLOUI**, thèse magister: «L'usage des concepts du web sémantique dans le filtrage d'information collaboratif», Ministère de l'enseignement Supérieur et de la recherche scientifique Institut National d'Informatique Alger, 2008.

[9] **ROZA LEMDANI**. Système hybride d'adaptation dans les systèmes de recommandation. Thèse de Doctorat. Université Paris-Saclay, 2016.

[10] **PAUL RESNICK, NEOPHYTOS IACOVOU, MITESH SUCHAK, PETER BERGSTROM, AND JOHN RIEDL.** Grouplens: An open architecture for collaborative filtering of netnews. In Proceedings of ACM Conference on Computer Supported Cooperative Work, pages 175–186, 1994.

[11] **UPENDRA SHARDANAND AND PATTIE MAES**. Social information filtering: algorithms for automating"word of mouth". In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '95, pages 210–217, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[12] **JOHN S. BREESE, DAVID HECKERMAN, AND CARL KADIE**. Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, UAI'98, pages 43–52, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

[13] **CHARU C. AGGARWAL, JOEL L. WOLF, KUN-LUNG WU, AND PHILIP S. YU. HORTING HATCHES AN EGG** : a new graph-theoretic approach to collaborative filtering. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '99, pages 201–212, New York, NY, USA, 1999. ACM.

[14] **BADRUL SARWAR, GEORGE KARYPIS, JOSEPH KONSTAN, AND JOHN REIDL**. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web (WWW), pages 285–295, 2001.

[15] **SYLVAIN CASTAGNOS, ARMELLE BRUN, AND ANNE BOYER**. Probabilistic association rules for item-based recommender systems. In Proceedings of the Fourth Starting AI Researchers' Symposium (STAIRS), pages 36–46, 2008.

[16] **XIN JIN AND BAMSHAD MOBASHER**. Using semantic similarity to enhance item-based collaborative filtering. In 2nd IASTED International Conference on Information and Knowledge Sharing, 2003.

[17] **G. LINDEN, B. SMITH, AND J. YORK. AMAZON.COM RECOMMENDATIONS**: item-to-item collaborative filtering. Internet Computing, IEEE, 7(1): 76–80, 2003.

[18] **ABHINANDAN S. DAS, MAYUR DATAR, ASHUTOSH GARG, AND SHYAM RAJARAM. GOOGLE NEWS PERSONALIZATION**: scalable online collaborative filtering. In Proceedings of the 16th international conference on World Wide Web, WWW '07, pages 271–280, New York, NY, USA, 2007. ACM.

[19] **BADRUL SARWAR, GEORGE KARYPIS, JOSEPH KONSTAN, AND JOHN RIEDL**. Analysis of recommendation algorithms for e-commerce. In Proceedings of the 2nd ACM conference on Electronic commerce (EC), pages 158–167, 2000.

[20] **KARAOUZENE MERYEM** «Système de recommandation des services web sémantiques» .mémoire Master Informatique .Université Abou Bakr Belkaid– Tlemcen *Juin 2015.*

[21] **BOUCHINDHOMME AND ROCHLITZ (1992).** Dans le flou artistique. Eléments d'une théorie de la "rationalité esthétique". page 203–238.

[22] **PIAMRAT, K., VIHO, C., BONNIN, J.-M., AND KSENTINI, A. (2009).** Quality of experience measurements for video streaming over wireless networks. In Sixth International Conference on Information Technology New Generations, 2009.ITNG '09, page 1184 –1189.

[23] **BURKE, R. (2002).** Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4):331–370.

[24] **ARNAUTU, O. R. (2012).** Mures : Un système de recommandation de musique. Master's thesis, La Faculté des arts et des sciences Universite de Montreal.

[25] **NORA KOCH**. Software engineering for adaptive hypermedia systems : Reference model, modeling techniques and development process. Softwaretechnik- Trends, 21(1), 2001.

