



HAL
open science

Personnalisation et Désindividuation : les deux faces des Systèmes de Recommandation dans les Réseaux Sociaux Numériques. Application à la recommandation cinéphile en France et en Chine

Zhenfei Feng

► **To cite this version:**

Zhenfei Feng. Personnalisation et Désindividuation : les deux faces des Systèmes de Recommandation dans les Réseaux Sociaux Numériques. Application à la recommandation cinéphile en France et en Chine. Sciences de l'information et de la communication. Université de Lille, 2021. Français. NNT : 2021LILUH014 . tel-03714131

HAL Id: tel-03714131

<https://theses.hal.science/tel-03714131>

Submitted on 5 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE LILLE

École Doctorale Sciences de l'Homme et de la Société
Laboratoire GERiICO

THÈSE

Présenté par

Zhenfei FENG

pour obtenir le titre de

Docteur de l'Université de Lille

Discipline : Sciences de l'information et de la communication

**Personnalisation et Désindividuation : les
deux faces des Systèmes de Recommandation
dans les Réseaux Sociaux Numériques.
Application à la recommandation cinéophile en
France et en Chine**

*Soutenue publiquement le 22 janvier 2021
devant le jury composé de*

Directeur : Pr. Laurence FAVIER, Université de Lille
Présidente : Pr. Widad EL HADI, Université de Lille
Rapporteur : Pr. Philippe VIALON, Université de Strasbourg
Rapporteur : Pr. Mohamed SIDIR, Université de Picardie- Jules Verne
Examinatrice : Pr. Chirine GHEDIRA GUEGAN, Université de Lyon 3
Examinateur : M. Julien MESANGEAU, Université de Paris 3
Invité : Pr. Luc QUONIAM, Université Fédérale du Mato Grosso do Sul

REMERCIEMENTS

Je tiens à adresser toute ma reconnaissance à ma Directrice de thèse Laurence FAVIER qui m'a permis de réaliser cette thèse pour son encadrement, sa disponibilité, ses encouragements, ses conseils et sa confiance pendant ces années de thèse.

Je remercie également Widad MUSTAFA EL HADI pour tout le temps qu'elle m'a consacré, pour son esprit d'écoute, pour les échanges intéressants qu'on a eu pendant la thèse et pour tous les conseils qu'elle m'a prodigué. Qu'elle trouve ici l'expression de ma reconnaissance.

J'adresse mes remerciements également aux membres du jury Philippe VIALON, Mohamed SIDIR pour avoir accepté d'être les rapporteurs de ma thèse, Chirine GHEDIRA GUEGAN, Julien MESANGEAU d'avoir été examinateurs de ma thèse et Luc QUONIAM d'avoir accepté d'être professeur invité à cette soutenance.

Mes remerciements s'adressent également à Delphine SPILEERS pour son aide notamment pour toutes les procédures administratives qui étaient liées à ma thèse.

Mes remerciements vont aussi à tous les membres du laboratoire de GERiiCO. J'ai beaucoup apprécié l'ambiance de travail et les moments agréables passés avec eux qui étaient riches tant sur le plan professionnel que personnel.

RESUME

Le développement de l'Internet et de la technologie Web 2.0 qui ajoute à la facilité de publication le contenu généré par l'utilisateur, mettent à la disposition des utilisateurs une variété d'informations dont le volume est sans cesse croissant. Face à ce problème de surcharge d'informations, il est difficile pour les utilisateurs de s'orienter et de repérer des informations qui répondent à leurs besoins.

De nombreux systèmes de filtrage de l'information sont développés pour faire face à ce problème : l'un d'entre eux est le système de recommandation. L'objectif principal des systèmes de recommandation est de fournir aux utilisateurs des propositions de contenus personnalisés. Le principe sous-jacent est de déduire les besoins d'information de l'utilisateur, puis d'identifier dans le système les informations qui répondent à des besoins et les lui recommander. Les systèmes de recommandation, largement utilisés dans divers domaines, peuvent aussi être intégrés à des réseaux sociaux. La plupart des réseaux sociaux se caractérisent à la fois par le nombre important d'interactions et par l'anonymat des utilisateurs. Ces caractéristiques correspondent aux conditions décrites en psychologie sociale pour qu'un état de désindividuation soit déclenché. Les utilisateurs des réseaux sociaux sont susceptibles de se trouver dans une situation où l'identité du groupe est significativement élevée et leur identité individuelle restreinte. Leurs pensées, leurs comportements et même leurs préférences sont fortement influencées par les normes de groupe, y compris, bien sûr, leurs rétroactions sur les informations reçues. Ces rétroactions pourraient être biaisées c'est-à-dire ne pas refléter les vraies préférences individuelles des utilisateurs. Ainsi les recommandations basées sur ces rétroactions biaisées seraient contraires à l'intention initiale des recommandations personnalisées.

Cette thèse est consacrée à l'exploration du phénomène de la désindividuation qui peut exister dans les réseaux sociaux et de son impact sur le comportement de notation des utilisateurs, tout en incluant les différences culturelles. Nous choisissons comme terrain d'étude les systèmes de recommandation de films, ce qui nous amène à examiner les utilisateurs de quatre plateformes pour les cinéphiles à travers leur comportement de notation de films. Les résultats confirment l'existence du phénomène de la désindividuation dans les réseaux sociaux son impact significatif sur le comportement de notation des utilisateurs. La différence culturelle

est également un facteur important qui influence le comportement de notation. Sur cette base, nous arguons que les systèmes de recommandation appliqués dans les réseaux sociaux doivent y faire attention et que certaines mesures visant à individualiser les utilisateurs devraient être prises avant de recueillir et d'analyser les réactions des utilisateurs.

ABSTRACT

The development of the Internet and Web 2.0 technology makes available to users a wide variety of information whose volume is constantly increasing. Faced with this problem of information overload, it is difficult for users to find information that meets their needs.

Many information filtering systems have been developed to deal with this problem, and one of them is the recommendation system. The main objective of recommendation systems is to provide users with personalized recommendations that meet their individual needs in order to facilitate access to information that may be of interest to them. The underlying principle is to infer the user's information needs based on the information feedback received from the user, and then to identify in the system the information that meets their needs and recommend it. Recommender systems are widely used in various fields, including online social networks. However, most social networks are characterized by anonymity and there is plenty of interactions between users. These correspond to the conditions described in social psychology for a deindividuation state to be triggered. Social network users are likely to fall into this state where their group identity is significantly higher, while their personal identity is restricted. Their thoughts, behaviours and even preferences are strongly influenced by group norms, including, of course, their feedback on the information they have received. Such biased information feedback would not reflect users' true individual preferences, and recommendations based on such biased feedback are contrary to the original intent of personalized recommendations.

This thesis is devoted to exploring the phenomenon of deindividuation that may exist in social networks and its impact on users' rating behaviour, while including cultural differences. We choose the traditional domain of recommendation systems, movie recommendations, and examine users of four platforms for movie fans and their rating behaviour. The results confirm the existence of the phenomenon of deindividuation in social networks and that it has a significant impact on users' rating behaviour. Cultural difference is also an important factor in influencing rating behaviour. On this basis, we argue that recommendation systems applied in social networks should pay attention to it and that some measures to individualize users should be taken before collecting and analyzing user feedback.

LISTES DES FIGURES

Figure 1 Folksonomie Large	37
Figure 2 Folksonomie Restreinte	38
Figure 3 Une Définition Problématique des RSN	44
Figure 4 Exemples des RSN avec des objectifs divers (Thelwall, 2009)	48
Figure 5 Panorama des médias sociaux 2019 (Cavazza, 2019)	49
Figure 6 Panorama des Média Sociaux Chinois (Kantar, 2019)	52
Figure 7 Nombre d'utilisateurs des médias sociaux chinois.....	52
Figure 8 Les médias sociaux chinois les plus actifs.....	53
Figure 9 Nombre d'utilisateurs français actifs des médias sociaux (We Are Social & Hootsuite, 2019b).....	54
Figure 10 Médias sociaux les plus actifs en France (We Are Social & Hootsuite, 2019b)	54
Figure 11 Modèle du filtrage d'information et modèle du RI selon (Belkin et Croft, 1992)	59
Figure 12 Classification de SFI selon Malone (1987).....	60
Figure 13 Les étapes et les méthodes principales d'une exploration des données (Ricci et al., 2011).....	66
Figure 14 les approches principales du SR	67
Figure 15 le SR basé sur le contenu vu comme une boîte noire (Negre, 2015) traduction de (Jannach et al., 2010).....	68
Figure 16 Une architecture de haut niveau de ce processus (Lops, de Gemmis et Semeraro, 2011).....	69
Figure 17 Vecteurs moyens pour les documents pertinents et non pertinents.....	77
Figure 18 Faire Qi approcher aux documents pertinents (Manning, Prabhakar et Hinrich, 2008).....	78
Figure 19 : nouvel item sera classifié à différent class suit à la différente valeur de k.....	84
Figure 20 Le SR collaboratif vu comme une boîte noire (Negre, 2015) traduction de (Jannach, 2010)	88
Figure 21 La décomposition en valeurs singulières	104
Figure 22 La décomposition en valeurs singulières: une matrice items× caractéristiques peut être décomposée en trois éléments différents: un item × concepts, la force de chaque	

concept et un concept × caractéristiques (traduction de (Ricci et al., 2011)).....	105
Figure 23 Deux applications de méthodes hiérarchiques.....	114
Figure 24 Profil d'un utilisateur sur Douban	137
Figure 25 Page d'accueil de Douban Film (version desktop).....	138
Figure 26 Zone de recommandation dans la version mobile	139
Figure 27 Page d'un film particulier de la version desktop.....	140
Figure 28 Page d'un film particulier de la version desktop (suite).....	141
Figure 29 Profil de l'utilisateur sur Allociné.....	143
Figure 30 Liste d'abonnements et d'abonnés de l'utilisateur	143
Figure 31 Page profil d'un film sur Allociné	145
Figure 32 Page profil d'un film sur Allociné (suite).....	146
Figure 33 Profil de l'utilisateur sur SensCritique	147
Figure 34 Page profil d'un film sur SensCritique	148
Figure 35 Page profil d'un film sur SensCritique (suite).....	149
Figure 36 Page profil d'un film sur Allociné	149
Figure 37 Page d'accueil de Maoyan	150
Figure 38 Page d'un film sur Maoyan.....	151
Figure 39 Les recommandations fournies par Allociné	153
Figure 40 Liste des films similaires pour un film proposée par Allociné	153
Figure 41 Les recommandations fournies par Douban	154
Figure 42 Les recommandations personnalisées dans l'application mobile de Douban Film	155
Figure 43 Sur Allociné, l'explication de notation se présente lorsque l'utilisateur note un film	158
Figure 44 Sur Douban, l'explication de notation se présente lorsque l'utilisateur note un film	158
Figure 45 La messagerie privée de Douban	160
Figure 46 Un commentaire sous la critique du film rédigée par un utilisateur de Douban	161
Figure 47 Zone de discussion d'un film.....	161
Figure 48 les forums et la messagerie privée d'Allociné.....	162
Figure 49 On peut laisser des commentaires sur des critiques d'autrui	162
Figure 50 La messagerie sur SensCritique.....	163

Figure 51 Une critique de film sur SensCritique.....	164
Figure 52 Profil de l'utilisateur de Maoyan	165
Figure 53 L'espace de critique sur la page d'un film sur Maoyan.....	166
Figure 54 La répartition des répondants par âge et sexe	176
Figure 55 La répartition de l'ensemble utilisateurs de Douban par âge et sexe en 2017	176
Figure 56 La répartition des répondants par le nombre de films qu'ils ont noté sur Douban	177
Figure 57 Corrélations entre les influences d'évaluations positives des notes moyennes, des amis et des critiques de cinéma	180
Figure 58 Corrélations entre les influences d'évaluations négatives des notes moyennes, des amis et des critiques de cinéma.....	181
Figure 59 La répartition des notes sur SensCritique.....	187
Figure 60 La répartition des notes sur Allociné.....	187
Figure 61 La répartition des films par nationalité	188
Figure 63 Répartition des films par la différence de notes des mêmes films sur Allociné et sur SensCritique	189
Figure 64 Comparaison des notes des films par genre sur les deux plateformes.....	190
Figure 64 La répartition des films par nationalité	191
Figure 65 La répartition des notes sur Douban	192
Figure 66 La répartition des notes sur Maoyan.....	192
Figure 67 Répartition des films par la différence de notes des mêmes films sur Douban et sur Maoyan	194
Figure 68 Différence de notes des films des 10 pays ayant le plus grand nombre de films sur les deux plateformes	195
Figure 69 La répartition de nationalité des films qui sont mieux notés sur Douban.....	196
Figure 70 Comparaison des notes des films des dix pays ayant le plus grand nombre de films sur les deux plateformes.....	196
Figure 71 Comparaison des notes des films par genre sur les deux plateformes.....	197
Figure 72 Répartition des films par leur nationalité.....	199
Figure 73 La répartition des notes sur Allociné.....	199
Figure 74 La répartition des notes sur Douban.....	200
Figure 75 Répartition des films par la différence de notes des mêmes films sur Douban et sur Allociné	201

Figure 76 Comparaison des notes des films par genre sur les deux plateformes.....	202
Figure 77 Comparaison des notes des films des dix pays ayant le plus grand nombre de films sur les deux plateformes.....	203

LISTE DES TABLEAUX

Tableau 1	Formats et versions de RSS (Courrier, 2007)	33
Tableau 2	La matrice $m \times n$ de notes utilisateurs-items pour FC.....	90
Tableau 3	Les explication de l'échelle de notation sur 3 plateformes	156
Tableau 4	Explication de l'échelle de notation sur Douban et MovieLens : le critère objectif et le critère subjectif	157
Tableau 5	La répartition de films par le nombre d'utilisateurs qui les notent.....	185
Tableau 6	La corrélation entre les notes des mêmes films sur Allociné et SensCritique	189
Tableau 7	La corrélation entre les notes des mêmes films sur Douban et sur Maoyan ..	193
Tableau 8	La corrélation entre les notes des mêmes films sur Douban et Allociné	200

LISTE DES ABREVIATIONS

AIM : AOL Instant Messenger

Ajax : Asynchronous JavaScript and XML

CMC : Computer-Mediated Communication

DSI : Diffusion Sélective de l'Information

FI : Filtrage de l'Information

FC : Filtrage collaboratif

RSN : Réseau Social Numérique

RSS : Really Simple Syndication

SR : Système de Recommandation

SIDE : Social Identity model of Deindividuation Effects

RDF : Resource Description Framework

RSS 0.91 : Rich Site Summary

RSS 2.0 : Really Simple Syndication

SFI : Système de Filtrage de l'Information

RI : Recherche de l'Information

VSM : Vector Space Model

TF-IDF : Term Frequency-Inverse Document Frequency

k NN : k -Nearest Neighbors

SVD : Singular Value Decomposition

MSD : Mean Square Difference

LSA : Latent Semantic Analysis

LSI : Latent Semantic Indexing

PCA : Principal Component Analysis

LFM : Latent Factor Model

pLSA : probabilistic Latent Semantic Analysis

LDA : Latent Dirichlet Allocation

EM : Expectation Maximization

DAG : Directed Acyclic Graph

TABLE DES MATIERES

Remerciements	2
Résumé	3
abstract	5
Listes des figures	6
Liste des tableaux	10
Liste des abreviations	11
Introduction	16
PARTIE 1 Du Web 1.0 au Web 2.0	22
Chapitre I Web 2.0.....	22
I.1 Origine du Web 2.0	22
I.2 Les sept principes du Web 2.0	25
I.3 La technique principale du Web 2.0 : Ajax	28
I.4 Les applications et les services du Web 2.0 typiques	29
Chapitre II Réseau social numérique	43
II.1 Définition.....	43
II.2 Évolution des réseaux sociaux numériques	45
II.3 Typologie des réseaux sociaux numériques.....	47
II.4 RSN en Chine et en France.....	51
PARTIE 2 Le ciblage de l'individu personnalisation et filtrage collaboratif	56
Chapitre III Le filtrage d'information : le système de l'information personnalisable... ..	57
III.1 Définition du filtrage de l'information.....	57
III.2 Système de filtrage de l'information et système de recherche de l'information	58
III.3 Typologie de système de filtrage de l'information.....	60
Chapitre IV Système de recommandation personnalisé.....	65
IV.1 Définition et typologie du système de recommandation	65

IV.2 Approche de filtrage basé sur le contenu.....	67
IV.3 Approche de filtrage collaboratif.....	87
Partie 3 Les théories de l'identité sociale et de la désindividuation à l'épreuve des plateformes de cinéphilie	116
Chapitre V La théorie de l'identité sociale.....	116
V.1 L'étude des relations intergroupes	117
V.2 Les approches principales	119
V.3 Approche de l'identité sociale.....	121
Chapitre VI La théorie de la « désindividuation »	124
VI.1 La théorie de la désindividuation	124
VI.2 Le modèle SIDE.....	128
VI.3 La désindividuation dans l'environnement CMC	132
Chapitre VII Désindividuation et identité sociale à l'épreuve de quatre plateformes numériques de notations de films. Méthodologie de la recherche	134
VII.1 Les 4 plateforme pour les cinéphiles de notre recherche.....	135
VII.3 SR dans les plateformes pour les cinéphiles.....	152
VII.3 Les trois RSN et un non-RSN pour les cinéphiles.....	159
VII. 4 Allociné et Douban : 2 RSN pour les cinéphiles issus de 2 cultures différents	167
Partie 4 Synthèse des résultats, conclusion et perspectives	169
Chapitre VIII Enquête par questionnaire auprès des utilisateurs de Douban.....	171
VIII.1 Problèmes de recherche et les hypothèses.....	171
VIII.2 Méthode.....	175
VIII.3 Résultat	177
VIII.4 Conclusion.....	181
Chapitre IX Études comparatives.....	182
IX.1 Question de recherche et hypothèses	183
IX.2 Chiffre statistique	185

IX.3 Étude comparative: la comparaison entre Allociné et SensCritique	186
IX.4 Étude comparative entre RSN et non RSN : Douban et Maoyan	190
IX.5 Étude comparative entre deux RSN : Douban et Allociné.....	198
IX.6 Conclusion	203
Chapitre X Interprétation des résultats.....	205
X.1 Deux critères pour évaluer un produit culturel.....	205
X.2 L'influence normative et culturelle	207
X.3 Le RSN et le SR : désindividuation et personnalisation	208
Conclusion générale	210
Bibliographie.....	215
Annexes.....	240

INTRODUCTION

De nos jours, nous sommes de plus en plus envahis par des données et des informations instantanées et en continu. Suite au développement des techniques du web et des communications numériques depuis le début des années 1990, la surcharge informationnelle s'aggrave avec l'arrivée du Web 2.0. Les nouvelles plateformes du Web 2.0, telles que les réseaux sociaux numériques (RSN) et les médias sociaux, permettent aux individus de générer les contenus eux-mêmes. Une immense quantité d'informations est générée, communiquée et partagée à chaque instant sans arrêt. À chaque minute en 2019, on compte 18,1 millions messages envoyés par sms, 41,6 millions messages envoyés par Facebook Messenger et par WhatsApp, 3,8 millions requêtes de recherche sur Google, 188 millions Emails envoyés et 4,5 millions vidéo regardés sur YouTube *etc.*¹ D'ici à 2020, on estime que pour chaque habitant de la Terre, 1,7 Mo de données sera créée chaque seconde². Jamais autant d'informations n'avaient été mises à disposition au risque d'assiéger plus les individus plutôt que de les libérer.

Les systèmes de recommandation (SR) sont des moyens de filtrage l'information considérés comme efficaces et intéressants car ils permettent de fournir aux utilisateurs des informations personnalisées, soit celles qui correspondent le mieux à leurs préférences personnelles. Depuis la publication du premier article sur le filtrage collaboratif (FC) dans les années 1990 (Resnick *et al.*, 1994; Hill *et al.*, 1995; Shardanand et Maes, 1995), les SR ont connu un grand succès et attirent l'attention des chercheurs provenant de domaines variés. Plusieurs approches ont été proposées et de nombreux algorithmes ont été développés et améliorés en continu (Adomavicius et Tuzhilin, 2005; Su et Khoshgoftaar, 2009; Park *et al.*, 2012; Bobadilla *et al.*, 2013).

Les SR ont été appliqués avec succès aux utilisateurs des contenus aussi variés que : ceux de l'e-commerce (par exemple Amazon³), des films (par exemple, MovieLens⁴, IMDb⁵, Allociné⁶), de la vidéo à la demande (par exemple Netflix⁷), de la musique (par exemple

¹ Source : <https://www.visualcapitalist.com/what-happens-in-an-internet-minute-in-2019/>

² Source : https://www.domo.com/assets/downloads/18_domo_data-never-sleeps-6+verticals.pdf

³ <https://www.amazon.com/>

⁴ <https://movielens.org/>

⁵ <https://www.imdb.com/>

⁶ <http://www.allocine.fr/>

⁷ <https://www.netflix.com>

LastFM⁸, Spotify⁹), des vidéos (par exemple YouTube¹⁰), des livres (par exemple Goodreads¹¹, Babelio¹²), des articles et des ressources scientifiques (par exemple ResearchGate¹³, Academia¹⁴) et de bien d'autres.

Malgré leur popularité et leur réussite, les SR ont certains inconvénients dont la plupart se rattachent au problème du manque de données (Adomavicius et Tuzhilin, 2005). Pour les SR, le défi consiste à obtenir suffisamment de données pour déterminer ce qui est personnellement pertinent pour chaque utilisateur. Dans la pratique, les SR s'appuient sur les données personnelles des utilisateurs et surtout sur les données explicites (celles que l'on demande à l'utilisateur), sans lesquelles les recommandations seraient imprécises, voire impossibles. Cependant, les données personnelles des utilisateurs sont souvent insuffisantes pour le bon fonctionnement des SR (Sarwar *et al.*, 2000a) car les données explicites imposent un fardeau aux utilisateurs en leur demandant des efforts de déclaration de leur profil. Or, les utilisateurs ne veulent pas toujours fournir de telles informations (Amatriain, Pujol et Oliver, 2009; Jannach *et al.*, 2010).

Le SR est particulièrement vulnérable à ce type de problème de « sous contribution » (*under-contribution*) (Butler, 2001; Harper *et al.*, 2005) qui pourrait conduire au problème du démarrage à froid (*cold-start problem*) (Schein *et al.*, 2002; Lam *et al.*, 2008) ou le problème de sérendipité (*sparsity problem*) (Billsus et Pazzani, 1998; Good *et al.*, 1999), car il n'y aurait pas assez de données explicites disponibles (e.g, les notations, les commentaires). Par conséquent, les SR exploitent souvent d'autres types d'informations, telles que les caractéristiques démographiques des utilisateurs, provenant des médias sociaux dans le but d'améliorer la précision des recommandations (Bobadilla *et al.*, 2013; Yang *et al.*, 2014; Bernardes *et al.*, 2015).

En revanche, les médias sociaux¹⁵ d'aujourd'hui disposent d'une immense quantité

⁸ <https://www.last.fm>

⁹ <https://www.spotify.com>

¹⁰ <https://www.youtube.com>

¹¹ <https://www.goodreads.com/>

¹² <https://www.babelio.com/>

¹³ <https://www.researchgate.net/>

¹⁴ <https://www.academia.edu/>

¹⁵ Le terme de médias sociaux désigne les plateformes numériques accessibles par Internet (web et applications) et permettant à leurs membres d'établir ou d'intégrer des réseaux d'amis ou connaissances professionnelles et de participer à la vie de ces réseaux à travers la mise à disposition d'outils et interfaces de présentation, de communication et d'interaction

d'informations générées par les utilisateurs. En effet, chaque minute en 2019, plus de 474,000 tweets sont générés, 300 heures de vidéo sont téléchargées sur YouTube et les utilisateurs regardent 4,333,560 vidéos, 69,444 millions de postes sont créés sur Instagram. Sur Facebook, 510,000 commentaires sont publiés, 293,000 statuts sont mis à jour et 136 000 photos sont téléchargées chaque minute *etc.*¹⁶ Les SR bénéficient de nouveaux types de données et de métadonnées publiques introduites par les médias sociaux, telles que les hashtags, les notes, les commentaires et les relations interpersonnelles explicites, qui peuvent être utilisés pour améliorer les recommandations (Bobadilla *et al.*, 2013; Jannach et Friedrich, 2013; Yang *et al.*, 2014; Bernardes *et al.*, 2015; Aggarwal, 2016). En revanche, les médias sociaux ont besoin des techniques de recommandation personnalisée pour bien organiser les informations et proposer à chaque utilisateur les contenus les plus pertinents et les plus intéressants.

Il n'est pas surprenant que les SR soient largement appliqués dans les réseaux sociaux numériques (RSN), et surtout dans les RSN de contenu (qui partagent des critiques de produits culturels, tels que Goodreads, Babelio, SensCritique¹⁷, Allociné, IMDb, Douban¹⁸ *etc.*). Ce type de RSN propose de découvrir, de noter et d'écrire des critiques des produits culturels appartenant aux domaines des films, séries télévisées, livres, bandes dessinées, albums et morceaux musicaux, et jeux vidéo. Les SR s'intègrent au RSN pour aider les utilisateurs à découvrir des produits qui correspondent le mieux à leurs préférences personnelles (Resnick et Varian, 1997; Burke, 2007; Ricci *et al.*, 2011). Néanmoins, la crédibilité et la fiabilité des évaluations fournies par les utilisateurs dans les RSN restent encore douteuses. En effet, le SR dans le RSN a deux faces complètement contradictoires : la personnalisation et la désindividuation.

D'une part, l'objectif principal du SR est de fournir à chaque utilisateur des recommandations personnalisées qui correspondent à ses préférences personnelles. Ces recommandations personnalisées sont basées sur des rétroactions implicites (les comportements de l'utilisateur dans le passé) ou sur des rétroactions explicites (les notes et les commentaires précédents faites par l'utilisateur) (Bobadilla *et al.*, 2013). Dans la pratique, la rétroaction explicite est notamment exploitée pour apprendre les préférences de l'utilisateur (Amatriain, Pujol et Oliver, 2009) car elle est considérée plus fiable que la rétroaction implicite. Parmi toutes les méthodes existant pour recueillir les rétroactions explicites, la plus précise et efficace

¹⁶ Source : <https://www.visualcapitalist.com/what-happens-in-an-internet-minute-in-2019/>

¹⁷ <https://www.senscritique.com/>

¹⁸ <https://www.douban.com/>

est celle qui demande directement à l'utilisateur de noter un item. La note se situe souvent sur une échelle de 1 à 5 étoiles qui précisent à quel niveau l'utilisateur aime un item (Jannach *et al.*, 2010; Aggarwal, 2016). L'hypothèse fondamentale du SR est que les évaluations des films faites par les utilisateurs reflètent leurs préférences : les utilisateurs attribuent des notes élevées aux films qu'ils aiment et des notes faibles aux films qu'ils n'aiment pas. Les systèmes utilisent des notes allant de 1 étoile (je n'aime pas du tout) à 5 étoiles (j'aime beaucoup) qui sont ensuite transformées en valeurs numériques afin de pouvoir être utilisées par les algorithmes de recommandation (Sparling et Sen, 2011)¹⁹. Il est donc essentiel que le SR obtienne des rétroactions de l'utilisateur qui ne sont pas influencés par d'autrui et qui reflètent ses opinions personnelles sur des items.

D'autre part, cependant, les RSN sont pleins d'interactions entre les utilisateurs qui sont susceptibles de les désindividualiser et d'influencer leurs comportements et leurs pensées. En effet, selon le « modèle d'identité sociale des effets de désindividuation (modèle SIDE, *the Social Identity model of Deindividuation Effects*) (Reicher, Spears et Postmes, 1995), l'immersion dans le groupe social et l'absence d'indices de personnalisation peuvent renforcer l'identité sociale aux dépens de la personnalisation. En outre, l'anonymat dans ces RSN améliorerait le respect des normes du groupes²⁰ par les utilisateurs lorsque ceux-ci avaient une identité sociale (du groupe) saillante ou une identité personnelle moins saillante (Spears, Lea et Lee, 1990). Une série d'études a montré que l'anonymat dans la communication virtuelle favorise et renforce l'influence des facteurs sociaux au sein du groupe et que l'interaction de la CMC (*Computer Mediated Communication*) est fortement influencée par les normes du groupe dans un média où l'utilisateur est relativement anonyme ou désindividualisé (Spears, Lea et Lee, 1990; Lea et Spears, 1991; Postmes, Spears et Lea, 1999; Postmes *et al.*, 2001a).

Dans cette recherche, nous nous consacrons aux SR du domaine culturelle, spécifique de cinéma. MovieLens et Netflix sont des applications typiques qui visent à fournir des services de recommandations personnalisées du film sans utiliser de lien social. En apprenant les historiques de notation de l'utilisateur, ils peuvent inférer ses préférences et lui recommander des films. Dans ces applications, les utilisateurs sont indépendants et ne sont liés de façon implicite les uns aux autres que par les produits qu'ils ont évalués en commun (Jannach *et al.*, 2010). Par conséquent, les utilisateurs ne sont guère influencés par les opinions des autres

¹⁹ En 2017, Netflix a abandonné son système de notation par étoiles au profit d'un système binaire « thumb-up » ou « thumb-down » qui permettra d'indiquer si l'on a apprécié ou non un film ou un épisode de série.

²⁰ Les normes du groupe sont

utilisateurs et évaluent les films de manière subjective en ne se focalisant que sur leurs propres préférences. Dans les RSN pour les cinéphiles tels que IMDb, Allociné, SensCritique et Douban *etc.*, au contraire, les utilisateurs peuvent non seulement exprimer leurs opinions en notant des films et en publiant leurs commentaires, mais aussi en communiquant avec les autres utilisateurs. Les interactions entre les utilisateurs pourraient conduire à l'établissement et au renforcement de certaines normes du groupe²¹ (Tom Postmes, Spears et Lea, 2000; Hogg et Reid, 2006). Cela implique que les utilisateurs de ces RSN seraient fortement influencés par les normes du groupe établies dans ces plateformes et auraient tendance à exprimer les opinions conformes à ces normes plutôt que ce qu'ils pensent subjectivement. Il est probable que les évaluations faites par les utilisateurs ne reflèteraient donc pas toujours les préférences réelles, mais seraient largement influencées par les normes du groupe auxquels ils sont affiliés. Dans ce cas, un algorithme de recommandation pourrait ne pas être assez précis. Par conséquent, des études sont nécessaires pour comprendre l'impact de la norme du groupe sur les comportements de notations des utilisateurs afin d'améliorer la précision des SR.

Dans cette recherche, nous cherchons à vérifier si ces deux faces contradictoires existent vraiment pour les SR dans les RSN et si l'effet de désindividuation serait un obstacle pour les SR qui cherchent à fournir des recommandations personnalisées. La question de recherche fondamentale que nous traitons à travers cette thèse est celle de savoir si les RSN disposent de certaines normes du groupe qui sont assez puissantes pour influencer les utilisateurs lorsqu'ils évaluent les films, ce qui les empêcherait d'exprimer leurs opinions personnelles. Dans le cadre de cette thèse, nous cherchons à savoir si l'influence sociale et la norme du groupe ont un impact sur les comportements de notation des utilisateurs dans les RSN pour les cinéphiles. Puis nous cherchons à identifier les facteurs clés qui constituent les évaluations des utilisateurs.

La thèse s'organise comme suit. Elle est composée de 4 parties.

La Partie 1 « du web 1.0 au web 2.0 » : présente l'état de l'art du développement du Web 2.0 et du RSN. Les chapitres 1 et 2 présentent respectivement le développement, les principales

²¹ « Le terme "standard du groupe" ou "norme du groupe" a été librement utilisé, soit pour expliquer, soit pour décrire un phénomène maintes fois observé : le fait que les membres d'un même groupe à relations directes présentent une relative uniformité d'opinions et de conduite en certains domaines. L'emploi du terme, pour expliquer comme pour décrire, signifiait généralement que cette uniformité observée était, en quelque sorte, le résultat des influences que le groupe est capable d'exercer sur ses membres. Le fait que les membres d'un même groupe social ont tous à peu près les mêmes goûts dans le choix de leurs distractions par exemple, a souvent été expliqué par l'existence d'influence de groupes ou interpersonnelles plutôt que par des circonstances similaires produisant des réactions similaires chez un certain nombre de personnes. » (Festinger, Schachter et Back, 1964)

technologies et applications du Web 2.0 et RSN. Le Web 2.0 et les réseaux sociaux constituent le contexte général de cette thèse. Leur développement rapide a conduit au problème de la surcharge d'informations, qui est à l'origine de l'émergence et du développement du système de filtrage de l'information (SFI) et du SR qui seront présentés dans la deuxième partie.

La Partie 2 « Le ciblage de l'individu personnalisation et filtrage collaboratif » : est consacrée à la présentation des concepts de base de technologie de filtrage de l'information et de système de recommandation personnalisée. Chapitre III et IV présentent respectivement les définitions, les typologies et les techniques principales de SFI et SR. En présentant le fonctionnement du SR, nous révélons le fait qu'il repose largement sur les rétroactions des utilisateurs qui reflètent véritablement et précisément leurs besoins et leurs préférences.

La Partie 3 « Les théories de l'identité sociale et de la désindividuation à l'épreuve des plateformes de cinéphilie » : examine les théories de la désindividuation qui cadrent cette thèse. Chapitre V présente la théorie de l'identité sociale et les approches principales. Chapitre VI est consacré aux théories de la désindividuation, les approches principales et le modèle de l'identité sociale des effets de la désindividuation. Ces théories impliquent que les utilisateurs de RSN sont très susceptibles de se désindividualiser et que cette désindividuation entraîne des rétroactions d'information biaisés qui ne reflète pas fidèlement leurs opinions et leur préférence, ce qui affecte le fonctionnement des systèmes de recommandation personnalisée. Dans le Chapitre VII, nous présentons les 4 plateformes : Douban, Allociné, Maoyan et SensCritique qui concernent notre recherche. Parmi eux, Allociné et SensCritique sont deux RSN pour les cinéphiles français alors que Douban est un RSN pour les cinéphiles chinois. Maoyan est une plateforme pour les cinéphiles chinois qui n'est pas un RSN. Nous l'avons introduite pour comparer le comportement de ses utilisateurs en matière de notation avec celui des utilisateurs de RSN.

La Partie 4 « Synthèse des résultats, conclusion et perspectives » : expose les résultats de enquêtes portant sur Douban, Allociné, Maoyan et SensCritique. Ces résultats sont ensuite discutés.

Dans la dernière partie, nous présentons une synthèse des résultats que nous avons obtenus, un résumé sur nos différentes contributions et leurs extensions possibles.

PARTIE 1 DU WEB 1.0 AU WEB 2.0

Dans cette partie, nous présentons l'état de l'art du développement du Web 2.0 et du RSN qui constituent le contexte général de cette thèse. Leur développement rapide a conduit au problème de la surcharge d'informations, qui est à l'origine de l'émergence et du développement du SF et du SR. Les chapitres 1 présente les définitions, le développement, les principales technologies et applications du Web 2.0. Dans le chapitre 2, nous présentons les définition, le développement de RSN ainsi que l'état de l'art des RSN en Chine et en France.

Chapitre I Web 2.0

I.1 Origine du Web 2.0

Inventé par Tim Berner-Lee en 1989, le World Wide Web (abrégé www ou le Web) est un système hypertexte public fonctionnant sur Internet qui permet aux utilisateurs de consulter, avec un navigateur spécifique, des pages web accessibles sur des sites en ligne. Sur le Web, les ressources Internet peuvent être exprimées de manière plus intuitive dans une page Web. Ces ressources peuvent être liées les unes aux autres sur la page Web. Le Web n'est qu'une des applications d'Internet, distincte d'autres applications comme le courrier électronique, la messagerie instantanée et Usenet, *etc.*

La première étape de l'évolution du Web est appelée rétrospectivement le Web 1.0 sur lequel « les créateurs de contenu étaient peu nombreux » et « la majorité des utilisateurs étaient simplement en tant que consommateurs de contenus » (Cormode et Krishnamurthy, 2008). La plupart des sites web des années 90 avaient été initialement construits avec des pages HTML statiques et quelques styles simples intégrés dans le balisage HTML. Le web ne comprenait que des pages statiques reliées entre elles par des liens hypertextes rarement mises à jour. À la fin des années 1990 et au début des années 2000, les fonctionnalités du site Web interactif ont redéfini ce qui pouvait être accompli dans un navigateur web et ont marqué un point majeur d'évolution dans le monde du développement de web.

L'expression et le concept du Web 2.0 ont été mentionnés pour la première fois par Darcy Di Nucci dans son article « *Fragmented Future* » (DiNucci, 1999). Tim O'Reilly et Dale Dougherty les ont rendus populaires lors d'une conférence brainstorming entre la société O'Reilly Média et la société MediaLive International (O'Reilly, 2009). Depuis lors, le concept du Web 2.0 s'est répandu à un rythme incroyable dans le monde entier.

Comme Tim O'Reilly souligne que «le Web 2.0 n'a pas de frontière claire» (O'Reilly, 2009), la notion du Web 2.0 reste très floue et difficile à définir car : 1) il n'y a pas de technologies nouvelles et révolutionnaires qui rendent les applications ou les services Web 2.0 disponibles, 2) la compréhension du Web 2.0 est très variée, et 3) le Web 2.0 est un vaste sujet qui englobe de nombreux concepts et méthodologies concernées (Kim *et al.*, 2009).

Effectivement, il existe de nombreuses définitions et les avis divergents. Selon Tim O'Reilly, le Web 2.0 est le réseau en tant que plateforme, utilisant tous les moyens de connexion. Les applications du Web 2.0 sont celles qui tirent les avantages intrinsèques les plus élevés de cette plateforme. Il s'agit de fournir un logiciel qui offre des services continuellement améliorés, dont la qualité croît d'autant plus qu'un plus grand nombre d'utilisateurs y ont recours. (O'Reilly, 2009). Pour lui, il concerne en particulier les interfaces et les échanges permettant aux internautes ayant peu de connaissances techniques de s'approprier de nouvelles fonctionnalités du web. Les internautes peuvent d'une part contribuer à l'échange d'informations et interagir (partager, échanger, *etc.*) de façon simple, à la fois au niveau du contenu et de la structure des pages, et d'autre part, établir des connexions avec les autres et finalement créer des « réseaux sociaux » sur le web. Les applications typiques du Web 2.0 sont Wikipédia, Facebook, YouTube, Flickr, Twitter qui encouragent et facilitent la participation et la collaboration. Sans surprise, l'une des caractéristiques principales qu'ils partagent en commun est qu'ils se présentent sous la forme de réseaux sociaux. Les utilisateurs de ces applications du Web 2.0 peuvent définir les autres comme leurs amis avec lesquelles ils peuvent partager et communiquer. L'internaute devient, grâce aux outils mis à sa disposition, une personne active sur le web (O'Reilly, 2007).

Tenenbaum et Hughes proposent une description générale du Web 2.0 en tant qu'un ensemble de technologies et de méthodologies du web qui rendent le web plus participatif, plus sémantique et plus en temps réel. Le plus important est que le Web 2.0 est un phénomène culturel. Les développeurs partent d'une idée simple mais utile et la sortent rapidement, afin que d'autres puissent l'affiner et l'embellir. Le processus est connu sous le nom de collaboration de

masse - des milliers de personnes s'appuient progressivement sur le travail des autres. (Tenenbaum et Hughes, 2006)

Funk décrit le Web 2.0 comme « une transformation sociale qui a mis plus d'interactivité et de contrôle du contenu entre les mains des utilisateurs réguliers, pas seulement des grands propriétaires de sites » (Funk, 2009).

Pierre Kosciusko-Morizet déclare que le Web 2.0, est le web de chacun, le web de tous, pour tous et par tous. C'est le « vrai » Internet qu'on attendait depuis le début, sans savoir quelle forme il aurait, celui qui commence à utiliser le potentiel unique de la toile, c'est-à-dire le potentiel de mise en relation (Fayon, 2010).

Wilson, Lin, Longstreet et Sarker définissent le Web 2.0 en tant que la deuxième génération du web où les applications et services du web interopérables et centrés sur l'utilisateur favorisent la connectivité sociale, le partage des médias et des informations, le contenu créé par l'utilisateur et la collaboration entre les individus et les organisations (Wilson *et al.*, 2011).

D'après Weber et Rech, « les services du Web 2.0 sont centrés sur l'utilisateur, le partage de contenu (téléchargement et modification), les réseaux sociaux ou les services de mashup statiques basés sur des techniques prenant en charge des micropages dynamiques qui exploitent l'intelligence collective. Ils peuvent prendre en charge une interface de programmation ouverte avec des données fermées et une identité fermée afin d'utiliser le web comme système de fichiers distribué (contenu généré par l'utilisateur) ou système de collaboration (effets de mise en réseau)» (Weber et Rech, 2009).

En résumé, les définitions autour du Web 2.0 visent à différencier ce qui se positionne comme un World Wide Web de « deuxième génération » de son prédécesseur, le Web 1.0. Le Web 2.0 ne peut pas être résumé à une technologie ou une technique, c'est un ensemble des techniques, des fonctionnalités et des usages qui suivent une évolution logique du Web 1.0 vers l'interactivité (sociabilité), le partage et une simplicité d'utilisation ne nécessitant pas des connaissances techniques et informatiques pour les internautes.

I.2 Les sept principes du Web 2.0

Tim O'Reilly souligne que « comme de nombreux concepts majeurs, le Web 2.0 n'a pas de frontière claire. Vous pouvez voir le web 2.0 comme un ensemble de principes et de pratiques qui, à la manière d'un système planétaire, verrait des sites exposant tout ou partie de ses préceptes graviter à des distances variables du centre en question » (O'Reilly, 2009). Ayant la difficulté de définir le terme, Tim O'Reilly et ses collègues résument et suggèrent les 7 grandes principales du Web 2.0.

Le web en tant que plateforme

Le premier principe du Web 2.0 est que le web est une plateforme²². Comme Macmanus déclare que « pour les entreprises, le web est une plateforme pour les entreprises. Pour les marketeurs, le web est une plateforme de communication. Pour les journalistes, le web est une plateforme pour les nouveaux médias. Pour les geeks, le web est une plateforme de développement logiciel » (Macmanus, 2005). Les utilisateurs ne sont plus que les lecteurs qui consultent les contenus gérés uniquement par les éditeurs du site web. Au contraire, ils occupent une position centrale dans le site web et ont plus de pouvoir sur la production et la distribution de l'information. Les contenus qu'ils téléchargent seront acceptés et gérés de manière efficace par le site web. Ces contenus, tels que des informations et des divertissements basés sur divers médias sont l'essence du service.

Tirer parti de l'intelligence collective

Pour O'Reilly, c'est le principe central du succès des géants nés du web 1.0 qui ont survécu pour maintenant mener l'ère du web 2.0. Le développement de technique facilite l'interactivité et la collaboration entre les utilisateurs qui permettent aux sites du Web 2.0

²² Une plateforme est un service occupant une fonction d'intermédiaire dans l'accès aux informations, contenus, services ou biens édités ou fournis par des tiers. Au-delà de sa seule interface technique, elle organise et hiérarchise les contenus en vue de leur présentation et leur mise en relation aux utilisateurs finaux.

d'exploiter efficacement la force de l'intelligence collective que recèle potentiellement le web. Dans cette nouvelle vision, l'implication des utilisateurs dans le réseau devient le facteur-clé pour la suprématie du marché (O'Reilly, 2009).

La puissance est dans les données

O'Reilly déclare que « la gestion de base de données est le cœur de métier des sociétés du web 2,0 » (O'Reilly, 2009). Les entreprises ont déjà fait beaucoup attention sur la possession et la gestion de base de données même si avant l'arrivée du Web 2.0. Effectivement la maîtrise de base de données a conduit à la domination du marché et à d'énormes bénéfices économiques. À l'époque du Web 2.0, les nouvelles techniques simplifier l'utilisation de web et facilitent aux utilisateurs à créer et à gestion des contenus. Les utilisateurs deviennent à la fois les consommateurs et les fournisseurs des données. Les systèmes du Web 2.0 sont conçus pour encourager la participation d'utilisateurs.

La fin des cycles des releases

O'Reilly remarque deux changements fondamentaux dans le modèle d'affaire des sociétés de logiciels en entrant l'époque du Web 2.0: les traitements deviennent le cœur de métier et les utilisateurs doivent être traités comme des co-développeurs, en référence aux pratiques de l'open-source (O'Reilly, 2009). Les applications du Web 2.0 offrent aux utilisateurs des services au lieu des logiciels classiques qui sont sorties régulièrement pendant une période assez longue (toute les deux ou trois ans, les logiciels de Microsoft par exemple). Effectivement, les logiciels du Web 2.0 sont mise à jour très rapidement. Les sites web observent les comportements des utilisateurs en temps réel pour comprendre leurs préférences et besoins. Cela permet les sites web 2.0 de développer et de mettre à jour leurs produits avec de nouvelles fonctionnalités apparaissant de manière mensuelle, hebdomadaire voire quotidienne. Les logiciels du Web 2.0 ne sortent plus de nouvelle version mais une longue période de test, dite « la bêta perpétuelle », dans laquelle l'efficacité de nouvelles fonctionnalités peuvent être évaluées en temps réel.

Des modèles de programmation légers

À l'époque du web 2.0, le pragmatisme est préféré qu'une conception idéale. Par rapport aux technologies lourdes telles que les web services, SOAP, la mise en place de modèles de programmations légers permet la création de systèmes faiblement couplés dont l'intérêt repose sur la rapidité et la simplicité d'utilisation (Chaimbault, 2007). O'Reilly remarque que c'est la simplicité apparente des nouveaux services qui fait leur popularité et leur succès. L'exemple est le RSS (sigle venant de l'anglais « *Really Simple Syndication* ») qui devient rapidement l'un des services web les plus déployés du fait de sa simplicité, alors que la complexité des web services des grandes sociétés condamna ceux-ci à un déploiement limité (O'Reilly, 2009). Les applications du Web 2.0 sont « bidouillable » et « remaniable », c'est-à-dire que les applications doivent être programmées pour qu'ils puissent être réutilisés de manière facile. Cela permet un modèle de programmation léger.

Le logiciel se libère du PC

Les logiciels se libéreront du PC (*Personal Computer*) qui était un espace incontournable pour utiliser des logiciels. Les applications du Web 2.0 offrent des services de gestion de données qui « tirent parti de la puissance de la plateforme web, en faisant une partie permanente et pratiquement invisible de leurs infrastructures » (O'Reilly, 2009). Effectivement, de plus en plus d'appareils (téléphone mobile, tablette, liseuse, montre *etc.*) sont connectés à l'application internet que constitue le web. Les utilisateurs peuvent non seulement consulter les contenus via ces appareils variés mais également émettre des données en temps réel. Le web est maintenant partout.

Enrichir les interfaces utilisateur

Le dernier principe est « l'enrichissement des interfaces utilisateurs ». En effet, depuis l'aube des navigateurs et la possibilité d'y délivrer des « applets », les développeurs cherchent toujours à apporter de manière plus légère intelligence côté client et richesse d'interface. Le

langage Ajax se voit être un élément clé des applications web 2.0 (par exemple., Flickr et Gmail), grâce auquel les développeurs sont capables de réaliser des applications web aussi riches que les logiciels locaux classiques.

I.3 La technique principale du Web 2.0 : Ajax

« Le web en tant que plateforme » est le principe fondamental qui est réalisable grâce à l'émergence d'une nouvelle génération de technologies et de normes liées au web. Dans l'époque du Web 1.0, les applications logicielles s'exécutaient sur la machine de l'utilisateur, gérées par un système d'exploitation de bureau tel que MacOS, Windows ou Linux, alors que les applications du Web 2.0 sont sous la forme d'une plateforme web dont les services logiciels sont exécutés dans la fenêtre réelle du navigateur, communiquant avec le réseau et serveurs distants. Ce changement est ce que O'Reilly appelle « le logiciel se libère du PC ».

L'une des conséquences du « web en tant que plateforme » est que l'accent est mis notamment sur les applications fournissant des services au lieu sur les logiciels en paquet. Il exige que les applications soient mises à jour régulièrement et rapidement. En fait, de nombreux services Web 2.0 bien connus restent dans une version de « beta perpétuelle ». C'est « la fin des cycles de release » qualifié par O'Reilly.

Ajax est l'une des nouvelles technologies la plus utilisée, mais il existe des alternatives qui sont principalement basées sur la technologie Flash. L'architecture informatique JavaScript et XML asynchrones (Asynchronous JavaScript et XML) dont l'acronyme est Ajax, est un ensemble de technologies de développement du web permet de construire des applications et des sites web dynamiques interactifs sur le poste client en se servant de différentes technologies ajoutées aux navigateurs web entre 1995 et 2005.

Le terme Ajax a été introduit par Jesse James Garrett (informaticien américain), le 18 février 2005, dans son article « *Ajax : A New Approach to Web Applications* (Ajax : une nouvelle approche des applications web) » (Garrett, 2005) basé sur les techniques utilisées sur les pages Google. Depuis, Ajax a rapidement gagné en popularité et a été largement utilisé dans le développement d'applications web. Google investit massivement dans le développement d'Ajax. Il déploie à grande échelle Ajax et réussit à l'appliquer à des applications web telles que Gmail

(2004) et Google Maps (2005).

Ajax est destiné à réaliser des mises à jour rapides du contenu d'une page web, sans qu'elles nécessitent le moindre rechargement visible par l'utilisateur de la page web. Avec Ajax, les applications web peuvent envoyer et récupérer des données à partir d'un serveur de manière asynchrone sans interférer avec l'affichage et le comportement de la page existante. En découplant la couche d'échange de données de la couche de présentation, Ajax permet aux pages web et, par extension, aux applications web, de changer le contenu de manière dynamique sans avoir à recharger la page entière.

Ajax n'est pas une technologie nouvelle ou un langage différent, mais un ensemble de techniques existantes utilisées de manière nouvelle. Des applications web asynchrones emploient de nombreuses techniques du web existantes, mais d'une manière générale, le JavaScript est constamment présent. D'autres langages sont pris en compte tel que le HTML et le CSS, qui servent à l'affichage. La page web peut ensuite être modifiée par JavaScript pour afficher dynamiquement et permettre à l'utilisateur d'interagir avec les nouvelles informations. Les requêtes du type « XMLHttpRequest » sont couramment utilisées pour exécuter Ajax sur des pages web permettant aux sites de charger du contenu sur l'écran sans rafraîchir la page. Le transfert de données est géré exclusivement par le JavaScript et utilise certaines technologies de formatage de données, comme le XML ou le JSON. L'utilisation d'Ajax fonctionne sur tous les navigateurs web courants : Google Chrome, Safari, Mozilla Firefox, Internet Explorer, Microsoft Edge, Opera, *etc.*

I.4 Les applications et les services du Web 2.0 typiques

Il existe de nombreux de services et applications web qui illustrent le concept fondamental du Web 2.0: en matière de publication et d'agrégation personnelles, il existe le Blog et le RSS ; en collaboration sociale, le wiki ; en matière de participation des utilisateurs, le commentaire en ligne, le "tag" (indexation collaborative), le réseau social numérique, le social bookmarking et le service de partage de multimédia *etc.* Dans la section suivante, nous présentons ces applications et services couramment utilisés en vue de fournir une base commune pour une discussion ultérieure.

I.4.1 Blog et microblog

Le blog est l'un des succès majeurs du web 2.0 qui est une forme de publication, de communication et d'expression en ligne. Grâce à une interface simplifiée, le blog est facile à utiliser et a gagné en popularité depuis son émergence à la fin des années 1990 (Blood, 2002).

Le blog moderne a évolué à partir du journal en ligne personnel. Il est généralement reconnu que le premier blog était Links.net, créé par Justin Hall, alors qu'il était étudiant au *Swarthmore College* en 1994. Bien sûr, à cette époque, le terme « blog » n'était pas encore utilisé, et il l'avait simplement intitulé « page d'accueil personnelle ». C'est en 1997 que Jorn Barger a proposé pour la première fois le terme « *weblog* » qui fait référence à une simple page web composée de brefs paragraphes d'opinion, d'informations, de journaux personnels ou de liens, appelés messages, classés par ordre chronologique à partir du plus récent en premier, dans le style d'une revue en ligne (Blood, 2002; Rosenberg, 2009). En 1999, Peter Merholz a utilisé l'abréviation « blog » pour la première fois : ce terme a progressivement remplacé celui de « *weblog* » et est devenu un terme courant aujourd'hui. En 1998, l'éditeur d'Infosift²³ Jesse Garrett a collecté des sites web de type blog et les a envoyés à Cameron Barrett qui a affiché la liste sur le site web de Camworld²⁴. De nombreux utilisateurs du site web ont ensuite envoyé l'URL du blog à Barrett. Une nouvelle communauté en ligne prenait forme. En 1999, Brigitte Eaton a créé un annuaire de blogs « *Eatonweb Portal*²⁵ » pour collecter les sites de blogs qu'elle connaissait. Le mois de Juin 1999 a marqué un tournant dans le développement rapide du blog. À cette époque, Pitas a commencé à fournir des services de blog gratuits. En août 1999, Pyra Lab a lancé « *Blogger*²⁶ » (Blood, 2000). Blogger fournit des instructions faciles à apprendre et la possibilité de publier blog directement sur des sites web personnels, ce qui apporte une grande commodité aux utilisateurs et montre initialement la « facilité » de la technique et des applications web2.0.

Les premiers microblogs sont nommés « *tumblelogs* » le 12 avril 2005 par l'auteur *Why the lucky stiff*, un écrivain prolifique ayant participé au langage de programmation Ruby, lorsqu'il décrivait Anarchaia de Christian Neukirchen. Jason Kottke réutilisait ce terme le 19

²³ <http://www.jjg.net/retired/infosift/>

²⁴ <http://camworld.org/>

²⁵ <http://portal.eatonweb.com/>

²⁶ <https://www.blogger.com/>

octobre 2005. Ce type de blog a ensuite été appelé « microblog » vers 2006. Il devint très populaire avec le service Twitter, créant une forme extrême de microblog en limitant les publications à 140 caractères de texte, un format proche de celui des SMS en téléphonie mobile. Ce service supporte tout de même les liens hypertexte. Tumblr, né en 2007, permet à l'utilisateur de poster du texte, des images, des vidéos, des liens et des sons.

Il existe de nombreux types de blogs qui diffèrent par la manière dont le contenu est diffusé ou écrit : le blog personnel, le blog collaboratif, le blog d'entreprise et d'organisation *etc.* Le blog est utilisé dans les domaines très variés : le journalisme, la commerce, la pédagogie, la littérature *etc.*

Outre des blogs de texte, il existe également autres types de blog pour diffuser des fichiers (audio, image, vidéo ou autres), soit le podcasting. La plus grande différence entre le podcasting et la radio en ligne est que le podcasting utilise le mode d'abonnement RSS 2.0. Les individus peuvent diffuser facilement leurs propres programmes. Comparé avec la radio ou avec la télévision traditionnelle, le podcasting ne nécessite pas de canal ni d'un grand nombre d'appareils spécialisés. Les caméras vidéo numériques, les logiciels d'enregistrement, les microphones et autres appareils couramment utilisés à la maison permettent aux individus de devenir des « annonceurs » ou des « animateurs ». Par rapport aux médias de masse traditionnels, la production de contenu et la diffusion des podcastings sont très personnalisées.