[26] **DIEUDONNE TCHUENTE** « Modélisation et dérivation de profils utilisateurs à partir de réseaux sociaux : Approche à partir de communauté de réseaux k-égocentriques ». Thèse de Doctorat .Université de Toulouse, Janvier 2013.

[27] **BOUZEGHOUB M., KOSTADINOV D**., Personnalisation de l'information: aperçu de l'état de l'art et définition d'un modèle flexible de profils. Conférence en Recherche d'Informations et Applications, CORIA 2005, 201—218 (2005).

[28] **ZAYANI C**., Contribution à la définition et à la mise en œuvre de mécanismes d'adaptation de documents semi structurés. These de doctorat, Universite Paul Sabatier, mai 2008.

[29] **AMATO G., UMBERTO S**., User Profile Modeling and Applications to Digital Libraries. In ECDL 1999, European Conference on Research and Advanced Technology for Digital Libraries: pp184—197 (1999).

[30] **ABBAR S., BOUZEGHOUB M., KOSTADINOV D., LOPES S.,** A contextualization service for a Personalized Access Model. 9emes Journees Francophones Extraction et Gestion des Connaissances, EGC 2009:265-270, 2009.

[31] **TCHUENTE D., CANUT M-F., JESSEL N., PENINOU A., HADDADI A. E**., Visualizing the evolution of users' profiles from online social networks. International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2010), Odense, Danemark, 09/08/10-11/08/10, IEEE Computer Society, p. 370-374, aout 2010.

[32] **TCHUENTE D., CANUT M-F., JESSEL N., PENINOU A., SEDES F**., Visualizing the relevance of social ties in user profile modeling. Web Intelligence and Agent Systems, IOS Press, Vol. 10 N. 2, p. 261-274, mai 2012.

[33] **HOLLAND S., WERNER K**., Situated Preferences and Preference Repositories for Personnalized Database Applications, 23rd International Conference on Conceptual Modeling, ER 2004, Shanghai, China, (2004).

[34] **DROMZEE C., LABORIE S., ROOSE P**., Profil générique sémantique pour l'adaptation de documents multimédias. INFormatique des ORganisations et Systemes d'Information et de Decision (INFORSID 2012), 191-206, 2012.

**[**35] **BOUCHRA S**., Vers un système d'information Web fournissant des services Web sensibles au contexte. These de doctorat, Universite Paul Sabatier, avril 2010.

[36] **KOBSA A**., Generic User Modeling Systems, User Modeling and User Adapted International Journal, Vol. 11, 49-63, 2001.

[37] **GAO M., LIU K., WU Z**., Personalisation in web computing and informatics: Theories, techniques, applications, and future research. Information Systems Frontiers 12(5): pp. 607--629 (2010).

[38] **FRIAS-MARTINEZ E., MAGOULAS G., CHEN S., MACREDIE R**., Automated user modeling for personalized digital, International Journal of Information Management, 26, 234–248.

[39] **NANOPOULOS A., KATSAROS D., MANOLOPOULOS Y**., Effective prediction of web-user accesses: A data mining approach. WEBKDD, 1, 2001.

[40] **SARUKKAI R. R**., Link prediction and path analysis using Markov chains. Computer Networks, 33, 377–386,(2000).

[41] **KIM J. K., CHO Y. H., KIM W. J., KIM J. R., SUH J. H**., A personalized recommendation procedure for Internet shopping support. Electronic Commerce Research and Applications, 1, 301–313. (2002).

[42] **CHO Y. H., KIM J. K., KIM, S. H.**, A personalized recommender system based on web usage mining and decision tree induction. Expert Systems With Applications, 23, 329–342, (2002).

[43] **JUNG S. Y., HONG J. H., KIM, T. S**., A statistical model for user preference. IEEE Transactions on Knowledge and Data Engineering, 834–843, (2005).

[44] **SCHUBERT P., KOCH M**., The power of personalization: customer collaboration and virtual communities. Proceedings of the Eighth Americas Conference on Information Systems (AMCIS), 1953– 1965, (2002).