Le blog se base beaucoup sur la technique RSS qui sera décrite dans la section suivante. Les lecteurs peuvent s'abonner à un blog via un flux RSS afin de recevoir des notifications lorsque le blog est mis à jour. Les auteurs de blog utilisent RSS pour rendre leurs articles publiés faciles à diffuser. Du point de vue des modèles de communication, blogosphère établit une relation interactive positive entre les lecteurs et les auteurs.

À l'époque du Web 2.0, le blog donne aux internautes la compétence et la possibilité de produire du contenu. Cela satisfait largement le désir des individus de publier et d'exprimer leurs opinions.

I.4.2 RSS

Le blog facilite la publication des contenus personnels et la technique RSS permet aux

internauts d'obtenir des informations de manière très simple. RSS est une famille de formats de données utilisés pour la syndication de contenu web.

Le RSS est une ressource des sites web dont le contenu est produit automatiquement en fonction des mises à jour des sites web. Il est surtout utilisé pour regrouper des sites web qui publient fréquemment des contenus et qui sont régulièrement mis à jour, telles que les flux web des articles de blog, des actualités, de l'audio ou de la vidéo. Les flux RSS sont des fichiers XML qui sont souvent utilisés pour présenter le titre et un résumé ou le contenu intégral des dernières informations consultables.

Il existe trois formats de données qui peuvent être désignés par ces initiales : **RDF (Resource Description Framework) Site Summary** (RSS 0.90 et 1.0) sorti en 1999 ; **Rich Site Summary** (RSS 0.91) sorti en 1999 ; **Really Simple Syndication** (RSS 2.0) sorti en 2002.

RDF Site Summary (RSS 0.90) est la première version de RSS créée par Netscape en mars 1997. Quatre mois plus tard, une autre version RSS 0.91 était sortie en juillet 1999. RSS 0.91 simplifie le format en supprimant les éléments RDF et en incorporant des éléments du format de syndication d'actualités de Dave Winer. Cette version est également connue sous le nom de Rich Site Summary. Comme la version précédente, RSS 0.91 est également créée par Netscape ayant l'objectif de permettre à leurs clients de configurer « *MyNetscape*²⁷ », c'est-à-dire, de gérer les nouvelles sur leur page d'accueil.

Dans les années suivantes, plusieurs versions différentes de RSS sont sorties (voir Tableau 1) en tombant dans deux branches principales : RDF et Really Simple Syndication dont les versions initiales de ces deux branches étaient respectivement RSS 0.90 et RSS 0.91. Dans la famille de RDF, le groupe de travail RSS-DEV a réalisé une version nommée RSS 1.0 se basant sur RDF mais incompatible avec RSS 0.92. De l'autre côté, Dave Winer publie une version modifiée de la RSS 0.91 sur le site web « *UserLand* » en 2000. Six mois plus tard, une extension du format RSS 0.91 est sortie appelée RSS 0.92. En 2002, RSS 2.0 est publié que Dave Winer nomme « Really Simple Syndication »²⁸.

²⁷ <http://my.netscape.com/>

²⁸ <http://www.rssboard.org/rss-history>

Formats de syndication			
Date	Format	Version	Créateurs
15/03/1999	RDF Site Summary	0.90	Netscape (Ramanathan V. Guha)
10/07/1999	Rich Site Summary	0.91	Netscape (Dan Libby)
09/06/2000	Really Simple Syndication	0.91	UserLand (Dave Winer)
25/12/2000	Really Simple Syndication	0.92	UserLand (Dave Winer)
06/12/2000	RDF Site Summary	1.0	RSS-DEV Working Group
19/08/2002	Really Simple Syndication	2.0	UserLand (Dave Winer)

Tableau 1 Formats et versions de RSS (Courrier, 2007)

Depuis la naissance du blog, un très grand nombre de contenus ont été créés par les internautes. Par rapport aux navigateurs traditionnels, aux portails web et même aux moteurs de recherche, RSS montre un grand avantage dans l'acquisition, l'organisation et la gestion de l'information énormes. RSS est largement utilisé dans les blogs, les wikis et les sites d'actualité.

Bien que RSS soit un moyen utile de collecter des informations, les informations en ligne dont la quantité s'accroît sans cesse restent inépuisables. Au fur et à mesure qu'un individu s'abonne à plus en plus de flux RSS, il retombe dans une situation de confrontation d'une quantité énorme d'informations que nous ne pouvons pas traiter. Il est nécessaire que l'on développe des outils permettant de sélectionner automatiquement les informations en fonction de certaines règles prédéfinies par l'utilisateur.

En effet, RSS est moins en moins populaire depuis la deuxième décennie du 21^e siècle. Les lecteurs de flux d'informations au format RSS largement utilisés sont arrêtés tels que Google Reader²⁹. En outre, les nouvelles versions de navigateurs principaux tels que Firefox, Chrome et Microsoft Edge ont supprimé le support de RSS³⁰, bien qu'il soit possible d'y ajouter les extensions pour utiliser RSS. Selon Steve Gillmore, les RSN tels que Facebook et Twitter fournissent un flux contenant toutes les dernières nouvelles sur Internet. Les RSN ont succédé à RSS parce qu'ils étaient tout simplement de meilleurs flux (Gillmor, 2009).

²⁹ <https://googleblog.blogspot.com/2013/03/a-second-spring-of-cleaning.html>

³⁰ <https://www.mozilla.org/en-US/firefox/64.0/releasenotes/>

I.4.3 Wiki

Le wiki représente parfaitement l'une des tendances du web 2.0 : tirer parti de l'intelligence collaborative. Le mot « wikiwiki » vient de l'hawaïen « wee kee wee kee » qui signifie « très rapide » et « très vite ». En 1995, Ward Cunningham l'a choisi pour nommer le premier système wiki : « Wikiwikiweb³¹ » qui était une fonctionnalité ajoutée aux « *Portland Pattern Repository* », une section du site web de Cunningham & Cunningham, Inc. Dans le but de faciliter l'échange d'informations entre programmeurs informatiques, Ward Cunningham a créé cette fonctionnalité et développe continuellement des outils auxiliaires pour soutenir l'écriture collaborative au sein d'une communauté, enrichissant ainsi le concept wiki.

Aujourd'hui, un wiki désigne un système qui permet la création, la modification et l'illustration de pages de manière collaborative 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 logiciel, ou autrement dit, un moteur de gestion de contenu, dont la structure implicite est minimale, tandis que la structure explicite émerge en fonction des besoins des utilisateurs.

Wiki est une application qui tire parti l'intelligence collective des internautes pour construire une base de connaissances publique. Un wiki sert à partager les connaissances d'un certain domaine au sein d'une communauté. Un navigateur web standard (sans aucune extension ajoutant) est suffisant pour que n'importe quel utilisateur puisse créer de nouvelles pages ou modifier n'importe quelle page dans le site web du wiki. Un wiki n'est pas un site web soigneusement conçu par les experts et les écrivains professionnels, mais plutôt un site « organique » qui s'évolue constamment grâce à la création collaborative persistante. Bien entendu, la création collaborative signifie plus d'interaction entre les utilisateurs. C'est pourquoi, les wikis sont considérés parfois comme les plateformes de réseaux sociaux³². Jusqu'en 2019, de nombreux sites gérés par wiki existent (voir la liste de sites web utilisant la technologie wiki³³) parmi lesquels Wikipédia³⁴ est le plus consulté³⁵.

³¹ <http://wiki.c2.com/>

³² https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_a_social_networking_site

³³ https://en.wikipedia.org/wiki/List_of_wikis

³⁴ <https://www.wikipedia.org/>

³⁵ <https://www.alexa.com/topsites>

I.4.4 Folksonomie

En tant que l'un des concepts importants du web 2.0, la folksonomie est un système de classification collaboratif décentralisé et spontané qui se base sur une indexation effectuée par des non-spécialistes. La folksonomie constitue la possibilité pour l'utilisateur d'indexer des ressources en ligne à l'aide des tags (étiquettes) afin qu'il puisse plus aisément les retrouver grâce à un système de mots-clés.

Le terme « folksonomie » est une adaptation française du mot anglais « *folksonomy* », mot-valise combinant les mots « *folk* (le peuple, les gens) » et « *taxonomy* (la taxinomie) ». Outre la « folksonomie », il existe une multitude de termes différents pour le même concept : « *tagging collaborative* (l'étiquetage collaboratif) » (Golder et Huberman, 2006), « *social classification system* (le système de classification sociale) » (Feinberg, 2006), ou encore « *collaborative classification system* (le système de classification collaborative) » (Schmidt, 2003) *etc.*

Les premiers systèmes de folksonomie sont développés et mis en place en 2003 (Peters *et al.*, 2011). En 2004, le terme « folksonomie » est mentionné pour la première fois par l'architecte de l'information Thomas Vander Wal. Puis, il gagne en popularité lorsque Gene Smith l'a cité dans son blog (Smith, 2004) sur le thème de l'architecture de l'information (Peters, 2009).

Vander Wal définit la folksonomie en tant que « résultat de l'étiquetage personnel des informations et des objets (tout ce avec une URL) pour sa propre recherche. L'étiquetage (*tagging*) se fait dans un environnement social (partagé et ouvert aux autres). L'étiquetage est effectué par la personne qui consomme les informations. La valeur de cet étiquetage externe est dérivée du fait que les gens utilisent leur propre vocabulaire et ajoutent une signification explicite, qui peut provenir d'une compréhension inférée de l'information / de l'objet. Les gens ne catégorisent pas tant que fournir un moyen de connecter des éléments (placer des crochets) pour donner leur sens à leur propre compréhension » (Vander Wal, 2007).

La folksonomie peut également être définie comme un hypergraphe tripartite (Lambiotte et Ausloos, 2005) : les utilisateurs effectuant l'étiquetage, les tags sélectionnés pour la description et les ressources à décrire. Une folksonomie émerge lorsque les utilisateurs

étiquettent du contenu ou des informations, telles que des pages web, des photos, des vidéos, des podcasts, des tweets, des articles scientifiques et autres. Les utilisateurs créent des tags qu'ils utilisent pour étiqueter des ressources d'information. Les tags sont des mots-clés sélectionnables qui peuvent être librement attachés à toute ressource d'information (Peters, 2009). Les tags peuvent être utilisés pour la gestion, la classification et la description des contenus en ligne. Le système de folksonomie utilise ces tags pour indexer les informations dans le but de faciliter la recherche et l'accès des ressources. Folksonomie comprend également un ensemble d'URL qui sont utilisées pour identifier les ressources auxquelles les utilisateurs de différents sites web ont fait référence. Le comportement étiquetage est une activité volontaire d'utilisateurs qui annotent des ressources d'information avec des tags choisis de façon libre dans un vocabulaire illimité et incontrôlé (Strohmaier, Körner, & Kern, 2012).

En général, nous pouvons différencier deux types de folksonomies (Vander Wal, 2005; Sen *et al.*, 2006; Peters, 2009): la folksonomie large (*broad folksonomy*) et la folksonomie restreinte (*narrow folksonomy*).

Dans une folksonomie large (Figure 1 Folksonomie Large), de nombreux utilisateurs différents (de groupe d'utilisateurs A à F) étiquettent une même ressource qui peut être trouvée en mode de recherche à l'aide des tags ultérieurement. Ainsi, le contenu de cette ressource est décrit par les étiquettes pareilles, similaires ou différentes (de tag 1 à 5) des perspectives variées grâce auxquelles les utilisateurs peuvent non seulement voir les tags qu'ils ont ajouté mais également des recommandations (les tags associés) dans l'ensemble de la base de données fournies par le système. En utilisant une large folksonomie pour l'étiquetage multiple, le service de bookmarking social del.icio.us est l'un des services les plus importants du Web 2.0 (Vander Wal, 2005; Peters, 2009).

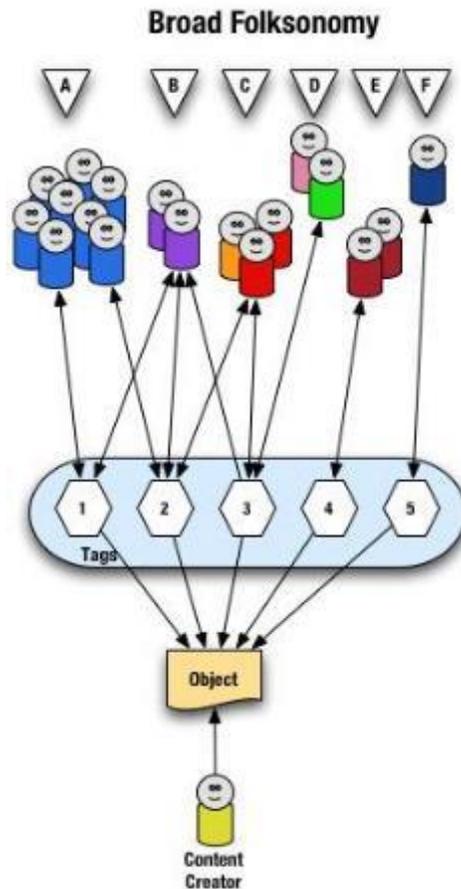


Figure 1 Folksonomie Large

Dans une folksonomie restreinte (Figure 2), quelques tags (souvent un ou deux tags) d'une ressource sont ajoutés par le créateur de ressource. Les nouveaux tags peuvent être ajoutés par quelques utilisateurs (groupe d'utilisateurs B) qui sont moins nombreux que dans une folksonomie large. Les utilisateurs du groupe C et D sont des consommateurs de ces tags qui n'ajoutent aucun tag. Les utilisateurs du groupe E ne peuvent pas trouver cette ressource car le vocabulaire qu'il utilise ne correspond à aucun des tags actuellement fournis. L'utilisateur du groupe F trouve cette ressource par d'autres moyens (un ami lui a envoyé un lien) et ajoute son propre tag (tag 3) pour cette ressource. La folksonomie étroite offre l'avantage d'étiqueter des ressources qui ne sont pas facilement consultables ou qui n'ont aucun autre moyen d'utiliser du texte pour décrire ou la trouver (Vander Wal, 2007).

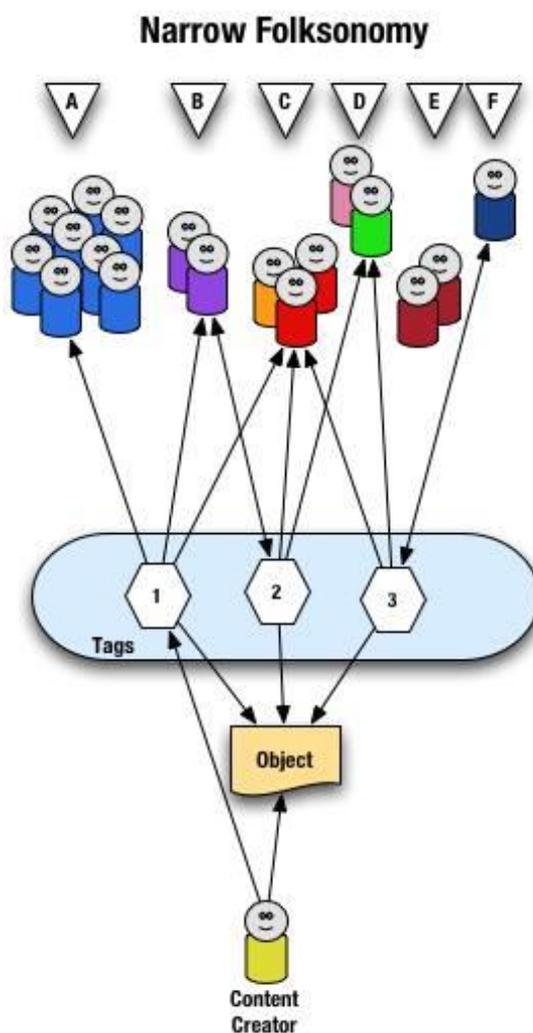


Figure 2 Folksonomie Restreinte

Contrairement à la taxinomie qui est une catégorisation hiérarchique dans laquelle des classes relativement bien définies sont imbriquées dans des catégories plus larges, les catégories établies par une folksonomie ne sont pas structurées. La folksonomie est tellement simple à utiliser que tout le monde peut contribuer à la formation du web. Elle facilite la diffusion, la recherche et le partage de l'information. La folksonomie est réellement centrée sur l'utilisateur : qui permet aux utilisateurs de choisir librement leurs propres mots-clés. Le vocabulaire dans une folksonomie peut refléter directement le vocabulaire et les façons de penser des utilisateurs (Marieke Guy, 2006; Spiteri, 2006; Gordon, 2007; Peters, 2009)

L'utilisation de tag et de folksonomie présente également plusieurs inconvénients. La simplicité de l'étiquetage peut entraîner des tags mal appliqués. Il n'y a aucun control du vocabulaire dans le système de folksonomie, par conséquent les tags sont souvent ambigus et

trop personnalisés. Le système de folksonomie manque souvent de mécanismes pour gérer les synonymes, les acronymes et les homonymes. Il manque également souvent de mécanismes pour gérer les variations d'orthographe telles que les fautes d'orthographe, la forme singulière / plurielle, les mots conjugués et composés. Les utilisateurs choisissent des tags spécialisés ou des tags sans signification pour les autres. (Hayman, 2007; Kipp et Campbell, 2007; Peters, 2009).

I.4.5 Social bookmarking

La folksonomie est une caractéristique importante du *social bookmarking* (marque-page social). Le social bookmarking est un service en ligne qui permet aux utilisateurs d'ajouter, de stocker, de classer, de chercher et de partager leurs signets de document en ligne (*bookmarks*) favoris. (Hammond *et al.*, 2005; Noll et Meinel, 2007; Aichner et Jacob, 2015).

Le concept de listes de signets partagées en ligne date d'avril 1996 avec l'apparition du site itList.com. Au cours des trois années suivantes, les services de social bookmarking sont devenus compétitifs, avec l'arrivée sur le marché d'entreprises financées par le capital-risque telles que Backflip, Blink, Clip2, ClickMarks, HotLinks, et d'autres. Manquant de modèles économiquement viables, la plupart de cette première génération de site de bookmarking social a disparu avec l'éclatement de la bulle Internet (D'Souza, 2007; Sankar et Bouchard, 2009; Sfetcu, 2017). Fondé en 2003, Del.icio.us³⁶ a popularisé les termes « social bookmarking » et « étiquetage (*tagging*) » (Gordon-Murnane, 2006; Hayman, 2007; Peters, 2009; Peters *et al.*, 2011). Depuis 2004, de nombreux services similaires ont été réalisés tels que Flickr³⁷, Furl³⁸, Reddit³⁹, Simpy⁴⁰, Connotea⁴¹, CiteULike⁴², BibSonomy⁴³, Ma.grolia (plus tard renommé

³⁶ <https://del.icio.us/>

³⁷ <https://www.flickr.com/>

³⁸ <https://en.wikipedia.org/wiki/Furl>

³⁹ <https://www.reddit.com/>

⁴⁰ <https://en.wikipedia.org/wiki/Simpy>

⁴¹ <https://en.wikipedia.org/wiki/Connotea>

⁴² <https://en.wikipedia.org/wiki/CiteULike>

⁴³ <https://www.bibsonomy.org/>

Gnolia)⁴⁴, Stumbleupon⁴⁵, Blue Dot (plus tard renommé Faves)⁴⁶, Diigo⁴⁷, Pinboard⁴⁸ et Pinterest⁴⁹. Depuis 2012, certains services tels que Furl, Simpy, Connotea, Gnolia, Faves et Stumbleupon ne sont plus actifs⁵⁰.

Le système de bookmarking social permet aux utilisateurs d'enregistrer des listes de liens vers des ressources en ligne qu'ils trouvent utiles et qu'ils veulent partager. Ces signets sont généralement publics, mais ils peuvent être enregistrés en privé, partagés uniquement avec des personnes ou des groupes spécifiés, partagés uniquement à l'intérieur de certains réseaux, ou une autre combinaison de domaine public et privé. Les utilisateurs autorisés peuvent généralement afficher ces signets de manière chronologique ainsi que par sujet, par catégorie, par tag ou même de façon aléatoire. En dehors des favoris web, on peut trouver d'autres services spécialisés sur un sujet particulier (alimentation et vins, livres, vidéos, commerce en ligne, cartographie *etc.*).

Le bookmarking social n'enregistre pas les ressources elles-mêmes, simplement des signets (un lien vers la page). Les utilisateurs peuvent ajouter des descriptions à ces signets sous la forme de métadonnées afin que toutes les ressources soient catégorisées. Des mots-clés ou des tags peuvent être ajoutés par les utilisateurs pour que les autres utilisateurs puissent comprendre le contenu de la ressource. Les utilisateurs peuvent également noter une ressource en fonction de sa qualité ou du nombre d'utilisateurs qui l'apprécient.

La plupart des services de social bookmarking encouragent les utilisateurs à organiser leurs signets avec des tags informels qu'à utiliser système de dossiers traditionnel basé sur un navigateur. Les utilisateurs peuvent chercher des signets en fonction de leur association avec des mots-clés. De nombreux services de bookmarking social ont développé des algorithmes qui permettent de tisser des liens entre les mots-clés attribués aux ressources en examinant des groupes de mots-clés particuliers et leur rapport avec d'autres mots-clés. Ils sont capables d'afficher les signets associés à un tag choisi en comprenant des informations sur le nombre d'utilisateurs qui les ont marqués. Certains services de bookmarking social tirent également des

⁴⁴ <https://en.wikipedia.org/wiki/Gnolia>

⁴⁵ <https://en.wikipedia.org/wiki/StumbleUpon>

⁴⁶ <https://en.wikipedia.org/wiki/Faves.com>

⁴⁷ <https://www.diigo.com/>

⁴⁸ <http://pinboard.in/>

⁴⁹ <https://www.pinterest.com/>

⁵⁰ https://en.wikipedia.org/wiki/List_of_social_bookmarking_websites

conclusions de la relation entre les tags pour créer des groupes de tags ou de signets.

Certains services de bookmarking social fournissent des flux web concernant leurs listes de signets, y compris des listes organisées par tags. Cela permet aux utilisateurs de prendre connaissance des nouveaux signets lorsqu'ils sont enregistrés, partagés et balisés par d'autres utilisateurs.

Le succès grandissant et la concurrence permettent de développer le nombre de services, il ne s'agit plus uniquement de partager des signets mais également des fonctionnalités supplémentaires telles que les évaluations et les commentaires sur les signets, la possibilité d'importer et d'exporter des signets à partir de navigateurs, l'envoi de signets par mail, les notifications automatiques ou s'abonner, des annotations web et des groupes ou d'autres fonctionnalités de réseaux sociaux (Hammond *et al.*, 2005).

Les services de bookmarking social améliore la fonction des applications conventionnelles de bureau en permettant à l'utilisateur d'enregistrer la liste des favoris du navigateur web en ligne, indépendamment du bureau, et ainsi de rendre les ressources déjà trouvées plus facilement récupérables (Gordon-Murnane, 2006). Les services de bookmarking social fournissent des listes de signets auxquels les utilisateurs peuvent ajouter des descriptions et des résumés. Ils permettent aux utilisateurs non seulement de gérer leurs propres liens en privé, mais aussi de rendre leurs collections de favoris accessibles au public aux autres utilisateurs du même service (Peters, 2009). Ainsi, l'ensemble de la communauté en ligne peut partager et profiter de l'activité d'un seul utilisateur (Hammond *et al.*, 2005). Le social bookmarking est très similaire à wiki en permettant aux internautes de partager leurs favoris et d'échanger des ressources entre eux.

Le social bookmarking subit des mêmes problèmes que la folksonomie à cause de l'ambiguïté de langue, des fautes d'orthographe, l'incapacité de traiter des synonymes ou des antonymes *etc.* En outre, le bookmarking social est également sujet à corruption et connivence. En raison de sa popularité, certains utilisateurs commencent à le considérer comme un outil pour rendre leur site web plus visible. Plus une page web est envoyée et étiquetée, plus elle aura de chance d'être trouvée. Les spammeurs utilisent un grand nombre de tags populaires pour faire référence de manière répétée à une même page et/ou à chaque page de leur site. Cela oblige les développeurs à améliorer constamment leur système de sécurité pour éviter les abus. C'est pour cette raison que certains sites de bookmarking social ont dû ajouter des tests de

« CAPTCHA », ce qui pose problème aux gens utilisant le site pour des raisons autres que le spamming (Hammond *et al.*, 2005).

Le Web 2.0 augmente l'interactivité au prix d'une complexification interne de la technologie mais permet plus de simplicité d'utilisation, les connaissances techniques et informatiques n'étant pas indispensables pour les utilisateurs. Les technologies du Web 2.0 préparent et créent les conditions qui favorisent l'émergence et le développement des RSN. Les applications du Web 2.0 sont orientées vers le contenu généré par l'utilisateur : elles ont ainsi augmenté la possibilité de publier, d'annoter et commenter (folksonomies et commentaires en ligne) ainsi que de diffuser (blog, fils RSS, wikis etc.). Elles n'offrent pas nécessairement de nouveaux moyens d'orientation dans les contenus, ni de personnalisation. L'idée d'améliorer l'orientation de l'utilisateur dans cet univers d'information « augmenté » par le web 2.0 repose sur le développement de l'exposition sélective à l'information. La « personnalisation » devient, dans ce contexte, un enjeu reposant sur les données des utilisateurs.

Chapitre II Réseau social numérique

II.1 Définition

La notion de « réseau social » a été introduite pour la première fois dans un article « *Class and Committees in a Norwegian Island Parish* » de John A. Barnes, un anthropologue social britannique (Barnes, 1954). Et « depuis, le recours à la notion de réseau pour désigner des ensembles de relations entre personnes ou entre groupes sociaux s'est largement répandu, à l'intérieur des sciences sociales comme à ses marges » (Mercklé, 2004).

Selon Michel Forsé, « un réseau social est un ensemble de relations entre un ensemble d'acteurs. Cet ensemble peut être organisé (une entreprise, par exemple) ou non (comme un réseau d'amis) et ces relations peuvent être de nature fort diverse (pouvoir, échanges de cadeaux, conseil, *etc.*), spécialisées ou non, symétriques ou non » (Forsé, 2008).

D'après les deux définitions ci-dessus données par Mercklé et Forsé, nous pouvons en déduire les trois concepts clés d'un réseau social : les individus, leurs liens et l'environnement social (une entreprise, par exemple).

Les RSN ne sont apparus qu'avec l'évolution d'internet depuis les années 1990 et surtout avec l'arrivée du Web 2.0 vers 2005. Depuis, les RSN ont connu un développement et une propagation très rapide.

Boyd et Ellison donnent leur définition des RSN comme des services en ligne (*social network sites*) qui permettent à leurs utilisateurs (Boyd et Ellison, 2007) :

- (1) de construire un profil public ou semi-public au sein d'un système ;
- (2) de gérer une liste des utilisateurs avec lesquels ils partagent un lien ;
- (3) de voir et naviguer sur leur liste de liens et sur ceux établis par les autres au sein du système.

Dans cette définition, ils emploient la terminologie « *Social Network Sites* » pour qualifier les sites web qui aident leurs utilisateurs à socialiser des relations déjà existées entre des amis,

alors que le terme « *Social Networking Sites* » désignent les sites web qui cherchent plutôt à initier des relations, c'est-à-dire, à rencontrer des étrangers.

Cependant cette définition n'est pas parfaite. Selon Beer, elle est trop large car elle regroupe un vaste nombre d'applications souvent bien différentes (Beer, 2008). Par exemple, YouTube correspond bien la définition ci-dessus, mais son objectif principal n'est pas l'amitié. Thelwall (2009) propose que l'on puisse garder la définition de Boyd et Ellison en introduisant une typologie de RSN qui regroupe réseau social de socialisation (*Social Network Site*), réseau social de réseautage (*Social Networking Site*) et réseau social de navigation (les sites comme YouTube) (Thelwall, 2009).

Andreas Kaplan et Michael Haenlein soulignent que les RSN (Facebook, YouTube *etc.*) sont un sous-ensemble des médias sociaux (Figure 3). Ils définissent les médias sociaux comme un groupe d'applications en ligne qui se fondent sur l'idéologie et la technologie du Web 2.0 en permettant la création et l'échange du contenu généré par les utilisateurs (Kaplan et Haenlein, 2010).



Figure 3 Une Définition Problématique des RSN

La notion du Web 2.0 désigne l'ensemble des usages, des outils et des fonctionnalités du

web qui permettent à tous les internautes de publier et d'échanger d'informations. Les applications du Web 2.0 regroupent tous les plateformes numériques sur lesquelles les contenus sont créés, publiés et modifiés par les internautes au lieu des webmasters (O'Reilly, 2007).

Dans notre travail, nous gardons la définition de Boyd et Ellison et la typologie de Thelwall. Les réseaux sociaux numériques sont une partie des médias sociaux qui sont inclus dans les applications du Web 2.0.

II.2 Évolution des réseaux sociaux numériques

Inventé dans les années 1960, l'internet n'est accessible au grand public qu'à partir des années 1990. Dès lors il a accéléré son évolution et a donné la naissance de services réseaux sociaux.

Lancé en 1995, le « *Classmates*⁵¹ » permettait à ses utilisateurs de s'affilier à leur école et de surfer avec les autres qui s'y affilient aussi. En 1997, AIM (*AOL Instant Messenger*)⁵² est apparu, grâce auquel les utilisateurs peuvent communiquer avec les autres en envoyant les messages instantanés. Dès ce moment-là, il est déjà possible pour un individu d'avoir un lien avec les autres en ligne. Cependant, selon la définition donnée par Boyd et Ellison, le « vrai » RSN n'existe qu'après le lancement du « *Six Degrees*⁵³ », car « l'on pouvait y ajouter son propre profile, avoir lien avec les autres et se communiquer. » En bref, il possédait presque toutes les propriétés d'un RSN (Boyd et Ellison, 2007).

Pendant les années suivantes, de nombreux RSN ont apparu : « *Asian Avenue*⁵⁴ » (1997), « *Xanga*⁵⁵ » (1998), « *Live Journal*⁵⁶ » (1999), « *Cyworld*⁵⁷ » (1999), « *LunarStorm*⁵⁸ » (2000),

⁵¹ <https://www.classmates.com/>

⁵² https://fr.wikipedia.org/wiki/AOL_Instant_Messenger

⁵³ <http://sixdegrees.com/>

⁵⁴ Asian Avenue (ou AsianAve) est un RSN destiné à la communauté américaine d'origine asiatique. En mars 2019, AsianAve.com a commencé à rediriger vers BlackPlanet.com, qui appartient également à la société mère Radio One.

⁵⁵ Xanga est un site Web des weblogs, des photoblogs et des profils de réseaux sociaux. <http://xanga.com>

⁵⁶ LiveJournal (souvent abrégé en LJ) est une communauté virtuelle dont les utilisateurs peuvent tenir un blog ou un journal. <https://www.livejournal.com/>

⁵⁷ Cyworld est un service de réseau social sud-coréen exploité par SK Communications. <http://cyworld.co.kr/>

⁵⁸ LunarStorm (ou Lunar) était un site Web de réseautage social financé par la publicité suédoise pour les adolescents et fermé le 18 août 2010 en raison d'un manque d'activité. <https://en.wikipedia.org/wiki/LunarStorm>

« Ryze⁵⁹ » (2000) et « BlackPlanet⁶⁰ » (2000) par exemple. Lancés respectivement en 2002 et en 2003, « Friendster⁶¹ » et « MySpace⁶² » ont connu un grand succès. Il ne faut pas négliger certainement les autres comme « Meetup⁶³ » (2002), « BeautifulPeople⁶⁴ » (2002), « Hi5⁶⁵ » (2003), « CouchSurfing⁶⁶ » (2003), « Tribe⁶⁷ » (2003). On constate aussi la naissance de « LinkedIn⁶⁸ » en 2003.

Jusqu'à 2003, les RSN se sont développés progressivement. La plupart d'entre eux sont lancés aux États Unies sauf « Cyworld » (en Corée du Sud) et « LunarStorm » (en Suède). Ils se limitent même à certaines régions et certains types d'individus. C'est grâce à « Facebook » que les RSN se sont déployés dans le monde entier et ont montré leurs diversités : en 2004, « Orkut⁶⁹ » au Brésil, « MiXi⁷⁰ » au Japon ; en 2005, « Hyves⁷¹ » aux Pays-Bas et « Bebo⁷² » en Angleterre, en Nouvelle-Zélande et en Australie ; en 2006, « VK⁷³ » en Russie par exemple. Et grâce au développement de la technologie du Web 2.0, les RSN évoluent : les utilisateurs peuvent ainsi partager les photos (comme Flickr en 2004 et Printrest en 2012), les vidéos (YouTube en 2005). Les utilisateurs peuvent créer son propre RSN (Ning⁷⁴ en 2004).

Depuis 2006, l'évolution du Smartphone a fortement influencé les RSN. Le « micro blog » s'est largement répandu ; « Twitter⁷⁵ », par exemple, est devenu un autre synonyme de RSN après Facebook. Il n'est pas difficile d'en énumérer d'autres comme « Tumblr⁷⁶ » (2007), «

⁵⁹ Ryze est un RSN gratuit et payant conçu pour relier les professionnels, en particulier les nouveaux entrepreneurs. <https://www.ryze.com/>

⁶⁰ BlackPlanet est un RSN afro-américain. <https://www.blackplanet.com/discover/message/1>

⁶¹ Friendster était un RSN essentiellement pour les jeux basés à Kuala Lumpur en Malaisie. Il est fermé le 14 juin 2015. <https://en.wikipedia.org/wiki/Friendster>

⁶²<https://myspace.com/>

⁶³ Meetup est un RSN utilisé pour créer des groupes qui organisent un événement. <https://www.meetup.com>

⁶⁴ BeautifulPeople est un RSN de rencontres en ligne. <https://www.beautifulpeople.com>

⁶⁵ <https://secure.hi5.com>

⁶⁶ CouchSurfing est un RSN dont l'objet social est d'assurer un service d'hébergement temporaire et gratuit, de personne à personne. <https://www.couchsurfing.com/>

⁶⁷ Tribe était une communauté en ligne et est fermé en 2007. <https://en.wikipedia.org/wiki/Tribe.net>

⁶⁸ LinkedIn est un RSN professionnel. <https://www.linkedin.com/>

⁶⁹ Orkut est fermé le 30 septembre 2014. <https://en.wikipedia.org/wiki/Orkut>

⁷⁰ <https://mixi.jp/>

⁷¹ Hyves était RSN localisé aux Pays-Bas, principalement composé de membres et visiteurs néerlandais. Le site change en plateforme de jeux en ligne dès fin 2013. <https://hyvesgames.nl>

⁷² <http://www.bebo.com/>

⁷³ <https://vk.com/>

⁷⁴ <https://www.ning.com/>

⁷⁵ <https://twitter.com>

⁷⁶ <https://www.tumblr.com/>

Foursquare⁷⁷ » (2009), « Google Buzz⁷⁸ » (2010) et « Google+⁷⁹ » (2011).

II.3 Typologie des réseaux sociaux numériques

En principe, les RSN sont des plateformes gratuites et ouvertes qui permettent l'interaction entre les utilisateurs. L'utilisateur peut y exploiter de données personnelles d'autres utilisateurs ; partager de contenus, d'opinions et de sentiments ; de regrouper d'autres utilisateurs autour de centres d'intérêt *etc.*

Thelwall catégorise les RSN selon leurs objectifs (Figure 4) qu'il nomme respectivement : *socialisation, réseautage et navigation* (Thelwall, 2009) :

1. les réseaux sociaux de **socialisation** sont conçus pour les loisirs de la communication sociale entre les membres. Les connexions sont souvent utilisées pour trouver et afficher des listes d'amis existants hors ligne. Nous pouvons citer MySpace, Hi5, Facebook, et Cyworld par exemple.
2. les réseaux sociaux de **réseautage** sont utilisés pour trouver de nouveaux contacts au travers de connexions d'amis déjà existantes. Ces nouveaux contacts comportent une proportion importante de connaissances et de personnes auparavant inconnues. LinkedIn est un exemple représentatif.
3. Les réseaux sociaux de **navigation** possèdent des caractéristiques de réseau social et sont utilisés dans le but d'aider les membres à trouver certains types spécifiques d'information et de ressources. Les connexions sont utilisées comme un outil pour déployer des listes de contacts, listes permettant l'accès à l'information et aux ressources associées à ceux-ci. Digg.com, Filckr.com en font partie.

Bien sûr certains sites peuvent recouvrir plusieurs objectifs. La classification ci-dessus est donc dynamique et dépend plus de l'intention d'un site ou les pratiques de ses membres que

⁷⁷ <https://foursquare.com/>

⁷⁸ Google Buzz a arrêté le 15 décembre 2011 pour laisser la place à Google+.
https://en.wikipedia.org/wiki/Google_Buzz

⁷⁹ Google+ est fermé progressive a été annoncée le 9 octobre 2018. <https://en.wikipedia.org/wiki/Google%2B>

de ses caractéristiques.

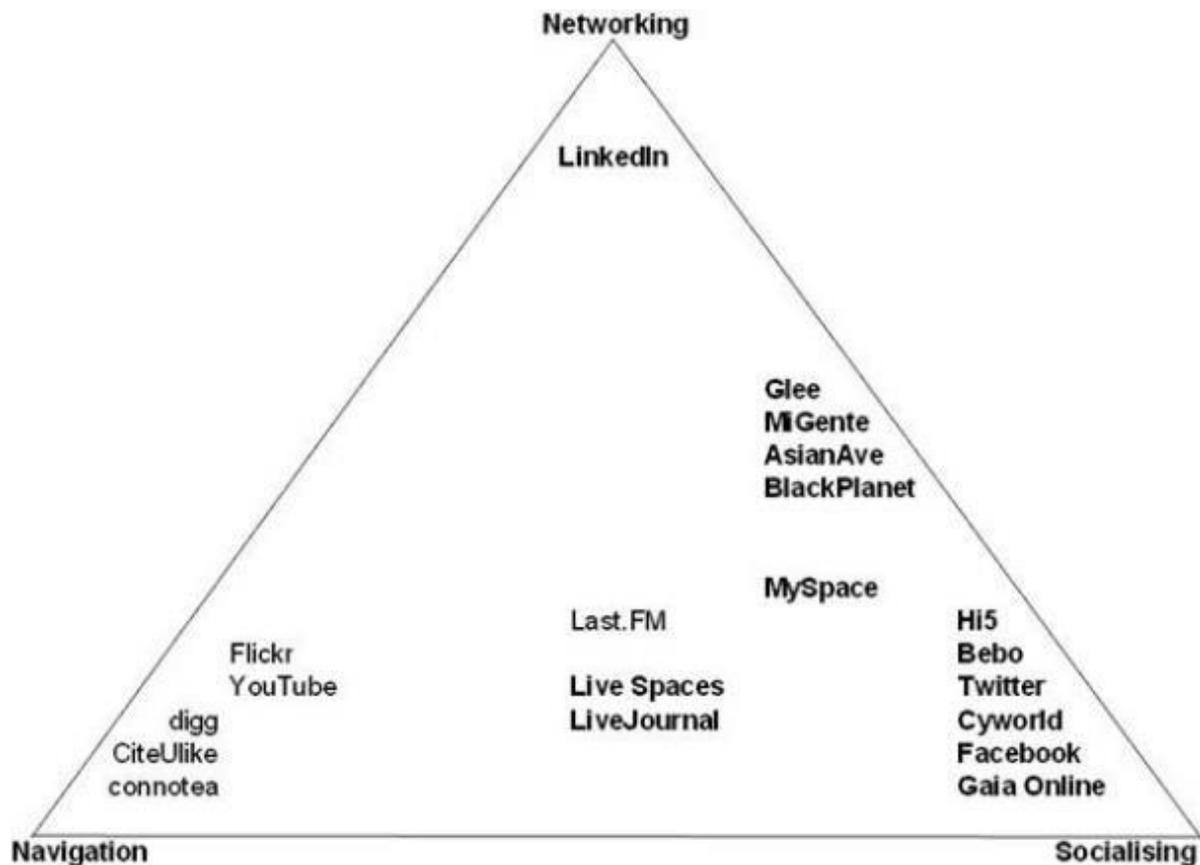


Figure 4 Exemples des RSN avec des objectifs divers (Thelwall, 2009)

Cavazza classe les médias sociaux selon les six grands usages et propose un panorama assez complet des médias sociaux existants actuellement (Figure 5). Ce panorama nous aide à avoir une image plus claire des RSN qui sont un sous-ensemble des média sociaux. Il propose une répartition des différents services en six grands usages : publication, partage, messagerie, discussion, collaboration et réseautage (Cavazza, 2019) :

- les outils de **publication** avec les plateformes d'hébergement de blog (WordPress⁸⁰, Blogger, LiveJournal, TypePad⁸¹ etc.), les wikis (Wikipedia, Fandom⁸²), les plateformes de partage (Houzz⁸³ ou SVBTLE⁸⁴), et les services hybrides de

⁸⁰ <https://fr.wordpress.com/>

⁸¹ <https://www.typepad.com/>

⁸² <https://www.fandom.com/>

⁸³ <https://www.houzz.fr/>

⁸⁴ <https://svbtle.com/>

publication et partage (Tumblr, MySpace, Mastodon⁸⁵ etc.);

Social Media Landscape 2019



Figure 5 Panorama des médias sociaux 2019 (Cavazza, 2019)

- les outils **de partage** de photos (Flickr, Google Photos⁸⁶, 500px⁸⁷ etc.), de vidéos (YouTube, Vimeo⁸⁸, Dailymotion⁸⁹, etc.), de musique (Spotify⁹⁰, Deezer⁹¹, SoundCloud⁹²), d'avis (TripAdvisor⁹³, Yelp⁹⁴) etc. ;
- les outils de **discussion** avec les plateformes de questions/réponses (Quora⁹⁵, Ask⁹⁶,

⁸⁵ <https://mastodon.social/about>

⁸⁶ <https://photos.google.com/>

⁸⁷ <https://web.500px.com/>

⁸⁸ <https://vimeo.com/fr/>

⁸⁹ <https://www.dailymotion.com/fr>

⁹⁰ <https://www.spotify.com/fr/>

⁹¹ <https://www.deezer.com/en/>

⁹² <https://soundcloud.com/>

⁹³ <https://www.tripadvisor.com/>

⁹⁴ <https://www.yelp.fr/lille>

⁹⁵ <https://fr.quora.com/>

⁹⁶ <https://ask.fm/>

etc.) ; les plateformes de conversation (Reddit⁹⁷, GitHub⁹⁸, StackExchange⁹⁹) ; les systèmes de gestion de commentaires et débats (Disqus¹⁰⁰, Discourse¹⁰¹, Muut¹⁰² *etc.*) ; les espaces de partage / discussion (4Chan¹⁰³, 8Chan¹⁰⁴, WizardChan¹⁰⁵, *etc.*) ;

- les outils de **communication** : les applications de messagerie mobile (iMessages, Google Messages, Viber *etc.*) ; les messageries traditionnelles (Gmail, Outlook, Yahoo Mail) ; les messageries professionnelles (Slack, TalkSpirit, Hangouts Chat, *etc.*) ;
- les plateformes de **collaboration** (Microsoft Teams¹⁰⁶, Jive¹⁰⁷, Jamespot¹⁰⁸ *etc.*) ; les espaces de social VR (Sansar¹⁰⁹, Facebook Spaces¹¹⁰, Oculus Rooms¹¹¹, *etc.*) ;
- et le **réseautage** avec les services / applications mobiles de rencontre (Tinder¹¹², Badoo¹¹³, OKcupid¹¹⁴ *etc.*), les plateformes d'évènements (Meetup et Eventbrite¹¹⁵), les RSN de proximité (Nextdoor¹¹⁶) et outils de création de réseaux sociaux (Ning¹¹⁷, Whaller¹¹⁸) *etc.*

⁹⁷ <https://www.reddit.com/>

⁹⁸ <https://github.com/>

⁹⁹ <https://stackexchange.com/>

¹⁰⁰ <https://disqus.com/>

¹⁰¹ <https://www.discourse.org/>

¹⁰² <https://muut.com/>

¹⁰³ <http://www.4chan.org/>

¹⁰⁴ <https://8ch.net/>

¹⁰⁵ <https://wizchan.org/>

¹⁰⁶ <https://teams.microsoft.com/start>

¹⁰⁷ <https://www.jivesoftware.com/>

¹⁰⁸ <https://www.fr.jamespot.com/>

¹⁰⁹ <https://www.sansar.com/>

¹¹⁰ <https://www.facebook.com/spaces>

¹¹¹ <https://www.oculus.com/experiences/gear-vr/1101959559889232/>

¹¹² <https://tinder.com/>

¹¹³ <https://badoo.com/>

¹¹⁴ <https://www.okcupid.com/>

¹¹⁵ <https://www.eventbrite.com/>

¹¹⁶ <https://fr.nextdoor.com/>

¹¹⁷ <https://www.ning.com/>

¹¹⁸ <https://whaller.com/en/>

II.4 RSN en Chine et en France

Dans le domaine des RSN, le cas de la Chine est particulier puisque certains RSN très populaires dans le monde ne sont pas accessibles en Chine. C'est le cas de Facebook, Twitter, Instagram par exemple. Toutefois, leurs équivalences chinoises existent (Kantar, 2019).

Fondé le 6 mars 2005 par Yang Bo, Douban¹¹⁹ est le premier RSN chinois. C'est un RSN dédié aux thèmes culturels qui se concentre notamment sur les critiques de livres, de musique et de films. En décembre 2005, Renren¹²⁰ (appelé à l'origine Xiaonei) est créé par un groupe d'étudiant de l'université de Tsinghua, Wang Xing, Wang Huiwen et Lai Binqiang. Il est une équivalence de Facebook. Un autre RSN comparable intitulé Kaixin001¹²¹ a été fondé en mars 2008. La plupart des utilisateurs sont des jeunes employées. Sina Weibo¹²² est considéré comme l'imitateur de Twitter et devient le RSN le populaire en Chine aussitôt après sa fondation en 2008.

Les RSN chinois (Figure 6) ont une influence plus importante sur des décisions d'achat des utilisateurs que ceux d'autres pays. Les consommateurs chinois avouent qu'ils sont davantage susceptibles d'acheter les produits qui ont des commentaires positifs sur les RSN et qu'ils sont plus tentés d'acheter un produit ou un service que leurs amis leur recommandent dans les RSN (Chiu, Ip et Silverman, 2012).

¹¹⁹ <https://www.douban.com/>

¹²⁰ <http://www.renren.com/>

¹²¹ <http://www.kaixin001.com/>

¹²² <https://www.weibo.com/>

Les médias sociaux chinois les plus actifs sont Wechat, Baidu Tieba¹²³, QQ¹²⁴, Sina Weibo, Youku¹²⁵, Qzone¹²⁶, Tencent Weibo¹²⁷ et Tudou¹²⁸ (Figure 8). La moitié d’entre eux sont des produits développés par l’entreprise Tencent.

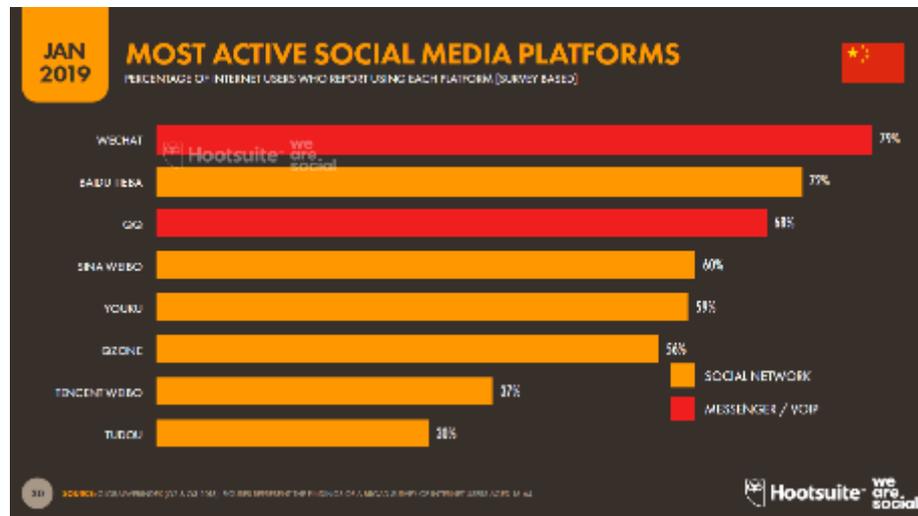


Figure 8 Les médias sociaux chinois les plus actifs

Sur 65,36 millions d’habitants, la France compte 60,42 millions d’internautes, soit un taux de pénétration de 92%. En ce qui concerne l’utilisation des médias sociaux, 38 millions de Français sont actifs sur les médias sociaux représentant 58% de la population française (Figure 9) (We Are Social & Hootsuite, 2019b).

¹²³ Baidu Tieba est un moteur de forum Internet chinois. <https://tieba.baidu.com/>

¹²⁴ QQ est le système de messagerie instantanée propriétaire. <https://im.qq.com/>

¹²⁵ Youku est un site web d’hébergement de vidéo. <https://www.youku.com/>

¹²⁶ Qzone est un RSN qui permet aux utilisateurs de publier des blogs, de tenir des journaux intimes, d’envoyer des photos, d’écouter de la musique et de regarder des vidéos. <https://qzone.qq.com/>

¹²⁷ Tencent Weibo est un site de microblogging chinois. <http://t.qq.com/>

¹²⁸ Toudou est un site web d’hébergement de vidéo. <https://www.tudou.com/>

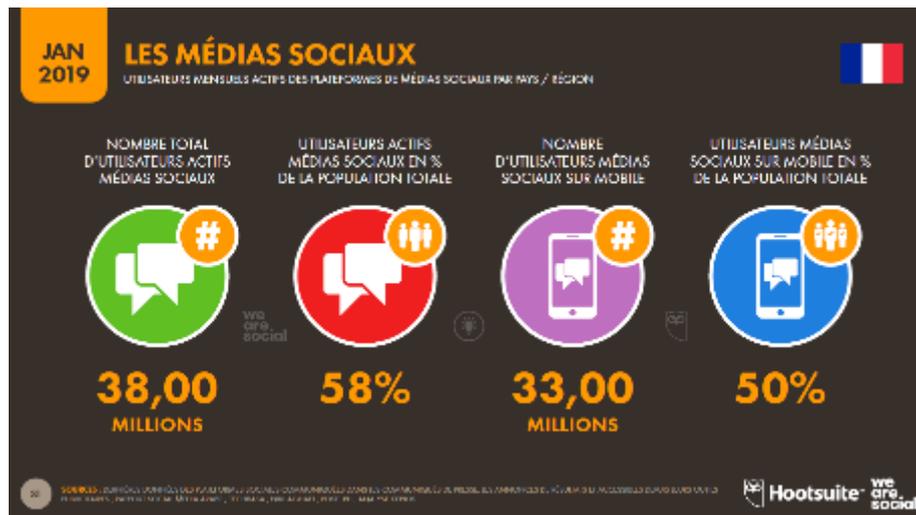


Figure 9 Nombre d'utilisateurs français actifs des médias sociaux (We Are Social & Hootsuite, 2019b)

Contrairement au cas de la Chine, les 10 RSN¹²⁹ les plus visités en France sont : YouTube, Facebook, Messenger, Instagram, WhatsApp, Snapchat, Twitter, Skype, Pinterest et LinkedIn (Figure 10). Nous constatons que ce sont des RNS internationaux provenant des États-Unis et qu'il n'y a pas beaucoup de RSN français.

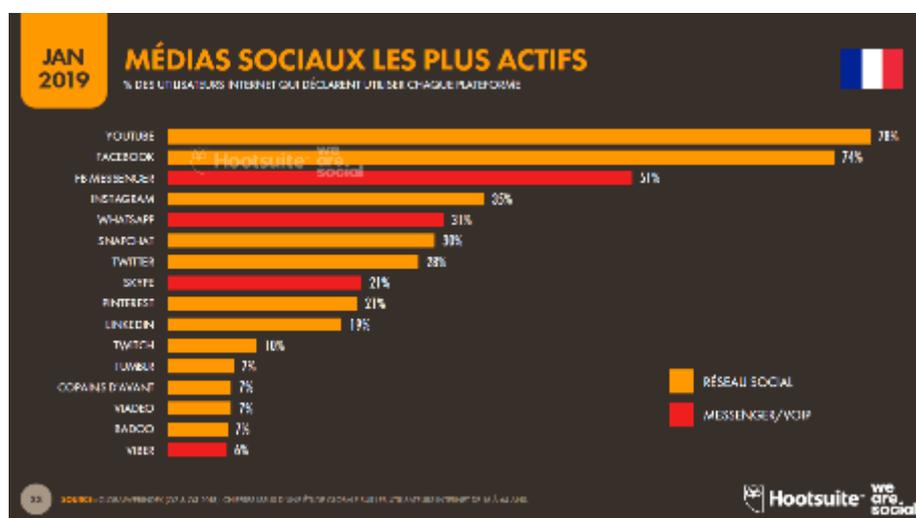


Figure 10 Médias sociaux les plus actifs en France (We Are Social & Hootsuite, 2019b)

La croissance et le développement des RSN dans le monde entier ont sans aucun doute

¹²⁹ <https://www.alexa.com/>

fortement aggravé le problème de la surcharge d'informations. La grande quantité d'informations générées par les utilisateurs est souvent désorganisée. Il devient indispensable de disposer les systèmes de l'information qui peuvent orienter efficacement les utilisateurs en les aidant à filtrer les informations impertinentes. Dans la partie suivante, nous allons présenter le système de filtrage de l'information et le système de recommandation qui sont les plus utilisés et largement appliqués dans les RSN.

PARTIE 2 LE CIBLAGE DE L'INDIVIDU

PERSONNALISATION ET FILTRAGE

COLLABORATIF

La multiplication des contenus disponibles sur le web au fur et à mesure de son évolution, comme nous venons de le voir, a fait des systèmes de recherche et d'orientation, les portes d'entrée très stratégiques dans l'univers informationnel du Web. Comme alternative aux moteurs de recherche qui se basent sur des caractéristiques générales des sites telles que la popularité, les systèmes de filtrage d'information et de recommandation proposent une présentation des informations basée sur le ciblage de l'individu. L'individu n'est plus limité à une requête mais à l'ensemble des données reflétant son histoire, son parcours sur les systèmes qu'il utilise, les interactions qui jalonnent son utilisation du site. Ces données étant supposées restituer sa « personne », la personnalisation devient une stratégie d'orientation dans l'univers informationnel.

Cette partie est consacrée à la présentation des concepts de base des technologies de filtrage d'information et de système de recommandation tels qu'ils reposent sur le ciblage de l'individu. Les chapitres III et IV présentent respectivement les définitions, les typologies et les techniques principales des systèmes de filtrage d'information (SFI) et des systèmes de recommandation (SR).