[45] **RUVINI J. D**., Adapting to the user's internet search strategy. Proceedings of the 9th International Conference on User Modeling (UM2003), Pittsburgh, 55–64, (2003).

[46] **HORVITZ E., BREESE J., HECKERMAN D., HOVEL D., ROMMELSE K**., The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, 256–265, (1998).

[47] **SIMON RENAUD-DEPUTTER,** Système de recommandations utilisant une combinaison de filtrage collaboratif et de segmentation pour des données implicites, Mémoire présenté au Département d'informatique en vue de l'obtention du grade de maître ès sciences (M.Sc.)

[48] **JONATHAN L. HERLOCKER, JOSEPH A. KONSTAN, LOREN G. TERVEEN ET JOHN T. RIEDL** :Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS), 22(1):5—53, janvier 2004.

[49] **SLIMANE LEMMOUCHI**. Etude de la robustesse des graphes sociaux emergents. Universite Claude Bernard - Lyon I, 2012. Francais.

[50] **CHIKHI, N. F. 2010.** "Calcul de centralité et identification de structures de communautés dans les graphes de documents". Thèse de doctorat. Université Toulouse 3 Paul Sabatier.

[51] **SAMIA BELDJOUDI, HASSINA SERIDI, ABDALLAH BENZINE**. Améliorer la Recommandation de Ressources dans les Folksonomies par l'Utilisation de Linked Open Data.

[52] **GUY M. ET TONKIN E. (2006)**, Folksonomies: Tidying up tags?, *D-Lib Magazine*, 12, 1 (http://www.dlib.org/dlib/january06/guy/01guy.html).

[53] **KIRCHHOFF, L., STANOEVSKA-SLABEVA, K., NICOLAI, T., FLECK, M**. "Using social network analysis to enhance information retrieval systems". Technical report. University of St.Gallen - Alexandria Repository (Switzerland). 2008.

[54] **FATIHA SADOUKI.** « Intégration de l'analyse des réseaux sociaux dans le processus de recherche d'information», mémoire de magister Informatique. Université BLIDA .avril 2015.

[55] **ELLENM. VOORHEES**. The philosophy of information retrieval evaluation. In Carol Peters, Martin Braschler, Julio Gonzalo, and Michael Kluck, editors, Evaluation of Cross-Language Information Retrieval Systems, volume 2406 of Lecture Notes in Computer Science, pages 355–370. Springer Berlin Heidelberg, 2002.

[56] **GUY SHANI AND ASELA GUNAWARDANA**. Evaluating recommendation systems. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, Recommender Systems Handbook, pages 257–297. Springer US, January 2011.

[57] **SONIA BEN TICHA** Recommandation Personnalisée Hybride THESE présentée et soutenue publiquement le 11 novembre 2015 pour l'obtention du Diplôme de Docteur en Informatique.

[58] **BILLSUS AND PAZZANI**, 1998. Learning collaborative infor-mation filters. In Proceedings of the Fifteenth International Conference on Machine Learning,ICML '98, pages 46–54, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

[59] **FREEMAN L. C. (1979).** Centrality in social networks: Conceptual clarification. Social Networks,vol. 1, no 3, p. 215-239.

[60] **WASSERMAN S., FAUST K. (1994).** Social network analysis, methods and application (4th edition:1998, 1999 éd.; C. U. Press, Ed.). Cambridge University Press, New York, USA.

[61] **HAKIMI S. L. (1964).** Optimum locations of switching centers and the absolute centers and medians of a graph. Operations Research, vol. 12, no 3, p. 450–459.

[62] **R. KUMAR, P. RAGHAVAN, S. RAJAGOPALAN AND A. TOMKINS**. Core algorithms in the CLEVER system. ACM Trans. Internet Technol., 6(2):131-152, 2006.

[63**] J. KLEINBERG**. Authoritative sources in a hyperlinked environment. In Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms, San Francisco, California, United States, Society for Industrial and Applied Mathematics, pages, 1998.

[64] **J. KLEINBERG**. Authoritative sources in a hyperlinked environment. J. ACM,46(5):604-632, 1999.