Chapitre III Le filtrage d'information : le système de l'information personnalisable

L'accès à l'information peut être effectué de manière volontaire à travers les moteurs de recherche, mais les outils de recherche d'information (RI) ne sont pas suffisants pour faire face aux flux d'informations générés. Les moteurs de RI requièrent de l'utilisateur une activité de formulation systématique de son besoin en formulant une requête et en utilisant des mots-clés. De plus, le résultat retourné à l'utilisateur contient souvent un grand nombre de documents non pertinents. L'utilisateur doit donc sélectionner manuellement ceux qui l'intéressent. Il s'agit d'une tâche coûteuse en temps et fastidieuse. Au lieu de cela, une autre façon « involontaire » d'accéder à l'information peut être effectuée par ce que l'on appelle un système de filtrage d'information (SFI) qui permet de faciliter la tâche de l'utilisateur en lui faisant parvenir continuellement l'information supposée l'intéresser. Les SFI pérennisent ce besoin d'information et permettent l'acheminement régulier des documents intéressants. Ils aident les utilisateurs qui n'ont pas suffisamment d'expérience ou la compétence nécessaire pour les évaluer et les trier alors qu'ils sont en nombre potentiellement important.

Nous présentons dans ce chapitre les notions relatives au filtrage d'information y compris l'architecture d'un SFI et les différents types de filtrage.

III.1 Définition du filtrage de l'information

Intuitivement, le filtrage de l'information (FI) est une procédure située entre l'utilisateur et la source d'information qui laisse passer certaines d'information et en bloque d'autres, après un examen du contenu de l'information selon certaines règles prédéfinies. Le FI n'est pas une nouvelle notion : il existait avant l'arrivée de l'internet quand on mettait en œuvre le contrôle et la restriction d'information. Avant Internet les règles selon lesquelles on filtrait des informations ne pouvaient être que prédéfinies par d'autres au lieu de l'être par l'audience elle-même : le gouvernement, l'éditeur, le journaliste, l'école et l'université (Palme, 1998). Ce type de FI n'est pas exact ce dont on parle dans cette recherche même s'il continue à être d'actualité sur Internet (dans les mécanismes de modération des sites par exemple).

Avec l'avènement de l'Internet, le FI décrit une variété de processus fournissant aux utilisateurs des informations pertinentes pour leurs intérêts. (Belkin et Croft, 1992). Le FI est devenu un domaine de recherche actif dans le but de concevoir une approche systématique pour l'extraction des informations qu'un utilisateur trouve pertinentes parmi un grand flux d'informations (Canavese et Besser, 1994). Il est souvent relié à la diffusion sélective de l'information (DSI) (Luhn, 1958) et au Filtrage Collaboratif (FC) (Malone *et al.*, 1987).

En général, SFI est un mécanisme destiné à faire parvenir à l'utilisateur l'information qui l'intéresse directement (Belkin et Croft, 1992). L'objectif primordial d'un SFI est d'aider l'utilisateur à extraire, d'une manière efficace, les informations pertinentes parmi un grand flux d'informations entrant. Plus précisément, il existe deux objectifs un peu différents : 1) un filtrage des informations impertinentes (Khan et Card, 1997) et 2) une sélection des informations pertinentes à partir d'un ensemble plus large de possibilités, puis de les présenter dans un ordre de priorité (Malone *et al.*, 1987). La hiérarchisation de ces deux objectifs est différente dans les différents SFI.

III.2 Système de filtrage de l'information et système de recherche de l'information

Belkin et Croft indiquent que le SFI (Belkin et Croft, 1992) (Figure 11):

- est conçu pour traiter des bases de données semi-structurées ou non-structurées
- est capable de traiter des données gigantesques
- traite principalement des données textuelles mais aussi des données multimédias
- est basé sur le profil de l'utilisateur
- dont l'objectif est d'enlever des données impertinentes sur un flux entrant.

Les caractéristiques de SFI sont partagées également par les systèmes de recherche de l'information (SRI). En effet, Belkin considère le FI comme un type particulier de RI, mais il distingue des nuances entre eux :

- la RI est conçu pour satisfaire un besoin d'information en une seule fois (au bout d'une seule utilisation de systèmes) alors que le FI nécessite plusieurs utilisations régulières du système pour satisfaire des besoins d'information à long terme.
- Dans les SRI, le besoin d'information de l'utilisateur est exprimé dans la forme d'une requête tandis que dans les SFI, le besoin d'information est représenté par le profil de l'utilisateur.
- Les SRI traitent les bases de données relativement statiques alors que FI prend en charge un flux d'informations dynamiques qui arrive continuellement.
- Les SRI concernent la collection et l'organisation de l'information pertinente selon les requêtes de l'utilisateur alors que le SFI concerne la distribution de l'information pertinente sur l'utilisateur selon son profil.
- L'utilisateur d'un SRI est toujours très motivé pour chercher d'information alors que l'utilisateur d'un SFI ne l'est pas.

Il signifie que les SFI sont plus actifs puisqu'il s'agit de connaître leurs utilisateurs (leurs préférences, leurs besoins d'information et leurs centres d'intérêt *etc.*) ; de chercher et collecter toutes les informations pertinentes sur un flux dynamique ; et de les distribuer aux utilisateurs différents selon leur profil.

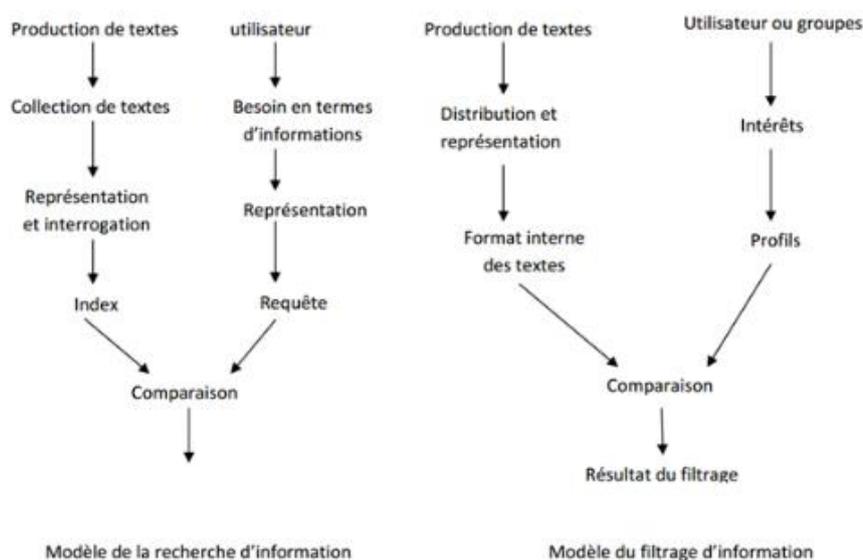


Figure 11 Modèle du filtrage d'information et modèle du RI selon (Belkin et Croft, 1992)

III.3 Typologie de système de filtrage de l'information

Le SFI se concentre sur le contrôle de l'information entrante. En 1987, Malone a proposé trois formes de FI : le filtrage cognitif, le filtrage économique et le filtrage social. Hanani et ses collègues présentent une typologie plus détaillée de SFI (Figure 12) qui classe les SFI selon quatre paramètres : l'initiative de l'opération, la position de l'opération, l'approche de filtrage et les méthodes d'acquisition de connaissances des utilisateurs (Hanani, Shapira et Shoval, 2001).

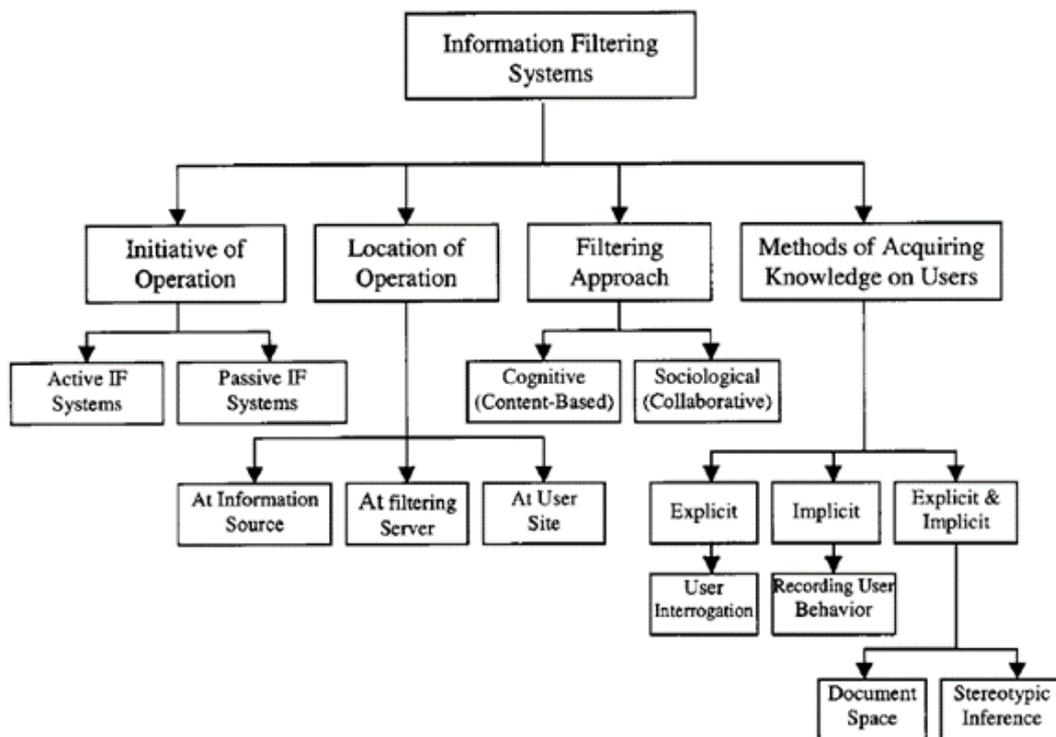


Figure 12 Classification de SFI selon Malone (1987)

III.3.1 L'initiative de l'opération

Selon l'objectif et la fonction, deux types de SFI peuvent être distingués : actif et passif. Le SFI actif désigne le système qui cherche activement les informations pertinentes afin de les envoyer aux utilisateurs qui en ont besoin selon leur profil. L'exemple est les « serveurs push (*Push Technologie*) ». De nos jours, nous pouvons ajouter un SR (système de recommandation) comme celui d'Amazon dans cette catégorie qui non seulement cherche activement les

informations pertinentes mais également étudie et essaie de comprendre activement ses utilisateurs. Le SFI passif désigne le système qui enlève les informations impertinentes sur un flux d'informations entrant. Contrairement à un système actif, il ne cherche pas activement des informations mais essaie de bloquer toutes les informations impertinentes entrantes ou/et de les classer dans différentes catégories définies par l'utilisateur. Le SFI passif est souvent utilisé dans les services de courrier électronique. Nous pouvons citer aussi les applications de blocage de publicité et les logiciels de contrôle parental *etc.*

III.3.2 La position de l'opération

Hanani, Shapira et Shoal distinguent 3 types de SFI selon la position où effectuent les filtres (Hanani, Shapira et Shoal, 2001) :

- à la source d'information : dans cette approche, un utilisateur doit déposer son profil à un fournisseur d'informations qui diffuse plus tard des informations pertinentes selon le profil en retour. L'exemple est l'Alert service lancé par Dialog (Bates, 1994). Nous pouvons observer aussi de nos jours cette forme de filtrage sur les sites d'actualité qui permettent aux utilisateurs de choisir des domaines d'actualité qui les intéressent (le sport, l'économie *etc.*)
- dans un serveur de filtrage : le serveur de filtrage est un intermédiaire entre les utilisateurs et les fournisseurs d'informations. Les utilisateurs déposent leur profil aux serveurs de filtrage qui leur distribuent, en retour, les informations pertinentes en provenance de fournisseur d'informations. Hanani et ses collègues citent les systèmes de filtrage SIFT (*Stanford information Filtering tool*) (Yan et Garcia-Molina, 1999), Rama (Agrawal et Srikant, 1994) *etc.* comme des exemples. Nous pouvons ajouter aussi les services de RSS comme autre exemple.
- au site d'utilisateur local : ce type de système est généralement passif et travaillent dans le but d'enlever les informations impertinentes entrantes puis de classer les informations selon leur pertinence. Ils sont particulièrement utilisés dans les services de courrier électronique, comme Outlook, Gmail par exemple.

III.3.3 Approches du filtrage

Selon la typologie de Malone (Malone *et al.*, 1987), le filtrage cognitif et le filtrage social sont les approches plus étudiées et utilisées. En réalité, la plupart des SFI sont hybrides car ils emploient plusieurs méthodes de filtrage en même temps.

Le filtrage cognitif

Le filtrage cognitif aussi connu sous le nom de *filtrage basé sur le contenu*. Il « caractérise le contenu du message et les besoins en information des destinataires potentiels du message, puis utilise ces représentations pour faire correspondre intelligemment les messages aux destinataires » (Malone *et al.*, 1987). Les documents entrants sont indexés par des termes puis comparés au profil de l'utilisateur représenté sous la forme d'un ensemble de thèmes décrivant ses centres d'intérêt. Le résultat de la comparaison est le filtrage des documents qui correspondent le mieux à son profil. Dans ce genre de système, chaque utilisateur opère indépendamment des autres comme s'il était en train de faire une recherche d'information classique (Gallardo-Lopez, 2005). La sélection se base sur l'analyse de leur contenu, il n'est proposé à l'utilisateur que des documents sur les mêmes thèmes que ceux qui l'intéressent. La fonction principale du filtrage par le contenu consiste en la sélection de documents pertinents par rapport au profil et en la mise à jour du profil en fonction des rétroactions de l'utilisateur concernant les documents qu'il a reçus. La mise à jour se fera par l'insertion des nouveaux thèmes abordés par les documents jugés pertinents.

Le filtrage social

Le filtrage social, connu aussi sous le nom de filtrage collaboratif, est défini par Malone comme le filtrage qui « travaille en soutenant les interrelations personnelles et organisationnelles des individus dans une communauté » (Malone *et al.*, 1987). Dans la plupart des communautés partageant des centres d'intérêt, les utilisateurs se recommandent ou s'échangent régulièrement des documents. Le filtrage social consiste à filtrer les documents du

flux entrant en se basant sur l'opinion de chaque utilisateur de la communauté. Tout document qu'un utilisateur juge intéressant sera diffusé à l'ensemble des utilisateurs ayant eu des opinions similaires par le passé. Le profil ne se présente plus sous la forme classique d'un ensemble de thème pondéré comme dans le filtrage de contenu, mais plutôt sous une forme d'un ensemble d'évaluations des documents que l'utilisateur a faites dans le passé. (Gallardo-Lopez, 2005).

III.3.4 Les méthodes d'acquisition de connaissance des utilisateurs

Dans les SFI, les méthodes diverses sont employées pour acquérir les connaissances de leurs utilisateurs (Stadnyk et Kass, 1992). En général, trois approches peuvent être distinguées : l'approche explicite, l'approche implicite et l'approche hybride.

L'approche explicite

La technique la plus populaire de l'approche explicite est l'interrogation de l'utilisateur (*User Interrogation*). Le système exige généralement que leurs utilisateurs décrivent leurs domaines d'intérêt ou fournissent d'autres paramètres pertinents. Dans la pratique, diverses mesures sont employées. Par exemple, certains systèmes de filtrage fournissent à l'utilisateur un ensemble de profils prédéfini parmi lesquels il peut choisir le profil le plus approprié. D'autres systèmes fournissent à l'utilisateur un ensemble de termes qui représente chaque domaine, à partir duquel il peut construire un profil personnel. Certains systèmes plus sophistiqués permettent aux utilisateurs de déterminer les termes et leurs poids d'importance. Il y a bien d'autres méthodes utilisées pour que les SFI puissent connaître leurs utilisateurs en les interrogeant directement. Cependant, cette approche nécessite davantage d'efforts de la part des utilisateurs de sorte qu'ils ne sont pas toujours disposés à répondre.

L'approche implicite

L'approche implicite ne nécessite pas la participation active de l'utilisateur. En revanche, le système enregistre la réaction de l'utilisateur à chaque document entrant dans le but de

connaître ses préférences et ses besoins. Dans la pratique, les comportements de l'utilisateur servent à indiquer ses options sur les documents, par exemple, le temps que l'utilisateur a passé à lire un document et s'il enregistre, supprime, imprime ou transfère les documents *etc.* Néanmoins, cette méthode implicite risque d'être biaisée (par exemple, perturbation pendant la lecture) et ne doit être utilisée qu'en complément d'une sorte d'interrogation de l'utilisateur.

L'approche hybride

L'une des méthodes de l'approche hybride est « l'espace de document (*Document Space*) ». Cette méthode crée un espace de documents que l'utilisateur a précédemment jugé pertinent. Les nouveaux documents entrants sont testés et triés selon leur similitude avec les documents de cet espace. Si la similitude du nouveau document dépasse un certain seuil de pertinence, ce document est considéré comme pertinent. L'utilisateur n'a pas besoin de définir un profil, mais seulement d'évaluer la pertinence du document. L'un des pièges de cette méthode est qu'une distorsion continue des intérêts de l'utilisateur peut se produire si certains domaines d'intérêt ne sont pas couverts par l'espace de documents initial. De nombreuses autres méthodes existent qui combinent l'approche explicite et l'approche implicite. En bref, l'approche hybride nécessite une participation minimale de l'utilisateur.

Chapitre IV Système de recommandation personnalisé

Le SRI et le SFI partagent beaucoup de points communs car leur but fondamental est d'établir un lien entre les utilisateurs et les informations pertinentes *pour eux*. Néanmoins, des différences existent. L'une des différences les plus importantes est la motivation de l'utilisateur qui est une différence décisive entre ces deux systèmes au début de leur conception. Dans un SRI, l'utilisateur est toujours fortement motivé, il connaît précisément son besoin d'informations et cherche activement les informations susceptibles de le satisfaire. Alors que dans un SFI, l'utilisateur est souvent moins motivé, son besoin d'informations est souvent trop général et vague. Il indique plusieurs domaines dans son profil au début et il attend que le système enlève les informations impertinentes et lui distribue les informations pertinentes.

Dans le chapitre précédent, nous avons présenté la typologie de SFI proposée par Hanani et ses collègues. Ce type de système est notamment basé sur les algorithmes qui cherchent à comprendre le besoin d'informations de l'utilisateur et à calculer les similarités des informations par rapport à ses besoins d'informations afin de faire des recommandations. Il est largement utilisé dans le commerce électronique et dans les RSN. Comparativement, le SR est plus complexe et plus compliqué que le SFI passif. Les chercheurs et les entreprises portent davantage leur attention sur les SR et de nombreux approches et algorithmes ont été développés.

IV.1 Définition et typologie du système de recommandation

Le SR est apparu comme un domaine de recherche indépendant au milieu des années 1990 (Adomavicius et Tuzhilin, 2005). Dans le contexte de marketing, les SR sont des logiciels et des techniques qui fournissent des suggestions d'items utiles à l'utilisateur. Ces suggestions cherchent à aider les utilisateurs à prendre une décision dans les processus de consommation. Cela veut dire qu'un SR a pour objectif de fournir à l'utilisateur des ressources pertinentes en fonction de ses préférences. Par conséquent, le SR se prouve sa valeur pour les internautes face à la surcharge d'information et deviennent l'un des outils les plus puissants et les plus populaires dans le commerce électronique (Ricci *et al.*, 2011).

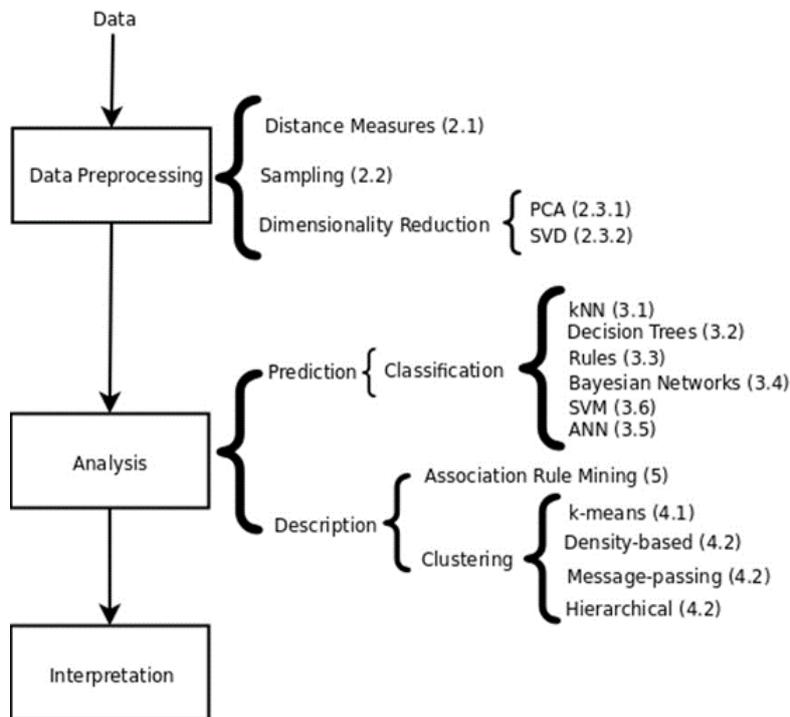


Figure 13 Les étapes et les méthodes principales d'une exploration des données (Ricci et al., 2011)

Dans le but d'identifier plus efficacement des items qui intéressent un utilisateur parmi un vaste ensemble de choix, le SR utilise les opinions des autres utilisateurs qui partagent les mêmes goûts que cet utilisateur (Resnick et Varian, 1997). Il faut donc étudier le profil de l'utilisateur (Montaner, López et De La Rosa, 2003) et « l'évaluation (*rating*) » faite par l'utilisateur dans le passé (Adomavicius et Tuzhilin, 2005). Il s'agit ensuite de travaux de fouilles de données consistant en 3 étapes : le prétraitement de données, l'analyse de données et l'interprétation de résultat (Figure 13). Selon Negre (Negre, 2015), « le prétraitement de données est souvent nécessaire pour aborder l'étape d'analyse avec des données « propres » et adaptées aux techniques qui seront utilisées pour l'analyse. » « L'analyse de donnée a deux objectifs distincts : prédictif et descriptif... Dans le premier cas, il s'agira d'utiliser des techniques de classification... ; dans le second cas, il s'agira d'utiliser des techniques à base de règle d'association ou de clustering ».

En fonction de différents algorithmes de filtrage, les SR peuvent être classifiés dans 3 approches : le filtrage collaboratif, le filtrage basé sur le contenu et le filtrage hybride (Figure 14).

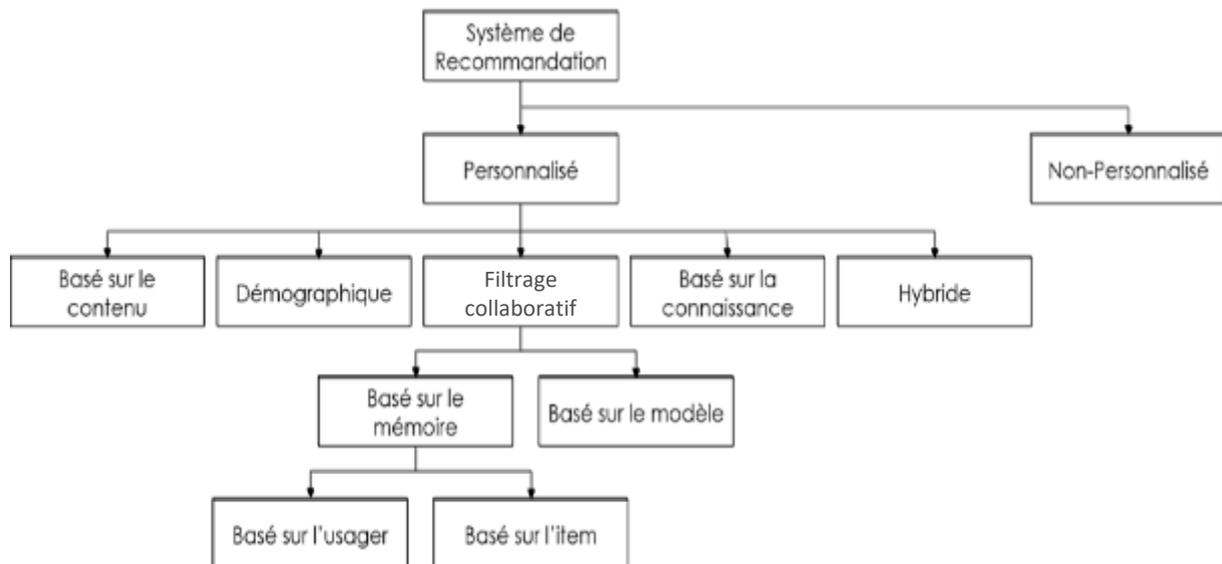


Figure 14 les approches principales du SR

IV.2 Approche de filtrage basé sur le contenu

Le SR basé sur le contenu propose à l'utilisateur des items qui ressemblent à ceux que l'utilisateur a aimés dans le passé (Adomavicius et Tuzhilin, 2005). L'idée de base est que les préférences de l'utilisateur soient assez stables et que les caractéristiques d'items puissent être extraites automatiquement (Jannach *et al.*, 2010). Cette approche traite les items et les caractéristiques d'items. Elle concentre sur la similarité entre les items en employant des techniques diverses.

Le SR basé sur le contenu s'appuie sur l'item et le profil de préférence de l'utilisateur (Burke, 2007). De manière générale, le système analyse les caractéristiques des items et lui recommande ceux dont les caractéristiques sont similaires à ceux qu'il a aimés d'après son profil. Ce profil est une représentation structurée de centres d'intérêt dans le but de recommander de nouveaux items (Lops, de Gemmis et Semeraro, 2011). Il s'appuie sur des évaluations effectuées par cet utilisateur sur un ensemble d'items (Béchet, 2012) qui représentent ses préférences et sur les intérêts qu'il a explicitement définis. Le système fait coïncider les caractéristiques de l'item avec le profil de l'utilisateur pour décider de sa pertinence. Ainsi le système recommande des informations similaires à celles que l'utilisateur a déjà appréciées sans prendre en compte les avis des autres utilisateurs (Peis, Morales del Castillo et Delgado-López, 2008). La tâche de recommandation consiste donc à déterminer quels items coïncident

le mieux avec les préférences de l'utilisateur (Negre, 2015).

Cette approche de filtrage montre sa grande efficacité du filtrage d'information si un profil représente précisément les préférences de l'utilisateur (Lops, de Gemmis et Semeraro, 2011). Mais il montre également des limites comme « sur-spécialisation ¹³⁰», problème de démarrage à froid pour les nouveaux utilisateurs *etc.*

IV.2.1 Processus du SR basé sur le contenu

Essentiellement, le processus du SR basé sur le contenu est un travail d'appariement. Dans un tel système représente les items dans une forme de description (un profil) en montrant leurs caractéristiques (par exemple, on pourrait utiliser le titre, le genre, le directeur/ la directrice, les acteurs/ actrices *etc.*, pour la recommandation de film). Il construit ensuite un modèle ou un profil qui représente les préférences et les centres d'intérêt de l'utilisateur basé sur les évaluations qu'il a effectuées. Il est capable de recommander de nouveaux items en analysant leurs caractéristiques puis les comparant avec le profil d'utilisateur. Negre illustre ce processus dans la Figure 15.