[65] **KRISTEN MORI,** trust-networks in recommender systems .December 2008.

[66] **J.S. BREESE, D. HECKERMAN, AND C. KADIE**. Empirical analysis of predictive algorithms for collaborative filtering. In Proceeding of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI), pages 43–52, Morgan Kaufmann, San Francisco, 1998. Madison, Wisconsin.

[67] **BAMSHAD MOBASHER, XIN JIN, AND YANZAN ZHOU**. Semantically enhanced collaborative filtering on the web. In Bettina Berendt, Andreas Hotho, Dunja Mladenic, Maarten van Someren, Myra Spiliopoulou, and Gerd Stumme, editors, Web Mining : FromWeb to SemanticWeb, volume 3209 of Lecture Notes in Computer Science, pages 57–76. Springer Berlin / Heidelberg, 2004.

[68] **MUKUND DESHPANDE AND GEORGE KARYPIS, (JANUARY 2004).** Itembased top-n recommendation algorithms. ACM Trans. Inf. Syst., 22(1), pp.143–177.

[69] **GUY SHANI AND ASELA GUNAWARDANA**. Evaluating recommendation systems. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, Recommender Systems Handbook, pages 257–297. Springer US, 2011.

[70] **BADRUL SARWAR, GEORGE KARYPIS, JOSEPH KONSTAN, AND JOHN RIEDL**. Itembased collaborative filtering recommendation algorithms. In Proceedings of the 10th International Conference on World Wide Web, WWW '01, pages 285–295, New York, NY, USA, 2001. ACM.

[71] **PAUL RESNICK, NEOPHYTOS IACOVOU, MITESH SUCHAK, PETER BERGSTROM, AND JOHN RIEDL**. Grouplens : an open architecture for collaborative filtering of netnews. In The 1994 ACM conference on Computer supported cooperative work, page 175–186, 1994.

[72] **JAMES BENNETT AND STAN LANNING**. The netflix prize. In Proceedings of KDD Cup and Workshop, 2007.

[73] **C. CLEVERDON AND M. KEAN**. Factors determining the performance of indexing systems. Aslib Cranfield Research Project, Cranfield, England, 1968.

[74] **JING SIANG NG (GERALDJS) TING AN IAN NGIAW (IANNGIAW) BILI XU (XBILI),**[General Machine Learning] Predicting Movie Popularities Using Their Genomes.

[75] **NAAK AMINE** Papyres : Un système de gestion et de recommandation d'articles de recherche Juillet, 2009.

# Annexe A - Movielens

Nous avons utilisé le jeu de données ***Movielens 20m (genome.socre, genome.tag) et 10m (tags, ratings, movies)*** pour extraire notre réseau social. Le jeu contient des votes pour des films appréciés par des utilisateurs. Les films sont évalués par une note entre 0 et 5.