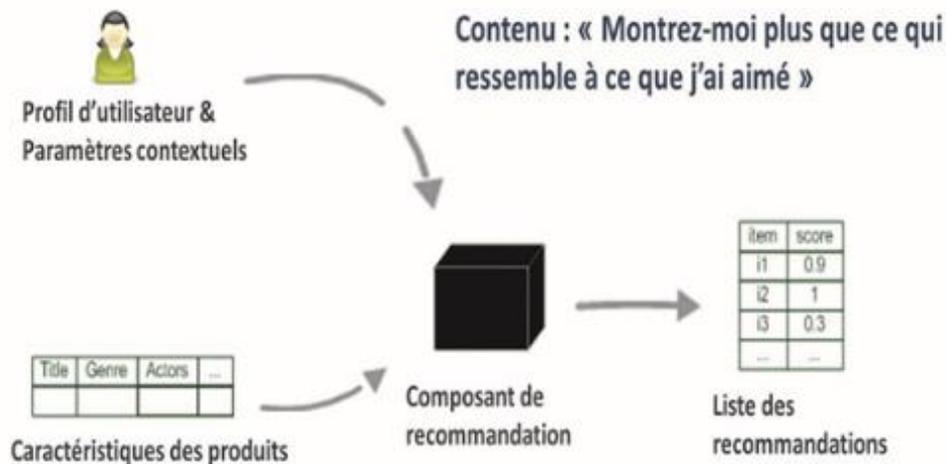


Figure 15 le SR basé sur le contenu vu comme une boîte noire (Negre, 2015) traduction de (Jannach et al., 2010)

¹³⁰ Les recommandations se limitent aux items homogènes

Lops, De Gemmis et Semeraro montrent la partie interne du « composant de recommandation » de ce processus et indiquent les 3 étapes principales de ce processus sont traitées respectivement par les 3 composants : l'analyse du contenu, l'apprentissage du profil et le composant du filtrage (Lops, de Gemmis et Semeraro, 2011) (Figure 16).

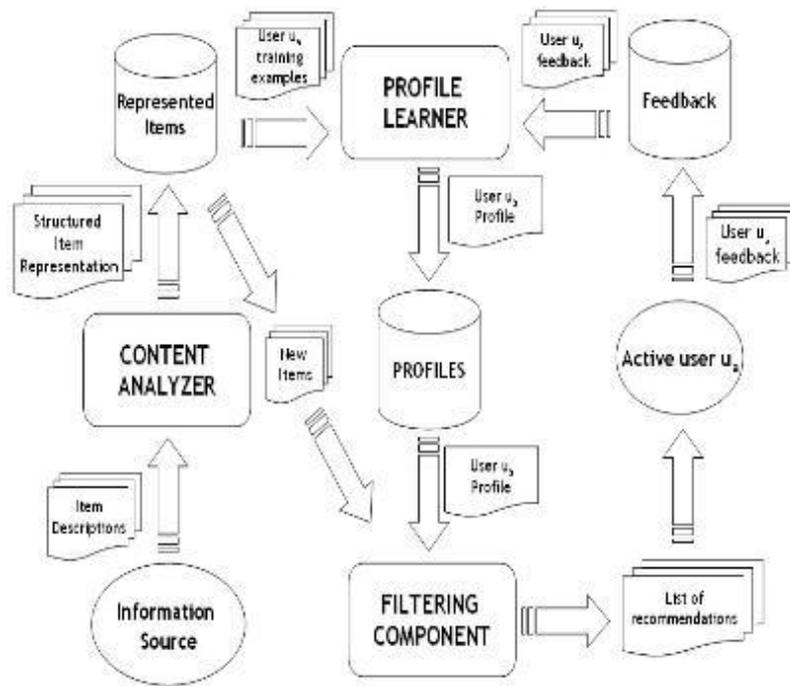


Figure 16 Une architecture de haut niveau de ce processus (Lops, de Gemmis et Semeraro, 2011)

La première étape est effectuée par le composant « l'analyse du contenu » qui traite les items provenant de la source d'information. Il consiste à analyser les contenus des nouveaux items et à extraire des informations pertinentes dans le but de représenter les contenus des items. Dans cette étape, le système recourt aux techniques de la RI (Adomavicius et Tuzhilin, 2005; Lops, de Gemmis et Semeraro, 2011).

Opérée par le composant « l'apprentissage du profil », la deuxième étape cherche à constituer le profil de l'utilisateur. Cette étape a pour l'objectif de comprendre les préférences de l'utilisateur en collectant des données représentatives. Ces données représentatives aussi appelées des rétroactions (*feedback*) peuvent être distinguées entre celles qui sont implicites (les comportements de l'utilisateur) et celles qui sont explicites (les évaluations faites par l'utilisateur). Les rétroactions aident le système à discriminer les items que l'utilisateur a aimés de ceux qu'il n'a pas aimés dans le passé. Dans cette étape, les techniques de l'apprentissage

automatique (*Machine Learning*) sont souvent utilisées (Adomavicius et Tuzhilin, 2005; Jannach *et al.*, 2010; Lops, de Gemmis et Semeraro, 2011).

La dernière étape est exécutée par « le component du filtrage » qui calcule la similarité entre des nouveaux items et ceux que l'utilisateur a évalués dans le but de proposer des recommandations. Plusieurs méthodes peuvent être utilisées pour calculer la similarité (Dice, 1945; Rocchio, 1971; Salton, Wong et Yang, 1975; Chuang et Sher, 1993; Billsus et Pazzani, 2000; Herlocker *et al.*, 2004; Pazzani et Billsus, 2007; Quinlan, 2014).

IV.2.2 Représentation d'item

La manière la plus simple de décrire un catalogue d'items est d'avoir une liste explicite des caractéristiques (également appelés l'attribut, profil d'item, la propriété *etc.*) de chaque item. Des items pourraient être recommandés à l'utilisateur en fonction de leurs caractéristiques. Quand le profil de l'utilisateur est exprimé sous forme d'une liste d'intérêts basée sur les mêmes caractéristiques, la tâche de recommandation consistera donc à faire coïncider les caractéristiques de l'item et le profil de l'utilisateur (Jannach *et al.*, 2010; Negre, 2015).

Beaucoup de systèmes de filtrage basé sur le contenu se concentrent sur des recommandations d'items contenant de l'information textuelle (Adomavicius et Tuzhilin, 2005).

Une méthode classique est d'extraire une liste d'informations (mots-clés) pertinentes à partir des informations textuelles contenues dans l'item lui-même. Pour traiter les items multimédias (le film, la musique et l'image *etc.*), les métadonnées et les tags sont beaucoup plus utilisés bien qu'ils ne soient pas les « contenus » d'items (Jannach *et al.*, 2010). Les techniques les plus utilisées sont le modèle vectoriel des mots-clés avec la pondération TD-IDF (Salton, Wong et Yang, 1975). Dans le but d'améliorer cette méthode, les chercheurs proposent également des techniques comme l'analyse sémantique en utilisant des ontologies ou des sources de connaissances encyclopédiques.

Modèle d'espace vectoriel et pondération TF-IDF

La plupart des SR basé sur le contenu a recours aux techniques de RI, parmi lesquelles, le modèle d'espace vectoriel (*Vector Space Model*, VSM) est la plus connue. Le modèle d'espace vectoriel est une méthode pour représenter les items (documents). Ce modèle est originellement utilisé dans la RI pour traiter les documents textuels, notamment pour la recherche documentaire, la classification ou le filtrage de données. Ayant été introduit dans le SR, il est utilisé non seulement pour traiter les items textuels mais aussi les items multimédias.

Dans un modèle d'espace vectoriel, les items sont représentés sous forme vectorielle dans un espace à n dimensions (n axes). Chaque dimension correspond à un terme (soit un mot, soit une phrase) dans un ensemble de la collection d'items. Dans ce cas, chaque item peut être représenté par un vecteur des termes. Soit un l'ensemble de documents $D = \{d_1, d_2, d_3 \dots d_n\}$ et l'ensemble de terme $T = \{t_1, t_2, t_3 \dots t_n\}$, un document $d_i = \{p_{1i}, p_{2i}, p_{3i} \dots p_{ni}\}$ dont le p_{ki} est la pondération du terme t_k dans le document d_i .

Cependant, tous les termes de l'item n'ont pas la même importance. Dans le but de représenter l'item de manière plus précise, il nécessite des techniques astucieuses pour pondérer les termes. Le TF-IDF (*Term Frequency-Inverse Document Frequency*) est l'une des pondérations les plus utilisées (Salton, Wong et Yang, 1975; Salton, 1989).

Term Frequency

La fréquence du terme (*Term Frequency*, TF) indique le nombre d'occurrences du terme dans un document. L'idée de base est que plus un terme est fréquent dans un document, plus il est représentatif de son thème. TF est un nombre absolu, c'est-à-dire un compte simple du nombre d'occurrences du terme dans un document. Pourtant il arrive souvent que le nombre d'occurrences du terme soit plus élevé dans un document long que dans un document court. Pour éviter ce biais, certaines normalisations de cet indicateur sont nécessaires (Salton and Buckley, 1988; Chakrabarti, 2002). Ainsi, TF (i, j) représente la fréquence du terme t_i dans un document d_j .

Inverse Document Frequency

La fréquence inverse de document (*Inverse Document Frequency*, IDF) indique la rareté d'un terme par l'inverse de sa fréquence dans l'ensemble de documents. L'idée essentielle est que plus un terme est rare dans la collection, plus sa présence dans un document est importante. Cet indicateur vise à réduire la pondération des termes qui apparaissent souvent dans tous les documents. En revanche, un poids plus important doit être donné aux termes ceux qui se présentent seulement dans certains documents particuliers parce que ces termes sont considérés comme plus discriminants. Soit t_i un terme, une collection D . Le nombre total de documents est $|N_D|$, le nombre de documents de la collection D avec t est n_i .

$$IDF(i) = \log \frac{|N_D|}{n_i}$$

La pondération TF-IDF du terme t_i dans un document d_j est calculé

$$TF - IDF(t_i, d_j) = TF(t_i, d_j) \times IDF(t_i)$$

Soit :

$$TF - IDF(t_i, d_j) = TF(t_i, d_j) \times \log \frac{|N_D|}{n_i}$$

Les vecteurs pondérés par le TF-IDF sont utilisés souvent comme la base de calcul des mesures de la similarité entre les items. Le TF-IDF permet d'évaluer et d'identifier l'importance d'un terme dans un document en donnant moins de poids aux termes banals (e.g., mot-vide) et en donnant plus de poids aux termes noyaux. Néanmoins, il n'est pas capable de traiter les documents dont les termes et les concepts noyaux ne se présentent pas. Certaines techniques additionnelles sont souvent utilisées dans le but d'améliorer la pondération TF-IDF telles que le mot-vide et la lemmatisation, la réduction d'une phrase à un terme *etc.*

Dans la pratique, plusieurs systèmes de ce type ont été développés concernant des domaines très variés. Pour la recommandation des sites web, on peut noter Letizia (Lieberman, 1995), Personal WebWatcher (Mladenic, 1999a; Mladenic, 1999b), Syskill & Webert (Pazzani, Muramatsu et Billsus, 1996; Pazzani et Billsus, 1997), ifWeb (Asnicar et Tasso, 1997),

Amalthea (Moukas, 1997), et WebMate (Chen et Sycara, 1998), Quant à la recommandation des actualités, on trouve NewT (Sheth et Maes, 1993), PSUN (Sorensen et McElligott, 1995), INFOrmer (O’Riordan et Sorensen, 1997), NewsDude (Billsus et Pazzani, 1999), Daily Learner (Billsus et Pazzani, 2000) et YourNews (Ahn *et al.*, 2007). Dans le domaine de la musique, Pandora est le plus représentatif. On peut trouver également LIBRA (Mooney et Roy, 2000), Re:Agent (Boone, 1998), Citeseer (Bollacker, Lawrence et Giles, 1998), INTIMATE (Mak, Koprinska et Poon, 2003), Movies2GO (Mukherjee *et al.*, 2001) et d’autres systèmes hybrides.

Cette approche basant purement sur le mot-clé est souvent considérée moins intelligent, car il subit des limites évidentes : il ne comprend pas vraiment la signification du terme et il néglige le contexte du terme. Alors elle n’est pas capable de recommander à l’utilisateur les items dont les contenus sont similaires mais qui emploient les termes différents. D’ailleurs, il est difficile de traiter les termes de polysémie et les synonymes. Dans ce cas, certaines approches ont été développés qui utilise l’analyse sémantique.

IV.2.3 Mesures de Similarité

Après avoir choisi une propre méthode à représenter les items, il faut des bonnes méthodes à mesurer la similarité des items. Plusieurs approches existent parmi lesquelles, la distance euclidienne, la similarité cosinus(Lang, 1995; Billsus et Pazzani, 1998, 2000; Sarwar *et al.*, 2000b), le coefficient de corrélation de Pearson (Lee Rodgers et Nicewander, 1988; Good *et al.*, 1999) et l’indice de Dice (Dice, 1945) sont les plus fréquemment employées.

Dans les sections suivantes, nous allons indiquer a et b pour désigner deux items qui sont respectivement représentés par deux vecteurs à n dimensions : a (a_1, a_2, \dots, a_n) et b (b_1, b_2, \dots, b_n) dont le a_k et le b_k sont respectivement la k -ième dimension (ou caractéristique, terme, attribut *etc.*) de l’item a et b . \bar{a} et \bar{b} représentent la moyenne d’ a et d’ b .

La distance de similarité

La méthode la plus simple est de calculer la similarité entre deux items comme la distance entre leurs représentations vectorielles ramenées à un seul point. **La distance euclidienne** est

l'une des mesures la plus utilisées dont la formule est :

$$Sim_{euclidienne}(a, b) = \sqrt{\sum_{k=1}^n (a_k - b_k)^2}$$

où n est le nombre total de dimensions (termes, attributs ou caractéristiques), a_k et b_k sont respectivement le k -ième termes d'item a et b .

La distance de Minkowski est considérée comme une généralisation de la distance euclidienne. La formule est :

$$Sim_{Minkowski}(a, b) = \sqrt[p]{\sum_{k=1}^n (a_k - b_k)^p}$$

Quand le $p=1$, c'est la **distance de Manhattan** (ou **distance L1**) ; lorsque $p=2$, c'est la **distance euclidienne** (ou **distance L2**) ; quand $p \rightarrow \infty$, c'est la **distance de Chebyshev** (ou distance **Lmax**).

La similarité cosinus

Une autre mesure est la similarité cosinus qui est fréquemment utilisée dans le domaine de RI et celui de fouille de textes en tant que mesure de ressemblance entre deux documents. Cette technique permet de calculer le cosinus de l'angle entre les deux vecteurs à n dimensions afin de mesurer la similarité entre eux. Dans le SR, le vecteur peut représenter non seulement l'item mais également l'utilisateur. Conséquemment, elle est largement utilisée dans toutes les approches de SR. Soit deux vecteurs \vec{a} et \vec{b} qui représentent deux items a et b , la similarité entre eux est calculée :

$$Sim_{cosinus}(\vec{a}, \vec{b}) = \cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$$

$$= \frac{\sum_{i=1}^n a_i \times b_i}{\sqrt{\sum_{i=1}^n (a_i)^2} \times \sqrt{\sum_{i=1}^n (b_i)^2}}$$

où θ désigne l'angle entre les deux vecteurs, le symbole « \cdot » est le produit scalaire des vecteurs et le $|\vec{a}|$ est la norme de vecteur \vec{a} qui est la longueur euclidienne de ce vecteur. La valeur $\cos\theta$, à savoir, la valeur de similarité est comprise dans l'intervalle $[0, 1]$. La valeur 0 indique des vecteurs indépendants (orthogonaux), alors que la valeur 1 indique des vecteurs similaires (colinéaires de coefficient positif).

La similarité entre des items ou des utilisateurs peut aussi être calculée en mesurant la corrélation linéaire entre eux, à savoir, leur corrélation. Plusieurs coefficients de corrélation existent parmi lesquels le coefficient de corrélation est le plus utilisé dans le SR. Ce coefficient calcule la similarité entre des items ou des utilisateurs comme le cosinus de l'angle entre leurs représentations vectorielles centrées réduites. La formule est :

$$\begin{aligned} Sim_{pearson}(a, b) &= \frac{Cov(a, b)}{\sigma_a \times \sigma_b} \\ &= \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum (a_i - \bar{a})^2} \sqrt{\sum (b_i - \bar{b})^2}} \\ &= Sim_{cosinus}(a - \bar{a}, b - \bar{b}) \end{aligned}$$

L'indice de Dice

L'indice de Dice est une autre méthode statistique qui mesure la similarité entre deux items en se basant sur le nombre de termes communs aux deux vecteurs à comparer. La formule est :

$$Sim_{dice}(a, b) = \frac{2 \times |motsclés(a) \cap motsclés(b)|}{|motsclés(a)| + |motsclés(b)|}$$

La mesure de similarité est au cœur de SR. Toutes les approches mesurent la similarité. L'approche de FC basé sur l'utilisateur (*user-user*) mesure la similarité entre les préférences des utilisateurs, l'approche de FC basé sur l'item (*item-item*) mesure la similarité entre les notes

des items données par les utilisateurs. Quant à l'approche basée sur le contenu, elle mesure la similarité entre les caractéristiques d'items.

IV.2.4 Prédiction selon profil de l'utilisateur

Le SR basé sur le contenu recommande à l'utilisateur des items qui sont similaires à ceux qu'il a appréciés dans le passé. Il se concentre sur la similarité des caractéristiques entre les items. La tâche est alors de mesurer à quel niveau les caractéristiques d'un nouvel item sont similaires à celles des items que l'utilisateur a aimés dans le passé. Il est nécessaire d'avoir les informations sur les préférences (les évaluations) de l'utilisateur sur des items précédents, plus précisément, ceux qu'il a aimés et ceux qu'il n'a pas aimés. C'est une tâche de catégorisation binaire : chaque item doit être classifié dans la catégorie pertinente et impertinente selon le profil de l'utilisateur. Ainsi, l'ensemble de catégories est $C = \{c+, c-\}$, où $c+$ représente tous les items positifs (que l'utilisateur aime) et $c-$ est tous les items négatifs (que l'utilisateur n'aime pas) (Lops, de Gemmis et Semeraro, 2011). Cela signifie que des techniques propres pour apprendre le profil de l'utilisateur sont nécessaires. En effet, il s'agit de techniques de catégorisation de textes et d'apprentissage numérique. Plusieurs méthodes existent: la méthode de Rocchio (Rocchio, 1971), la méthode probabiliste, les approches bayésiennes (Pazzani et Billsus, 2007), les arbres de décision (Quinlan, 2014), les techniques basées sur la similarité des espaces vectoriels (k NN) (Billsus et Pazzani, 2000).

IV.2.4.1 Algorithme de Rocchio

L'algorithme de Rocchio est originellement développé dans le SRI SMART (Salton, 1971) au début des années 1970. Reposant sur l'espace vectoriel, c'est un algorithme classique pour la mise en œuvre de la technique de rétroaction de pertinence (*Relevance Feedback*). L'idée de base est d'incorporer l'utilisateur dans le processus de recherche afin d'améliorer le résultat final (Manning, Raghavan et Schütze, 2008). Elle est fondée sur l'hypothèse que l'utilisateur soit capable de distinguer les items pertinents de ceux qui ne le sont pas. Plus concrètement, le système permet à l'utilisateur de donner des rétroactions sur la pertinence des items récupérés par rapport à leurs besoins d'information. Ainsi, le système acquiert des rétroactions explicites

qui servent ensuite à affiner progressivement les requêtes de l'utilisateur afin d'améliorer les résultats ultérieurs.

Dans la pratique, le système différencie les items (documents) déjà évalués dans deux groupes D_r et D_{nr} . Le groupe D_r comprend tous les items (documents) marqués pertinents par l'utilisateur et le groupe D_{nr} comprend tous ceux qui sont impertinents. Ensuite, le vecteur moyen des deux groupes est calculé. Il est considéré comme le barycentre du cluster des items (documents) pertinents ou impertinents (Figure 17).

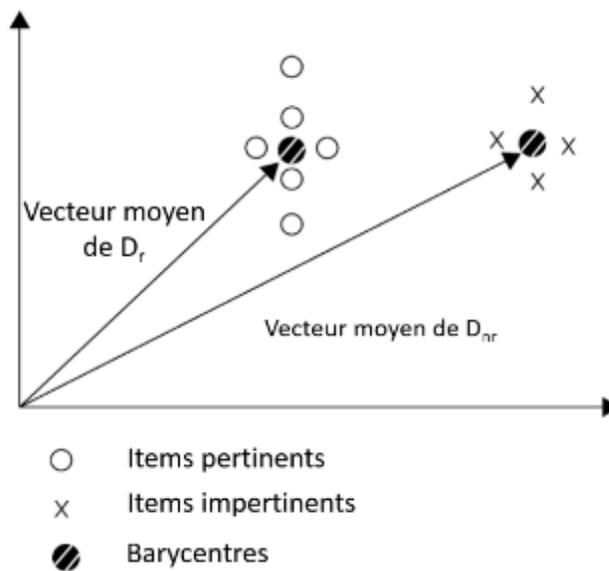


Figure 17 Vecteurs moyens pour les documents pertinents et non pertinents.

Dans le SRI, le but est de faire la requête originale Q_i approcher du barycentre des documents pertinents (Figure 18). La requête originale Q_i est représentée par le vecteur \vec{Q}_i , un vecteur multidimensionnel des termes. La requête modifiée Q_{i+1} est représentée par le vecteur \vec{Q}_{i+1} qui peut être calculé par l'algorithme de Rocchio en fonction des rétroactions de l'utilisateur. La formule de Rocchio est :

$$\vec{Q}_{i+1} = \alpha \cdot \vec{Q}_i + \beta \cdot \left(\frac{1}{|D_r|} \cdot \sum_{\vec{d}_j \in D_r} \vec{d}_j \right) - \gamma \cdot \left(\frac{1}{|D_{nr}|} \cdot \sum_{\vec{d}_k \in D_{nr}} \vec{d}_k \right)$$

où le \vec{Q}_i est le vecteur de la requête originale, le \vec{Q}_{i+1} est le vecteur de la requête modifiée. D_r et D_{nr} sont respectivement l'ensemble de documents pertinents et impertinents. Le \vec{d}_j et le \vec{d}_k sont

respectivement le vecteur du j-ième et k-ième document dans le D_r et le D_{nr} . Le α , β et γ sont les poids de chaque terme dont les valeurs sont définies en fonction de l'importance que l'on souhaite donner à chaque terme. Plus précisément, l' α définit l'importance de la dernière requête, le β et le γ indiquent respectivement l'importance des rétroactions positives et négatives. En fait, les rétroactions positives sont considérées plus importantes que celles de négatives, c'est-à-dire $\gamma < \beta$. Dans les systèmes de RI, les valeurs raisonnables sont $\alpha=1$, $\beta=0,75$ et $\gamma=0,15$ (Manning et al., 2008) ou $\alpha=1$, $\beta=2$ et $\gamma=0,25$ (Buckley et al., 1994). Ils sont responsables du mouvement (ou la modification) du \vec{Q}_i vers le \vec{Q}_{i+1} .

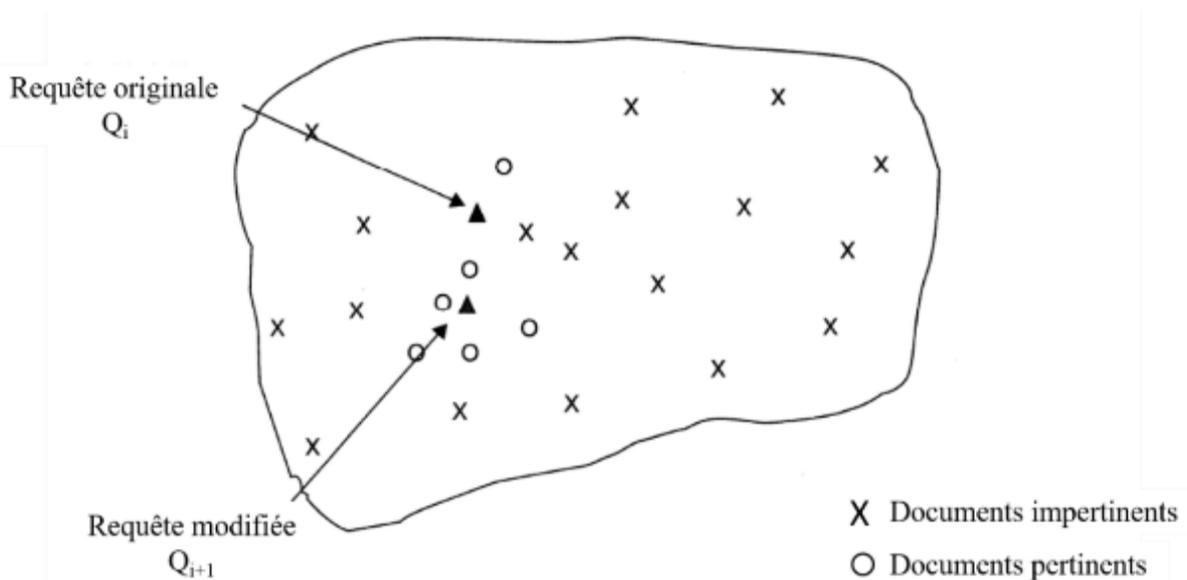


Figure 18 Faire Q_i approcher aux documents pertinents (Manning, Prabhakar et Hinrich, 2008)

La technique rétroaction de pertinence et l'algorithme de Rocchio ont été intégrés et largement utilisés dans les SR basés sur le contenu. L'utilisateur doit évaluer explicitement la pertinence des items en fonction de ses préférences. Ce type de rétroaction explicite est utilisé ensuite comme un classificateur qui affinera progressivement le profil de l'utilisateur et entrainera l'algorithme d'apprentissage inférant le profil de l'utilisateur (Lops, de Gemmis et Semeraro, 2011).

L'algorithme de Rocchio est souvent utilisé pour apprendre et inférer le profil (ou les préférences) de l'utilisateur. Dans un SR basé sur le contenu, les items sont représentés par des vecteurs multidimensionnels dont chaque dimension correspond à un terme de l'item. Ainsi les

items dont les caractéristiques sont similaires procèdent de vecteurs similaires qui sont calculés selon la pondération TF-IDF (Lops, de Gemmis et Semeraro, 2011). Le profil de l'utilisateur représenté par le vecteur \vec{w}_u peut être calculé :

$$\vec{w}_u = \beta \cdot \left(\frac{1}{|I_r|} \cdot \sum_{\vec{i}_j \in I_r} \vec{i}_j \right) - \gamma \cdot \left(\frac{1}{|I_{nr}|} \cdot \sum_{\vec{i}_k \in I_{nr}} \vec{i}_k \right)$$

où I_r et I_{nr} sont respectivement l'ensemble des items pertinents et impertinents que l'utilisateur a déjà évalué. i_j et i_k sont le j -ième et le k -ième items dans I_r et I_{nr} . Ainsi, la similarité entre le vecteur de l'utilisateur (soit \vec{w}_u) et le vecteur d'un nouvel item (soit \vec{i}_l) peut être calculé. Les items seront recommandés si leurs vecteurs sont similaires à ceux de l'utilisateur. La formule semble intuitive et l'algorithme est simple (Jannach *et al.*, 2010). Pourtant, il n'y a aucun fondement théorique ni la performance ou la convergence garantie (Pazzani et Billsus, 2007).

IV.2.4.2 Méthode probabiliste et approche bayésienne

Une autre méthode souvent adaptée dans le SR basé sur le contenu est la méthode probabiliste. L'approche probabiliste traite la tâche de recommandation comme un problème de catégorisation qui classifie des items avec des catégories prédéfinies (Tan, Steinbach et Kumar, 2006). Dans le cas de SR basés sur le contenu, le classement concerne souvent deux catégories : la catégorie pertinente et la catégorie impertinente en fonction du profil (ou des préférences) de l'utilisateur (Pazzani et Billsus, 1997; Jannach *et al.*, 2010).

Basé sur les données générées dans le passé, cette approche crée un modèle probabiliste qui estime la probabilité a posteriori $P(c|d)$ (*i.e.* la probabilité de l'item d appartenant à la catégorie c). Cette estimation repose sur les probabilités a priori : $P(c)$ (*i.e.* la probabilité de voir un item dans la catégorie c), $P(d|c)$ (*i.e.* la probabilité de voir l'item d dans la catégorie c) et $P(d)$ (*i.e.* c). L'item d sera attribué à la catégorie avec la plus grande probabilité c . La Théorème de Bayes est utilisée afin de calculer la probabilité a posteriori $P(c|d)$:

$$P(c|d) = \frac{P(d|c) \times P(c)}{P(d)}$$

Dans cette formule, la $P(d)$ est souvent supprimée comme elle est une valeur constante, c'est-à-dire que la probabilité de voir l'item d est égale pour tous les catégories c . La $P(c)$ peut être estimée grâce aux données d'entraînement qui sont souvent les données précédentes, plus concrètement, les notes des items données par l'utilisateur précédemment dans le contexte de SR basé sur le contenu. Quant à la $P(d|c)$, au lieu de traiter item d comme un ensemble et d'estimer la probabilité de le voir dans le catégorie c , on estime souvent la probabilité individuelle de chaque terme de l'item d apparaissant dans le catégorie c du fait qu'il est rare de voir un même item plus d'une fois (Lops, de Gemmis et Semeraro, 2011). Ainsi est introduite la classification naïve bayésienne.

La classification naïve bayésienne est l'un des types de classification probabiliste le plus connus. Reposant sur le théorème de Bayes, elle met en œuvre un classifieur bayésien naïf, appartenant à la famille des classifieurs linéaires (Lops, de Gemmis et Semeraro, 2011). Son hypothèse de base est l'indépendance conditionnelle : elle suppose que l'existence d'une caractéristique pour une classe est indépendante de l'existence d'autres caractéristiques.

Dans le contexte de SR basé sur le contenu, tous les termes dans l'item d sont indépendants d'autres termes. L'item d est représenté comme $d = \{t_1, t_2, t_3, \dots, t_k, \dots, t_n\}$. Ainsi, la $P(d|c)$ peut être représenté par $P(t_1, t_2, t_3, \dots, t_k, \dots, t_n|c)$, où le t_k représente le k -ième terme de l'item d . La $P(t_k|c)$ est la probabilité de voir le terme t_k dans la catégorie c . Étant donné que le t_k est une variable indépendante :

$$P(d|c) = \prod_i^n P(t_i|c)$$

La formule précédente devient :

$$\begin{aligned} P(c|d) &= \frac{P(d|c) \times P(c)}{P(d)} \\ &= \frac{P(c) \times P(c|t_1, t_2, t_3, \dots, t_n)}{P(d)} \\ &= \frac{P(c) \times \prod_{i=1}^n P(t_i|c)}{P(d)} \end{aligned}$$

Dans la pratique, McCallum et Nigam ont distingué deux approches principales existantes qui permettent de mettre en œuvre le classifieur bayésien naïf : le modèle (d'événement) multivarié de Bernoulli et le modèle (d'événement) multinomial (McCallum et Nigam, 1998). Le modèle multivarié de Bernoulli traite les items comme un vecteur binaire qui ne décrit que si un certain terme apparaît ou pas. Dans le modèle multivarié Bernoulli, la $P(c|d)$ est calculée :

$$P(c|d) = \frac{P(c) \times \prod_{i=1}^n P(t_i|c)}{P(d)}$$

où le t_i est le i -ième terme dans l'item d et n est le nombre de termes dans l'item d . On estime la probabilité de l'item d appartenant à la catégorie c en estimant les probabilités individuelles de chaque terme de l'item d appartenant à la catégorie c . Ce modèle montre l'avantage de simplicité. Pourtant, on perd les informations éventuellement importantes sur le nombre d'occurrences et la position d'un terme dans l'item (Jannach *et al.*, 2010). Les évaluations empiriques montrent aussi que le modèle multinomial possède une meilleure performance par rapport le modèle multivarié Bernoulli et surtout quand il s'agit des items avec un grand vocabulaire (McCallum et Nigam, 1998).

Dans le modèle multinomial, le nombre d'occurrences des termes est compté, c'est-à-dire que la probabilité conditionnelle d'un terme t_i apparaissant dans un item de la catégorie c doit aussi estimer par rapport la fréquence d'occurrence du terme t_i dans tous les items de catégorie c :

$$P(t_i|C = c) = \frac{NomTermes(t_i, items(c))}{TousTermes(items(c))}$$

où $NomTermes(t_i, items(c))$ est le nombre d'occurrences du terme t_i dans les items qui appartiennent à la catégorie c et $TousTermes(items(c))$ est le nombre de tous les termes dans ces items. Dans la pratique, lissage Laplace est utilisé pour éviter la valeur de zéro :

$$\hat{P}(t_i|C = c) = \frac{NomTermes(t_i, items(c)) + 1}{TousTermes(items(c)) + |T|}$$

où le $|T|$ est le nombre de tous les termes apparaissant dans tous les items. Enfin, la probabilité de l'item d_i appartenant à la catégorie c_j est calculée :

$$P(c_j|d_i) = P(c_j) \prod_{w \in V_{d_i}} P(t_k|c_j)^{N_{(d_i,t_k)}}$$

où V_{d_i} représente l'ensemble de vocabulaire de l'item d_i , t_k est le k-ième terme de l'item d_i , $N_{(d_i,t_k)}$ est le nombre d'occurrences du terme t_k dans d_i .

On trouve cette approche basée sur le classifieur bayésien naïf dans Syskill &Webert (Pazzani, Muramatsu et Billsus, 1996; Pazzani et Billsus, 1997), NewsDude (Billsus et Pazzani, 1999), Daily Learner (Billsus et Pazzani, 2000), LIBRA (Mooney et Roy, 2000) et ITR (Degemmis, Lops et Semeraro, 2007; Semeraro *et al.*, 2009).

Malgré leur modèle de conception dit « naïf » et ses hypothèses de base extrêmement simplistes, le classifieur bayésien naïf a fait preuve d'une efficacité plus que suffisante dans de nombreuses situations réelles complexes (Lewis et Rieman, 1993; Pazzani et Billsus, 1997). Il a de bons résultats (Jannach *et al.*, 2010) et une facilité de mise en œuvre (Lops, de Gemmis et Semeraro, 2011). En outre, du fait de son hypothèse de « l'indépendance conditionnelle », les composants du classifieur peuvent être facilement mis à jour lorsque de nouvelles données sont disponibles et le temps. Néanmoins, cette approche a besoin de données d'entraînement (des données précédentes) et ainsi il subit le problème du « démarrage à froid ¹³¹» (Pazzani et Billsus, 1997; Jannach *et al.*, 2010).

IV.2.4.3 Autres méthodes

D'autres algorithmes d'apprentissage sont également utilisés dans le SR basé sur le contenu. En fait, le problème d'apprentissage du profil de l'utilisateur est essentiellement un problème de catégorisation des items ou d'étiquetage d'item. En effet, cela peut être considéré comme un cas particulier du problème de la classification de texte dont le but est de classer des items en fonction du profil (ou les préférences) de l'utilisateur dans deux catégories : pertinente et impertinente (Jannach *et al.*, 2010).

Dans le contexte d'un SR basé sur le contenu, les techniques de la classification supervisée

¹³¹ Le problème du démarrage à froid (*cold-start problem*) désigne la difficulté que les SR rencontrent lorsque les recommandations sont nécessaires pour des items et/ou des utilisateurs pour lesquels ils n'ont aucune information explicite ou implicite.

(ou l'apprentissage supervisé) sont centrales. Parmi les techniques de la classification supervisée connues, nous pouvons citer la méthode Rocchio et l'approche probabiliste que nous avons déjà parlé dans les deux sections précédentes. Il existe bien d'autres techniques que nous allons maintenant présenter : c'est le cas de la méthode des k plus proches voisins (*k-Nearest Neighbors – kNN*) et des arbres de décision.

La méthode des k plus proches voisins

Dans le domaine des SR, la méthode des k plus proches voisins (*k-Nearest Neighbors, kNN*) est la plus simple et la plus utilisée pour l'approche de FC. Cette méthode repose sur l'idée intuitive que l'identification des items ou des utilisateurs similaires équivaut à l'identification des voisins d'un item ou d'un utilisateur. C'est un type d'algorithme de classification (d'apprentissage) basée sur l'instance (*Instance-based classifiers, Instance-based learning*) (Amatriain *et al.*, 2011). Ce type d'algorithme repose sur l'ensemble de données d'entraînement (ou les données précédentes stockées dans le système). Il mémorise toutes les données d'entraînement et classifie un nouvel item en le comparant à tous les items stockés. En conséquence, il est aussi appelé algorithme basé sur la mémoire (*memory based classifieur ou memory based learning*) (Daelemans, Van Den Bosch et Ubler, 2005).

Dans le contexte de SR basé sur le contenu, cette méthode compare un nouvel item de catégorie inconnue à tous les items stockés. Il s'agit ensuite de choisir pour ce nouvel item, la catégorie majoritaire parmi ses k plus proches voisins (Negre, 2015). Plus concrètement, dans le but de classifier un nouvel item (en général, les items sont représentés par les vecteurs dans le modèle d'espace vectorielle), l'algorithme kNN détermine les k plus proches items (soit les plus proches voisins, *nearest neighbors*) dans l'ensemble de données stockées en calculant la distance entre eux ou en calculant la similarité entre eux (les mesures de similarité dont nous avons parlé dans la section « IV.2.3 Mesures de Similarité »). Il compte les occurrences de chaque classe pour les k plus proches items et enfin attribue la classe la plus fréquente à ce nouvel item. L'idée de base est que si un item se trouve dans un voisinage (*neighborhood*) où la plupart de voisins (d'items) appartiennent à une classe particulière, c'est parce que ce nouvel item est susceptible d'appartenir à la même classe. (Amatriain *et al.*, 2011; Negre, 2015).

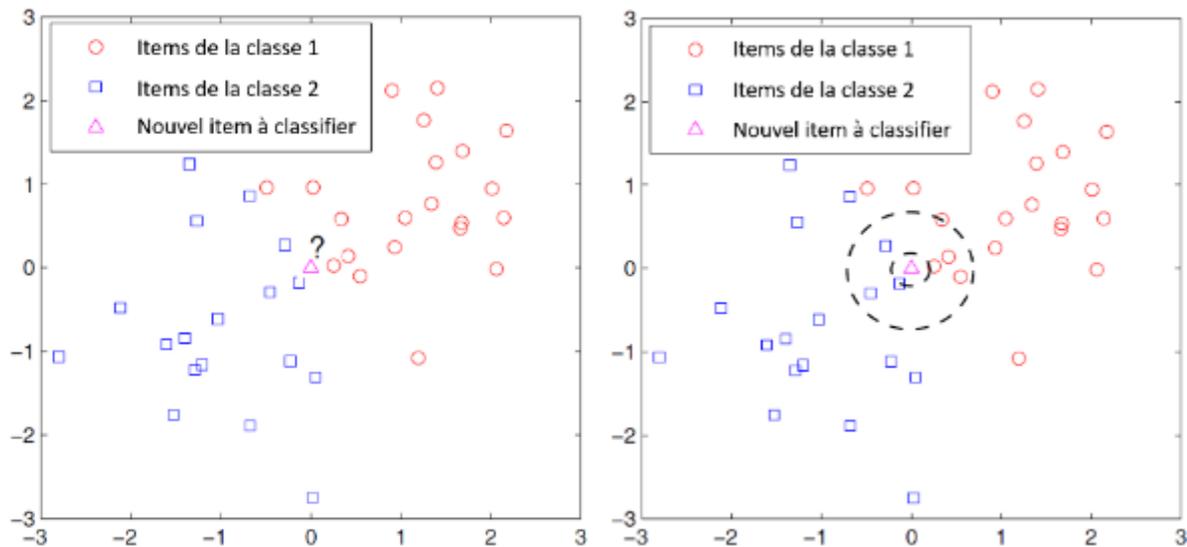


Figure 19 : nouvel item sera classifié à différent class suit à la différente valeur de k

Le choix de k plus proches voisins est le cœur de cette méthode qui « nécessite de connaître k , le nombre de voisins à considérer, qui peut être déterminé par validation croisée » (Negre, 2015). La valeur de k influencerait largement la classification finale : quand la valeur de k est trop petite, le classifieur serait trop sensible aux items impertinents et quand la valeur de k est trop large, le « voisinage » comprendrait trop d'items appartenant aux autres classes (Amatriain *et al.*, 2011). Comme Amatriain ses collègues montrent dans la Figure 19, la valeur de k doit être déterminé dans le but de classer le nouvel item représenté par le triangle. Si $k=1$, ce nouvel item sera attribué à la classe 2 dont les items sont représentés par le carré ; si $k=5$, cet item sera attribué à la classe 1 dont les items sont représentés par le cercle.

Cette méthode est relativement facile à mettre en œuvre. Elle s'adapte rapidement aux changements et exige relativement moins de données par rapport d'autres algorithmes (Zezula *et al.*, 2005; Jannach *et al.*, 2010; Amatriain *et al.*, 2011). Néanmoins, l'un de ses inconvénients est son inefficacité lors de classification car elle ne repose pas sur les donnée « entrainements » et fait des calculs en temps réels (Amatriain *et al.*, 2011). Évidemment, les calculs seront trop lourds au fur et à mesure de la croissance des items et des utilisateurs (Zezula *et al.*, 2005). D'ailleurs, selon Jannach, l'exactitude de prédiction d'une pure méthode des k plus proches voisins n'est pas idéale par rapport aux autres techniques plus sophistiqués (Jannach *et al.*, 2010).

Les arbres de décision

Un arbre de décision (Quinlan, 1986; Rokach et Maimon, 2014) est la représentation graphique d'une procédure de classification, plus particulièrement sous forme d'arbre, où les extrémités de chaque branche (ou feuilles, nœud) représentent les différents résultats possibles en fonction des décisions prises à chaque étape (Negre, 2015).

Les arbres de décision ont été largement appliqués aux problèmes pratiques comme à celui de l'exploration de donnée (*Data Mining*). Dans le cadre de SR, les nœuds internes (*inner nodes*)¹³² de l'arbre représentent les caractéristiques (les termes ou les mots-clés) des items et permettent de partitionner les données en fonction de l'existence ou de la non-existence d'un mot-clé dans un item (Jannach *et al.*, 2010; Negre, 2015). Plus particulièrement, dans le contexte de SR basé sur le contenu, le système préfère les « méta-caractéristiques » (i.e. le genre, l'auteur) aux termes et mots-clés. La raison est qu'il y a trop de caractéristiques (termes et mots-clés) à traiter dans les SR basé sur le contenu alors que l'arbre de décision fonctionne mieux seulement avec des items dont le nombre de caractéristiques est relativement restreint. En fait, la méthode de l'arbre de décision est rarement utilisée comme technique de base dans le SR basé sur le contenu. En revanche, elle est souvent appliquée combinée à d'autres techniques dans le but d'améliorer l'efficacité et la précision des recommandations (Jannach *et al.*, 2010).

IV.2.5 Avantages et limites

Nous constatons plusieurs avantages de l'approche de filtrage basé sur le contenu en comparaison avec d'autres et en particulier en comparaison avec le FI :

¹³² Un nœud interne (*internal node*) aussi appelé *inner node* ou *branch node* est un nœud d'arbre qui possèdent des fils (*sous-branches*) ; au contraire, un nœud externe (*external node*) aussi appelé les feuilles (*leaf nodes*) est celui qui ne possède pas de fils dans l'arbre.

- **L'indépendance de l'utilisateur** : cette approche n'utilise que les évaluations effectuées par l'utilisateur. Cela signifie qu'il ne requiert pas une grande communauté d'utilisateurs, car il ne nécessite pas de données provenant d'autres utilisateurs et qu'il possède la capacité de recommander des items similaires à ceux que l'utilisateur a aimés dans le passé ;
- **Le nouvel item** : cette approche montre aussi son avantage dans la découverte des « nouveaux items » : en fonction des mots-clés de description des items, le SR est capable d'en recommander de nouveaux qui n'ont pas été évalués au préalable par un utilisateur ou qui ne sont pas très populaires ;
- **La transparence** : il est facile d'expliquer le fonctionnement du système en listant explicitement des caractéristiques d'item. Cela donne à l'utilisateur une plus grande confiance dans le système. Malgré tous ses avantages, ses limites sont aussi évidentes.

Malgré tous ses avantages, ses limites sont aussi évidentes :

- **La limitation d'analyse du contenu** : cette approche s'appuie notamment sur les mots-clés et traite principalement les informations sémantiques textuelles. Par conséquent, le système n'est pas capable de traiter les items qui ne peuvent pas être représentés par les mots-clés et il est difficile de distinguer des items qui sont représentés par le même ensemble de mots-clés ;
- **Le problème de « sur-spécialisation »** : ce problème aussi appelé le problème de « sérendipité ». Cette approche est incapable de traiter d'autres critères de pertinence que les critères strictement thématiques. Les recommandations se limitent aux items similaires (réponses trop homogènes). Le système ne recommande que les items trop similaires à ceux que l'utilisateur a déjà aimés dans le passé ou même des items identiques (les nouvelles d'un même événement, ou les différentes versions d'un film) ;
- **Le problème du « nouvel utilisateur »** : quand un nouvel utilisateur commence à utiliser le système ou quand l'utilisateur n'a pas évalué suffisamment d'items pour que le système puisse comprendre ses préférences, il risque que le système

ne propose aucune recommandation pertinente. Selon Pazzani et Billsus, lorsque le nombre d'évaluations est entre vingt et cinquante, le système commence à proposer les bonnes recommandations (Pazzani et Billsus, 1997). Et pourtant, les utilisateurs n'ont souvent pas l'envie d'effectuer tellement évaluations (Jannach, 2011).

IV.3 Approche de filtrage collaboratif

L'approche du FC cherche à prédire l'utilité de l'item pour l'utilisateur en se basant sur les évaluations des items faites par d'autres utilisateurs précédemment (Adomavicius et Tuzhilin, 2005). Le système propose des recommandations en calculant la similarité entre les préférences d'un utilisateur et celles d'autres utilisateurs (Negre, 2015). Ce type de filtrage se base sur l'hypothèse qu'il existe des groupes d'utilisateurs qui se comportent de la même façon et qui ont des besoins et des préférences comparables (Jannach *et al.*, 2010) (Figure 20).

Dans ce cas, les algorithmes de FC cherchent à trouver les utilisateurs (les voisins proches) en fonction de leurs préférences et à les faire correspondre afin de produire des recommandations pour un utilisateur (Terveen et Hill, 2001). Cela veut dire que pour chaque utilisateur un ensemble de proches voisins est identifié, et que la décision de proposer ou non un item à l'utilisateur dépendra des appréciations des membres de son voisinage (Berrut et Denos, 2003).

Cette approche est largement investie et adoptée dans différentes applications (Oufaida et Nouali, 2009) telles que Tapestry (Goldberg *et al.*, 1992), GroupLens (Resnick *et al.*, 1994), Ringo (Shardanand et Maes, 1995), Amazon (Linden *et al.*, 2003), Netflix, et Barnes & Noble *etc.*

Les techniques de FC se répartissent généralement en deux approches (voir Figure 14): l'approche basée sur la mémoire (*memory based*) et l'approche basée sur le modèle (*model based*) (Breese, Heckerman et Kadie, 1998; Deshpande et Karypis, 2004; Adomavicius et Tuzhilin, 2005; Bell, Koren et Volinsky, 2007; Jannach *et al.*, 2010; Desrosiers et Karypis, 2011). L'approche basée sur la « mémoire » est aussi appelée l'approche basée sur le voisinage (*neighborhood based*) (Breese, Heckerman et Kadie, 1998) ou l'approche basée sur

l'heuristique (*heuristic based*) (Adomavicius et Tuzhilin, 2005). Elle est essentiellement heuristique et utilise directement toute la base de données d'évaluations des utilisateurs pour faire les prédictions. Deux sous approches principales existent : l'approche basée sur l'utilisateur (*user-based* ou *user-user*) et l'approche basée sur l'item (*item-based* ou *item-item*). Les systèmes basés sur l'utilisateur tels que GroupLens (Konstan *et al.*, 1997), Bellcore Video (Hill *et al.*, 1995) et Ringo (Shardanand et Maes, 1995), cherchent à prédire l'intérêt de l'utilisateur_u sur l'item_i en utilisant les notes faites par d'autres utilisateurs similaires (ses voisins). L'approche basée sur l'item (Linden, Smith et York, 2003; Deshpande et Karypis, 2004) consiste à prédire la note de l'item_i pour l'utilisateur_u en utilisant les notes faites par les autres utilisateurs sur les items qui sont similaires à l'item_i.

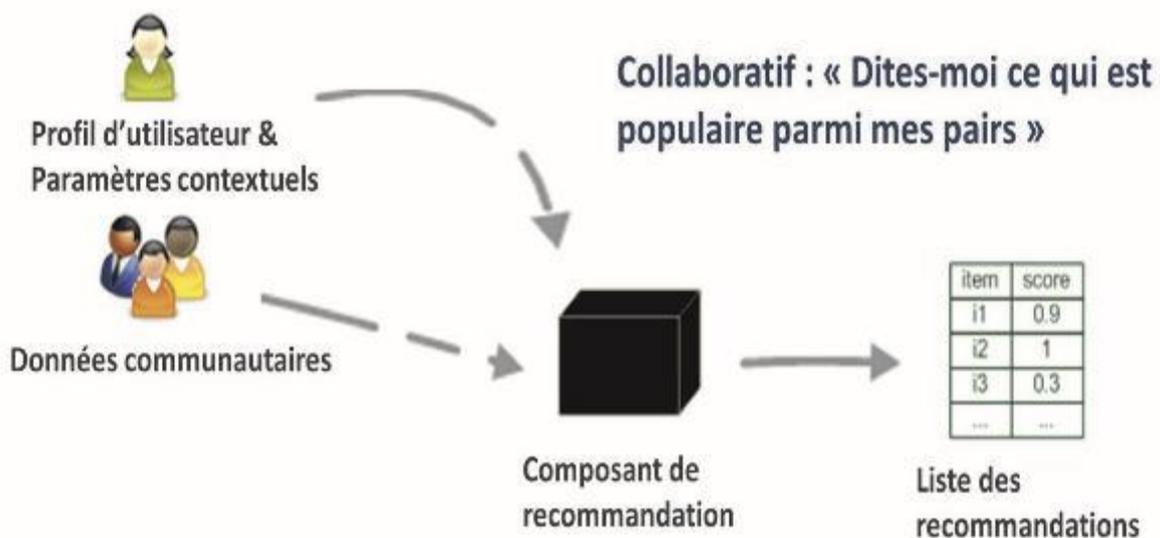


Figure 20 Le SR collaboratif vu comme une boîte noire (Negre, 2015) traduction de (Jannach, 2010)

L'approche basée sur modèle utilise les évaluations des items faites par les utilisateurs pour apprendre un modèle prédictif. Plusieurs techniques ont été proposées : le partitionnement (*Clustering*) (Breese, Heckerman et Kadie, 1998; George et Merugu, 2005), l'analyse sémantique latente (*Latent Semantic Analysis*) (Hofmann, 2004), l'allocation de Dirichlet latente (*Latent Dirichlet Allocation*) (Blei, Ng et Jordan, 2003), l'entropie maximale (*Maximum Entropy*) (Zitnick et Kanade, 2004), *Boltzmann Machines* (Salakhutdinov, Mnih et Hinton, 2007), Machine à vecteurs de support (Grčar *et al.*, 2006), and Singular Value Decomposition (Bell, Koren et Volinsky, 2007; Paterek, 2007; Koren, 2008; Takács *et al.*, 2008).

Dans les sections suivantes, nous allons présenter les approches principales et les techniques les plus utilisées dans le FC.

V.3.1 Définition du problème des SR

Le problème des SR est d'estimer la pertinence d'un nouvel item en fonction des informations historiques stockées dans le système et de proposer à l'utilisateur les items les plus pertinents (Desrosiers et Karypis, 2011; Negre, 2015). Dans le but de réaliser des recommandations, l'approche basée sur le contenu se concentre sur les similarités entre les caractéristiques du contenu des items différents tandis que l'approche de FC repose sur les similarités entre les préférences et les opinions de l'utilisateur_{*u*} et ceux des autres utilisateurs.

L'approche de FC se base notamment sur la matrice $m \times n$ de notes utilisateurs-items et $R_{u,i}$ est la note (*rating*) de *i*-ième item_{*i*} faite par *u*-ième utilisateur_{*u*} (Tableau 2 La matrice $m \times n$ de notes utilisateurs-items pour). Pour les approches purement collaboratives, cette matrice est la seule qui produise deux types de résultat : une prédiction numérique indiquant à quel niveau l'utilisateur_{*u*} aimera un item_{*i*} et une liste de *n* items recommandés (Jannach *et al.*, 2010). Ces deux types de sorties concernent en effet des problèmes les plus importants dans les SR : le problème de meilleur item (*best item*, appelé aussi *prediction problem*) et le problème des *N* meilleures recommandations (*top-N recommendation*) (Deshpande et Karypis, 2004). Quand des notes sont disponibles, c'est-à-dire, des évaluations numériques, l'objectif est de trouver l'item_{*i*} que l'utilisateur_{*u*} aimera le plus parmi les nouveaux items que l'utilisateur_{*u*} n'a pas vus. Dans ce cas, la note prédictive est calculée par une fonction $f(u,i)$ qui évalue les notes que l'utilisateur_{*u*} donnera à chaque item_{*i*}. L'item ayant la meilleure note sera recommandé à l'utilisateur_{*u*}. Quand les notes ne sont pas disponibles, il concerne le problème de *N* meilleures recommandation. Dans ce cas, il est impossible de calculer des notes prédictives faute de données calculables. On ne peut que proposer à l'utilisateur_{*u*} une liste d'items qui sont susceptibles de l'intéresser (Desrosiers et Karypis, 2011).

	Item A	Item B	Item C	...	Item m
Utilisateur 1	$R_{u1, A}$	$R_{u1, B}$	$R_{u1, C}$...	$R_{u1, m}$
Utilisateur 2	$R_{u2, A}$	$R_{u2, B}$	$R_{u1, C}$...	$R_{u1, m}$
Utilisateur 3	$R_{u3, A}$	$R_{u3, B}$	$R_{u1, C}$...	$R_{u1, m}$
...
Utilisateur n	$R_{un, A}$	$R_{un, A}$	$R_{un, C}$...	$R_{un, m}$

Tableau 2 La matrice $m \times n$ de notes utilisateurs-items pour FC

IV.3.2 Filtrage collaboratif basé sur le mémoire

Le FC basé sur le mémoire est aussi appelé le FC basé sur le voisinage (Jannach *et al.*, 2010) ou le FC basé sur l'heuristique (Adomavicius et Tuzhilin, 2005). Dans cette approche, l'ensemble des évaluations effectuées par l'utilisateur sont utilisées directement pour produire des recommandations. En général, deux sous-approches peuvent être distinguées : l'approche basée sur l'utilisateur (*user-user*) et l'approche basée sur l'item (*item-item*).