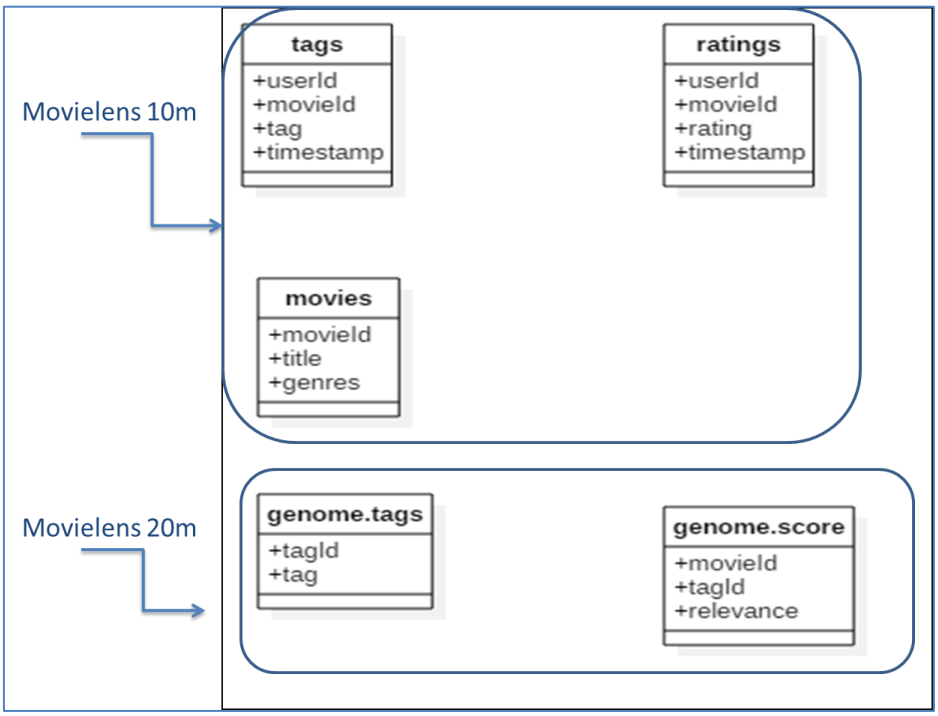



***Figure A.1. Movielens 20m et 10m***

Les relations entre les tables sont illustrées dans la figure suivant :

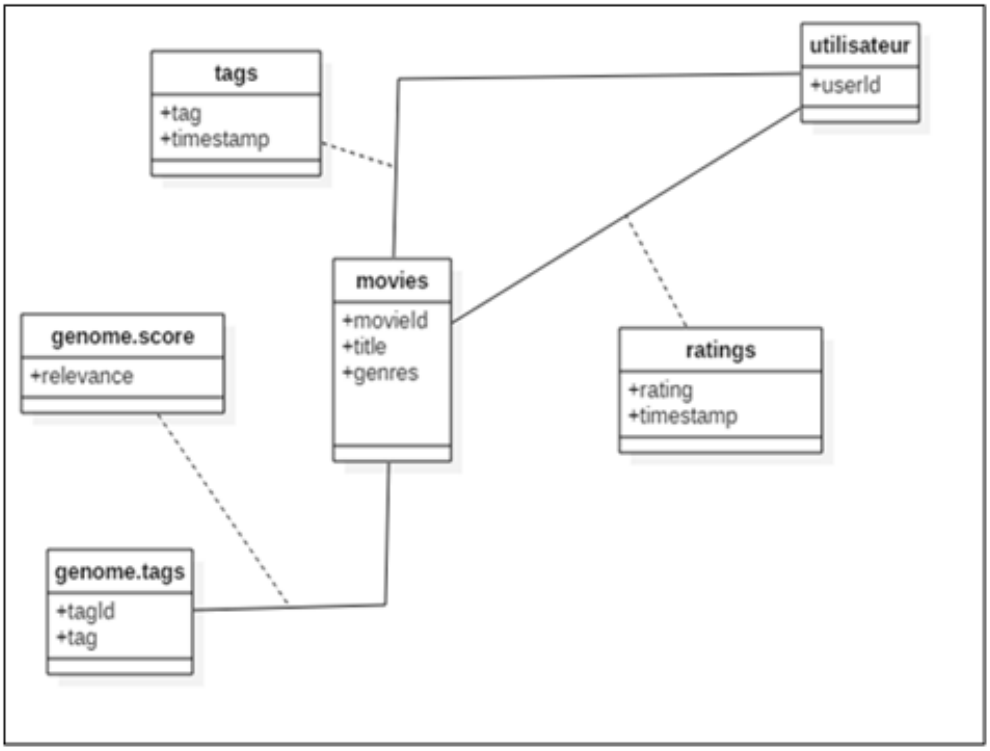



***Figure A.2. Relations entre les tables***

Le tableau suivant montre la structure des tables.

| champs | Signification | type | Valeur |
|---|---|---|---|
| userId | Identifiant utilisateur | Entier | - |
| movieId | Identifiant vidéo utilisé par https://movielens.org | Entier | Exemple : Si 1 alors https://movielens.org/movies/1 |
| Title | Titre vidéo | alphanumérique | |
| genres | Le type de vidéo | Chaine de caractère | • Action<br>• Adventure<br>• Animation<br>• Children's<br>• Comedy<br>• Crime<br>• Documentary<br>• Drama<br>• Fantasy<br>• Film-Noir |

| | | | |
|---|---|---|---|
| | | | • Horror<br>• Musical<br>• Mystery<br>• Romance<br>• Sci-Fi<br>• Thriller<br>• War<br>• Western<br>• (no genres listed) |
| tagId | Identifiant du tag | Entier | |
| Tag | Nom du tag d'une vidéo | alphanumérique | |
| timestamp | Représente les secondes depuis minuit Coordinated Universal Time (UTC) de janvier 1, 1970 | Réel | |
| rating | Note donnée par l'utilisateur à la vidéo | Réel | 0.5 étoile - 5.0 étoile |
| relevance | | réel | |

***Tableau A.1. Structure des tables***

# Annexe B - Programme Java

Le code ci-dessous permet d'exécuter un script contenant des requêtes SQL pour :

- Construire le réseau d'amis,
- calculer le poids entre deux amis,
- calculer l'importance des ressources,
- et enfin créer la liste des recommandations.

Il permet aussi de créer dynamiquement des tables SQL à X précision. Et enfin il calcule la précision à chaque X donné et génère des histogrammes.

**Code**

```
package recommandation;
/**
 *
 * @author PC-CLICK
 *
 */
import java.util.Random;
import org.jfree.chart.*;
import org.jfree.data.statistics.*;
import org.jfree.chart.plot.PlotOrientation;
import outils.connex;
import java.sql.*;
import java.io.*;
import com.mysql.jdbc.Driver;
import static java.lang.Integer.parseInt;
import java.util.Scanner;
import javax.swing.*;
```

```
import org.jfree.ui.RefineryUtilities;
public class Amitie {
    /**
     * @param args the command line arguments
     */
    public static void main(String[] args) {
        // TODO code application logic here
        try {
 Connection con = connex.conDB();
 Statement statement = con.createStatement();
script.executeDBScripts("d:\\good.sql", statement);
 Scanner sc = new Scanner(System.in);
 System.out.println("Donnez le nombre de précision à calculer: ");
 int nbr = sc.nextInt();
float valeurs1[] = new float[nbr];
float valeurs2[] = new float[nbr];
String[] label = new String[nbr];
  int j=0;
         for (int i = 0; i < nbr; i++)   {
             j+=1;
             System.out.println("La "+j+"ème précision est à @: ");
             int p = sc.nextInt();
             String nomT="";
              String nomTC="";
             //nom table
             nomT+="p"+p;
             nomTC+="pc"+p;
```

```
        //creation table @ p if not exists classique
       createTable.createTableCub(nomTC);
        //insertion donneés à top p classique
        insertion.insertTable(nomTC);//rajouter recommander
        //creation table @ p if not exists social
        createTable.createTableCub(nomT);
        //insertion donner à top p social
        insertion.insertTable(nomT);//rajouter recommander
        //calcul precision
      valeurs1[i] = eval.evaluation(p, nomTC, connex.conDB());
      System.out.print("v1"+valeurs1[i]);
      valeurs2[i] = eval.evaluation(p, nomT, connex.conDB());
      System.out.print("v2"+valeurs2[i]);
      label[i]="@"+p;
       }
      /* Histogram pie = new Histogram(valeurs2, label);
       pie.pack();
       pie.setVisible(true);
       pie.setDefaultCloseOperation(JFrame.EXIT_ON_CLOSE);*/
      final barChart demo = new barChart("Précision",label,valeurs1,valeurs2);
         demo.pack();
        RefineryUtilities.centerFrameOnScreen(demo);
       demo.setVisible(true);
    }catch(Exception sqlexp){
        System.out.println(sqlexp.getMessage());
    }
}
```