IV.3.2.1 Approche basée sur l'utilisateur

L'approche basée sur l'utilisateur (Resnick *et al.*, 1994; Shardanand et Maes, 1995; Konstan *et al.*, 1997; Breese, Heckerman et Kadie, 1998; Herlocker et Konstan, 1999) est considérée comme l'une des premières approches dans les SR. Elle concentre sur la similarité entre les préférences des utilisateurs différents. L'idée de base est d'identifier un ensemble d'utilisateurs (voisins) qui partagent les préférences similaires et de recommander à l'utilisateur des items aimés par ses voisins. Cette approche est fondée sur l'hypothèse que les préférences de l'utilisateur sont relativement stables et elles se synchronisent avec ses voisins (Jannach *et al.*, 2010).

Dans le but de réaliser la prédiction de l'item_i pour l'utilisateur_u, deux étapes principales peuvent être distinguées : l'identification des voisins de l'utilisateur_u et le calcul des prédictions. La première étape consiste à identifier les voisins de l'utilisateur_u. Les voisins sont des

utilisateurs qui ont fait des évaluations similaires à celles de l'utilisateur $_u$ sur certains items. Cela implique que l'utilisateur $_u$ et ses voisins partagent des préférences. Cette étape utilise les techniques statistiques pour mesurer la similarité entre les utilisateurs afin de sélectionner les k plus proches voisins. Une fois les k plus proches voisins identifiés, plusieurs algorithmes sont employés afin de calculer la prédiction pour l'utilisateur $_u$ (Sarwar *et al.*, 2001).

IV.3.2.1.1 Calcul des similarités et identification des voisins

Dans le but d'identifier les voisins de l'utilisateur $_u$, il est nécessaire de choisir une propre mesure de similarité qui calcule la similarité de préférences entre les utilisateurs. En effet, les mesures de similarité jouent un rôle important dans cette approche car elles permettent non seulement d'identifier les voisins avec qui l'utilisateur $_u$ partage des préférences, mais aussi d'attribuer les poids à l'importance des voisins. Le calcul de similarité est l'une des techniques les plus critiques de la construction d'un SR basé sur le voisinage car il influence à la fois la précision et la performance d'un SR de manière significative (Desrosiers et Karypis, 2011).

Dans le but de déterminer la proximité des utilisateurs (Jannach *et al.*, 2010), plusieurs mesures ont été exploitées parmi lesquelles nous pouvons citer : la similarité cosinus (Lang, 1995; Billsus et Pazzani, 1998, 2000; Sarwar *et al.*, 2000b), le coefficient de corrélation de Pearson (Herlocker et Konstan, 1999), le coefficient de corrélation de rang de Spearman (Gibbons et Kendall, 1990; Resnick *et al.*, 1994) et MSD (*Mean Square Difference*) (Shardanand et Maes, 1995).

La mesure de similarité cosinus

Comme nous avons déjà décrit dans la section « IV.2.3 Mesures de Similarité », la mesure de similarité cosinus permet de calculer le cosinus de l'angle entre les deux vecteurs à n dimensions afin de mesurer la similarité entre eux. Dans l'approche basée sur l'utilisateur, la similarité entre l'utilisateur $_u$ et l'utilisateur $_v$ peut être mesurée en calculant le cosinus de l'angle entre les deux vecteurs $\vec{r}_{u,i}$ et $\vec{r}_{v,i}$ qui représentent les notes de l'item $_i$ faites respectivement par

l'utilisateur_u et l'utilisateur_v. La similarité entre eux est ainsi calculée :

$$Sim_{cosinus}(\vec{r}_{u,i}, \vec{r}_{v,i}) = \cos\theta = \frac{\sum_{i \in I_{u,v}} \vec{r}_{u,i} \cdot \vec{r}_{v,i}}{\sqrt{\sum_{i \in I_u} (\vec{r}_{u,i})^2} \times \sqrt{\sum_{i \in I_v} (\vec{r}_{v,i})^2}}$$

où θ désigne l'angle entre les deux vecteurs ; $I_{u,v}$ est l'ensemble d'items que l'utilisateur_u et l'utilisateur_v ont tous notés. La valeur $\cos\theta$, à savoir, la valeur de similarité est comprise dans l'intervalle $[0, 1]$. La valeur 0 indique des vecteurs indépendants (orthogonaux) tandis que la valeur de 1 indique des vecteurs similaires (colinéaires de coefficient positif). Un problème de cette mesure est qu'elle ne tient pas compte des différences dans la moyenne et la variance des notes faites par les deux utilisateurs.

Le coefficient de corrélation de Pearson

Dans le cadre du FC basé sur l'utilisateur, l'utilisateur est représenté par les évaluations (les notes) qu'il a données sous la forme vectorielle. Alors, le coefficient de corrélation de Pearson mesure la similarité entre deux utilisateurs comme le cosinus de l'angle entre leurs représentations vectorielles centrées-réduites (Negre, 2015). La formule est :

$$Sim_{pearson} = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}}$$

où $r_{u,i}$ et $r_{v,i}$ sont la note d'item_i donnée respectivement par l'utilisateur_u et l'utilisateur_v ; \bar{r}_u (resp. \bar{r}_v) est la note moyenne de l'utilisateur_u (resp. l'utilisateur_v). La valeur de la corrélation de Pearson est entre +1 (corrélacion fortement positive) et -1 (corrélacion fortement négative). La similarité de cosinus ajustée est une modification de ce coefficient, qui est largement utilisée dans l'approche basée sur l'item. En tenant compte de la note moyenne de l'utilisateur, cette méthode évite le problème de l'interprétation différente de l'échelle des notes par les utilisateurs (certains utilisateurs ont tendance à donner de bonnes notes à la plupart des items alors que d'autres sont plus stricts et donnent rarement de bonnes notes).

Le coefficient de corrélation sur les rangs

Le coefficient de corrélation sur les rangs (Spearman rank correlation coefficient) est le premier test statistique réalisé à partir des rangs. Il mesure l'association entre deux variables mesurées au moins dans une échelle ordinale. Dans le cadre du FC, le coefficient de corrélation de rang de Spearman prend en compte les rangs des notes au lieu de la valeur des notes. Selon le coefficient de corrélation de Spearman, la similarité entre l'utilisateur $_u$ et l'utilisateur $_v$ est calculée ainsi :

$$Sim_{spearman} = \frac{\sum_{i \in I_{u,v}} (rg_{u,i} - \overline{rg}_u)(rg_{v,i} - \overline{rg}_v)}{\sqrt{\sum_{i \in I_{u,v}} (rg_{u,i} - \overline{rg}_u)^2 \sum_{i \in I_{u,v}} (rg_{v,i} - \overline{rg}_v)^2}}$$

où le $rg_{u,i}$ (resp. $rg_{v,i}$) est le rang de note de l'item $_i$ dans la liste des items que l'utilisateur $_u$ a notés (resp. $rg_{v,i}$) ; le \overline{rg}_u (resp. \overline{rg}_v) est le rang moyen des items notés par l'utilisateur $_u$ (resp. l'utilisateur $_v$) ; $I_{u,v}$ est l'ensemble d'items notés par l'utilisateur $_u$ et l'utilisateur $_v$ en commun. Il s'interprète de la même manière que le coefficient de corrélation de Pearson : une valeur positive (maximum = +1) indique une variation simultanée dans le même sens, une valeur négative (minimum = -1) une variation simultanée en sens inverse. En introduisant le rang, il évite le problème de la normalisation des notes. Cependant, il ne fonctionne pas bien lorsqu'il n'y a que très peu de valeurs. Par ailleurs, il est apparemment plus « cher » que le coefficient de corrélation de Pearson dans la réalité car il a besoin de davantage d'informations sur le rang de note d'items (Desrosiers et Karypis, 2011).

MSD (Mean Square Difference)

Le MSD est introduit pour la première fois dans le système de Ringo (Shardanand et Maes, 1995). Il évalue la similarité entre deux utilisateurs (l'utilisateur $_u$ et l'utilisateur $_v$) comme l'inverse de la différence quadratique moyenne entre les notes sur les mêmes items faites par l'utilisateur $_u$ et l'utilisateur $_v$. La similarité est calculée :

$$MSD(u, v) = \frac{|I_{u,v}|}{\sum_{i \in I_{u,v}} (r_{u,i} - r_{v,i})^2}$$

où $I_{u,v}$ est l'ensemble d'items que l'utilisateur $_u$ et l'utilisateur $_v$ ont tous notés ; $r_{u,i}$ (resp. $r_{v,i}$) est la note d'item $_i$ faite par l'utilisateur $_u$ (resp. l'utilisateur $_v$). Cette technique ne tient pas compte des corrélations négatives entre les préférences des utilisateurs ou l'appréciation de différents items. L'existence de telles corrélations négatives peut améliorer la précision de la prédiction de la note (Herlocker, Konstan et Riedl, 2002; Desrosiers et Karypis, 2011).

Les études montrent que le MSD et la mesure de similarité cosinus ne conduisent pas à des prédictions idéales par rapport au coefficient de corrélation de Pearson et au coefficient de corrélation de rang de Spearman. La corrélation de Pearson est un peu plus précise que la corrélation de rang de Spearman. Dans l'approche de FC basée sur l'utilisateur, le coefficient de corrélation de Pearson est considéré comme la meilleure mesure de similarité (Herlocker, Konstan et Riedl, 2002; Jannach *et al.*, 2010; Herlocker *et al.*, 2017).

Pourtant, les inconvénients existent dans la pratique. L'un des problèmes est que, dans la réalité, il y a des items aimés par tout le monde. Une mesure de similarité comme la corrélation de Pearson ne prend pas en compte le fait qu'un accord par deux utilisateurs sur un item moins apprécié par la plupart des gens (un item controversé) a plus de « valeur » qu'un accord sur un item aimé par tout le monde (Jannach *et al.*, 2010). La solution consiste soit à réduire l'importance relative des items aimés par tout le monde (la méthode « fréquence inverse utilisateurs (*inverse user frequency*)») (Breese, Heckerman et Kadie, 1998), soit à augmenter l'influence des items controversés (la méthode « facteur de variance pondérée (*variance weighting factor*)») (Herlocker *et al.*, 2017).

Un autre problème concernant la rareté de données (*The Sparse Data Problem*) est que dans la réalité, les données de notes sont souvent rares en comparaison avec le nombre d'utilisateurs et d'items dans un système. Il arrive souvent que l'utilisateur note très peu d'items en commun avec les autres ou que les items similaires soient notés par très peu des mêmes utilisateurs. Dans ce cas, le calcul des similarités risque d'être non fiable voire impossible. Par conséquent, les prédictions seront mauvaises (Herlocker *et al.*, 2017). Dans le but de résoudre ce problème, certaines techniques ont été proposées telles que la « pondération de signifiante (*significance weighting*) » (Herlocker, Konstan et Riedl, 2002; Herlocker *et al.*, 2017) et « la note par défaut (*default voting*)» (Breese, Heckerman et Kadie, 1998).

V.3.2.1.2 Sélection les k plus proches voisins

Dans la pratique, les SR possèdent souvent millions d'utilisateurs et d'items, il n'est pas possible de stocker tous les similarités entre chaque pair d'utilisateurs (ou d'items) à cause de la limitation de la mémoire de l'ordinateur (Desrosiers et Karypis, 2011). D'ailleurs, cette approche repose sur les données qui doivent être mises à jour assez fréquemment, et un tel calcul incluant tous les voisins prendra trop de temps (Jannach *et al.*, 2010). C'est pourquoi la sélection d'un groupe d'utilisateurs (un voisinage) au lieu de l'ensemble d'utilisateurs dans le système pour calculer les prédictions peut contribuer à la fois à la performance et à la précision.(Shardanand et Maes, 1995). Dans la littérature, deux stratégies principales sont proposées pour réduire la taille de voisinage lors de calculer la prédiction :

1. le filtrage par seuil de similarité : établit un seuil de similarité S_{min} à partir duquel les voisins qui sont plus similaires à l'utilisateur $_u$ pourront être sélectionnés (Shardanand et Maes, 1995; Breese, Heckerman et Kadie, 1998) ;
2. le filtrage de N meilleurs voisins (*Top-N filtering* ou *Best-n neighbors*) détermine une taille du meilleur voisinage qui ne laisse entrer qu'un groupe de voisins plus similaire à l'utilisateur $_u$. Le nombre de voisins est fixé à N (Herlocker *et al.*, 2017).

D'autres stratégies existent telles que le filtrage négatif (Desrosiers et Karypis, 2011) et le filtrage par seuil des items co-notés (Viappiani, Faltings et Pu, 2006) *etc.* Les voisins choisis sont les k plus proches voisins (k NN, l'abréviation de k *Nearest Neighbors*). Dans la perspective de sélectionner les k plus proches voisins parmi l'ensemble de voisins identifiés, la valeur de k doit être définie. La détermination de la valeur de k doit éviter les valeurs extrêmes. Si la valeur du seuil de similarité S_{min} (ou la valeur de N) est trop élevée, la taille de voisinage et la couverture seront trop petites. Cela implique qu'il y aura très peu d'items à recommander ; au contraire, si le seuil S_{min} (ou la valeur de N) est trop faible (petite), la taille de voisinage ne sera guère réduite (Anand et Mobasher, 2003; Herlocker *et al.*, 2017). Selon les analyses de MovieLens, la valeur raisonnable de k est entre 20 à 50 (Good *et al.*, 1999; Herlocker, Konstan et Riedl, 2002). Si le k est trop grand ($k > 50$), les voisins qui sont moins similaires à l'utilisateur $_u$ seront compris dans le calcul et ainsi les « bruits » influenceront la prédiction ; si le k est trop petit ($k < 20$), il risque que beaucoup d'items n'aient pas de chance d'être recommandés (Anand et Mobasher, 2003; Herlocker *et al.*, 2017). Selon Desrosiers et Karypis,

la valeur optimale de k doit être finalement déterminée par validation croisée (Desrosiers et Karypis, 2011).

IV.3.2.1.3 Prédiction de l'item pour l'utilisateur

Ayant identifié les k plus proches voisins de l'utilisateur u , l'étape suivante consiste à prédire l'opinion qu'aura l'utilisateur u sur un nouvel item i . La méthode la plus simple est de calculer la moyenne des notes données par ses k plus proches voisins pour l'item i :

$$P_{u,i} = \frac{\sum_{v \in V} r_{v,i}}{|V|}$$

où $P_{u,i}$ est la note de prédiction de l'item i pour l'utilisateur u , V est le nombre de k plus proches voisins de l'utilisateur u , $r_{v,i}$ est la note de l'item i faite par le voisin v .

Cette formule néglige la différence de la similarité entre l'utilisateur u et chacun de ses voisins v . En fait, les similarités entre l'utilisateur u et chacun de ses voisins ne sont pas pareilles, c'est-à-dire, certains voisins sont plus similaires (plus proches) à l'utilisateur u tandis que d'autres sont moins similaires (plus loin de) à lui. Naturellement, l'influence de chaque voisin ne doit pas être égale lors du calcul de prédiction. Une solution largement employée est de pondérer la contribution de chaque voisin en fonction de leur similarité à l'utilisateur. Les poids de chaque voisin peuvent être calculés à l'aide des mesures de similarité. Ainsi la formule devient :

$$P_{u,i} = \frac{\sum_{v \in V} sim_{u,v} \cdot r_{v,i}}{\sum_{v \in V} sim_{u,v}}$$

où le poids de voisin v est représenté par la $sim_{u,v}$ qui est la similarité entre l'utilisateur u et son voisin v .

Compte tenu de la différence dans l'interprétation de l'échelle de note par différents utilisateurs, les notes doivent être normalisées. Deux méthodes pour y parvenir sont le centrage sur la moyenne (*Mean centering*) et le Z-score. La méthode du centrage sur la moyenne (*Mean centering*) (Resnick *et al.*, 1994; Breese, Heckerman et Kadie, 1998) cherche à savoir si une note est positive ou négative en la comparant avec la moyenne des notes que l'utilisateur u a

faites. Dans ce cas, la formule est modifiée comme :

$$P_{u,i} = \bar{r}_u + \frac{\sum_{v \in V} sim_{u,v} \cdot (r_{v,i} - \bar{r}_v)}{\sum_{v \in V} |sim_{u,v}|}$$

où \bar{r}_v (resp. \bar{r}_u) est la moyenne des notes de l'utilisateur_v (resp. l'utilisateur_u). L'avantage de cette méthode est que l'on peut savoir immédiatement si un utilisateur aime un item et à quel niveau il l'apprécie. Cette formule de prédiction est relativement simple et largement utilisée. Une étude rapporte que la méthode du Z-score (Herlocker *et al.*, 2017) présente une meilleure performance en montrant la dispersion des notes d'un utilisateur et en intégrant l'écart type (Howe et Forbes, 2008) dont la formule de prédiction est :

$$P_{u,i} = \bar{r}_u + \sigma_u \frac{\sum_{v \in V} sim_{u,v} \cdot \frac{(r_{v,i} - \bar{r}_v)}{\sigma_v}}{\sum_{v \in V} |sim_{u,v}|}$$

où σ_v (resp. σ_u) est l'écart type des notes de l'utilisateur_v (resp. l'utilisateur_u). Néanmoins, cette méthode est si sensible que les notes qu'elle prédit ont souvent tendance à s'écarter de l'échelle. Une fois que les prédictions sont calculées, le système recommande à l'utilisateur u les items ayant les valeurs de prédiction les plus élevées.

IV.3.2.1.4 N meilleures recommandation (*Top-N recommendation*)

Les sections précédentes présentent l'approche FC basée sur l'utilisateur lorsque les notes sont disponibles, (*best item problem*). Dans cette approche, il reste le problème des N meilleures recommandations (*Top-N recommendation*), soit le problème des recommandations quand le système ne dispose que des votes binaires (aime, n'aime pas) ou unitaires (a acheté, a consulté) au lieu des notes numériques définies sur une échelle de valeurs.

La première étape consiste toujours à mesurer la similarité entre les utilisateurs et d'identifier les k plus proches voisins de l'utilisateur_u en recourant aux mesures de similarité. Ensuite, il faut déterminer pour chaque voisin_v, une liste L_v des items pertinents, soit une liste des items obtenus à partir des rétroactions positives de l'utilisateur_v (items aimés, achetés etc.). Les items sont par la suite triés selon la fréquence de leur présence dans les listes de tous les voisins. Les N items les plus fréquents (*Top-N item*) seront alors recommandés à l'utilisateur_u.

(Sarwar *et al.*, 2000a; Su et Khoshgoftaar, 2009).

Les algorithmes de recommandation Top- N basés sur les utilisateurs souffrent du problème du passage à l'échelle et du problème de performance pour les applications en temps réel (Karypis, 2001; Deshpande et Karypis, 2004). Car la complexité de calcul de ces méthodes croît linéairement avec le nombre d'utilisateurs, qui peut atteindre plusieurs millions dans les applications commerciales typiques. En outre, la matrice des similitudes d'utilisateur à utilisateur est assez dense. Les recommandations Top- N ne peuvent pas tirer parti des similitudes d'utilisateurs à utilisateurs précalculées.

IV.3.2.2 Approche Item-Item

L'approche basée sur l'utilisateur souffre de limitations lorsqu'on cherche à l'intégrer dans l'e-commerce. D'abord, elle ne fonctionne pas bien quand il s'agit de traiter un grand nombre d'utilisateurs et d'items. Dans le cadre de sites de vente, il y a souvent des millions d'items et d'utilisateurs à traiter. La complexité du calcul de chaque recommandation augmente de manière linéaire avec la croissance du nombre d'utilisateurs et d'items (Deshpande et Karypis, 2004). Ensuite, pour l'e-commerce, les calculs rapides en temps réel sont nécessaires car le profil de l'utilisateur change assez souvent. D'ailleurs, comme un utilisateur ne note qu'un très petit nombre d'items sur des millions, le problème du manque de données d'utilisateurs (*sparsity problem*) s'accroît dans les sites de vente. La similarité entre lui et ses voisins ne serait pas stable et fiable (Jannach *et al.*, 2010).

Dans ce cas, l'approche basée sur l'item est développée qui permet de calculer les recommandations pour une très grande matrice de notation en temps réel (Sarwar *et al.*, 2001). Elle concentre la similarité entre les notes obtenues par des items différents qui est plus stable que la similarité entre les préférences des utilisateurs. Cette similarité peut être calculée hors ligne. En comparant avec l'approche basée sur le contenu, cette approche s'occupe de la similarité entre les notes d'items différents faites par les utilisateurs au lieu de la similarité de caractéristiques et de contenu entre des items (Deshpande et Karypis, 2004). La prédiction peut être calculée en deux étapes : un précalcul (precomputation) de similarité entre les notes obtenues par les items hors ligne et un calcul de prédiction pour l'utilisateur.

Mesure de similarité

Dans l'approche du FC basée sur l'item, la mesure similarité cosinus est utilisée comme la méthode standard pour calculer la similarité entre les items. Tandis que le coefficient de corrélation de Pearson est considéré comme la meilleure mesure dans l'approche FC basée sur l'utilisateur, la mesure de similarité cosinus produit toujours les meilleurs résultats dans l'approche FC basée sur l'item (Sarwar *et al.*, 2001; Jannach *et al.*, 2010).

Dans l'approche FC basée sur l'item, les notes doivent être normalisées au début. Parmi les méthodes de normaliser des notes, nous pouvons citer le Z-score et le centrage sur la moyenne qui sont déjà présentées dans la section précédente. La normalisation du Z-score dans cette approche est :

$$\hat{r}_{u,i} = \frac{(r_{u,i} - \bar{r}_i)}{\sigma_i}$$

Et la normalisation du centrage sur la moyenne (*Mean centering*) est :

$$\hat{r}_{u,i} = r_{u,i} - \bar{r}_i$$

où le \bar{r}_i est la moyenne des notes d'item_i faites par tous les utilisateurs qui ont noté l'item_i. Alors, la similarité entre l'item_i et l'item_j est calculée :

$$Sim_{cosinus}(i, j) = \cos\theta = \frac{\sum_{u \in U_{i,j}} \hat{r}_{u,i} \cdot \hat{r}_{u,j}}{\sqrt{\sum_{u \in U_i} (\hat{r}_{u,i})^2} \times \sqrt{\sum_{u \in U_j} (\hat{r}_{u,j})^2}}$$

où θ désigne l'angle entre les deux vecteurs ; $\hat{r}_{u,i}$ (resp. $\hat{r}_{u,j}$) est la note normalisée d'item_i (d'item_j) faite par l'utilisateur_u, $U(i,j)$ désigne l'ensemble d'utilisateurs qui ont tous noté l'item_i et l'item_j. Cette mesure de similarité cosinus avec les notes normalisées fonctionne pareillement par rapport à la mesure de la similarité cosinus ajusté, elle soustrait \bar{r}_u , la moyenne des notes de l'utilisateur_u, et la formule est devenue :

$$Sim_{cosinus}(\vec{r}_{i,u}, \vec{r}_{j,u}) = \cos\theta = \frac{\sum_{u \in U_{i,j}} (\vec{r}_{i,u} - \bar{r}_u) \cdot (\vec{r}_{j,u} - \bar{r}_u)}{\sqrt{\sum_{u \in U_i} (\vec{r}_{i,u} - \bar{r}_u)^2} \times \sqrt{\sum_{u \in U_j} (\vec{r}_{j,u} - \bar{r}_u)^2}}$$

Ainsi, la valeur est dans l'intervalle (-1,1) où la valeur -1 indique les deux items sont fortement dissimilaires.

Calcul de Prédiction

La prédiction de l'item_i pour l'utilisateur_u est calculée :

$$P_{u,i} = \frac{\sum_{j \in I_u} sim_{i,j} \cdot \hat{r}_{u,i}}{\sum_{j \in I_j} |sim_{i,j}|}$$

IV.3.3 Le filtrage collaboratif basé sur le modèle

Tandis que l'approche FC basée sur la mémoire utilise l'ensemble des évaluations des items faites par les utilisateurs dans le passé afin de calculer directement les prédictions, l'approche FC basée sur le modèle (Shardanand et Maes, 1995; Billsus et Pazzani, 1998; Breese, Heckerman et Kadie, 1998; Ungar et Foster, 1998; Aggarwal *et al.*, 1999; Getoor et Sahami, 1999; Goldberg *et al.*, 2001; Hofmann, 2003; Pavlov et Pennock, 2003; Marlin, 2004) cherche à construire un modèle basé sur les évaluations hors ligne afin de réaliser les prédictions (Adomavicius et Tuzhilin, 2005; Jannach *et al.*, 2010). L'idée de base est de modéliser les interactions entre les utilisateurs et les items dans un système.

Bien que l'approche FC basée sur la mémoire soit théoriquement plus précise en utilisant l'ensemble de données disponibles pour générer les recommandations (Jannach *et al.*, 2010), ce type de système souffre de problèmes tels que la sensibilité aux données manquantes (Ekstrand, Riedl et Konstan, 2011) et le passage à l'échelle (Sarwar *et al.*, 2000a; Su et Khoshgoftaar, 2009; Desrosiers et Karypis, 2011; Ekstrand, Riedl et Konstan, 2011). La solution est simplement de compléter les notes manquées (Desrosiers et Karypis, 2011). Plusieurs méthodes ont été proposées afin d'aborder ces problèmes parmi lesquelles la réduction de la dimension est la plus populaire et la plus utilisée. Ayant trouvé des méthodes appropriées pour traiter les problèmes des données, plusieurs approches existent pour construire des modèles d'apprentissage afin de réaliser les recommandations. Les approches couramment utilisées sont le clustering Bayésien (*Bayesian Clustering*) (Breese, Heckerman et Kadie, 1998),

l'analyse sémantique latente (*Latent Semantic Analysis*) (Hofmann, 2004), l'allocation de Dirichlet latente (*Latent Dirichlet Allocation*) (Blei, Ng et Jordan, 2003), l'entropie maximale (*Maximum Entropy*) (Zitnick et Kanade, 2004), la machine de Boltzmann (*Boltzmann Machines*) (Salakhutdinov, Mnih et Hinton, 2007), la machine à vecteurs de support (*Support vector machine*) (Grčar *et al.*, 2006), et la décomposition en valeurs singulières (*Singular Value Decomposition, SVD*) (Bell, Koren et Volinsky, 2007; Paterek, 2007; Koren, 2008; Takács *et al.*, 2008).

IV.3.3.1 Modèle de facteur latent

Le modèle de facteur latent (*Latent Factor Model*) est considéré comme une approche la plus avancée dans le SR (Aggarwal, 2016). L'hypothèse de base est qu'il existe des représentations inconnues des utilisateurs et des items dans une dimension réduite où les affinités de l'utilisateur-item peuvent être modélisées plus précisément (Sammur et Webb, 2017). En reposant sur les techniques de modélisation statistique, cette approche a pour objectif de détecter ces représentations inconnues, à savoir, les facteurs sémantiques latents qui expliquent les notes observées. Ces facteurs sémantiques latents permettent de découvrir les classes latentes des objets (les items et les utilisateurs) ainsi que les relations latentes entre les utilisateurs et les items afin de construire des modèles pour réaliser les prédictions et proposer aux utilisateurs des recommandations (Su et Khoshgoftaar, 2009).

L'idée d'exploiter les facteurs sémantiques latents est établie dans le domaine de la RI dans les années 1980. Cette technique est connue sous le nom de l'analyse sémantique latente (*Latent Semantic Analysis, LSA*) qui est aussi appelée l'indexation sémantique latente (*Latent Semantic Indexing, LSI*) (Jannach *et al.*, 2010) pour aborder le problème de la sensibilité aux données manquantes. Elle est ensuite introduite dans le domaine des SR pour compléter la matrice utilisateur-items (Sarwar *et al.*, 2000b; Aggarwal et Parthasarathy, 2001; Goldberg *et al.*, 2001; Canny, 2002). La technique de LSA projette les données¹³³ dans une espace de la

¹³³ Dans le contexte de RI, les données sont les requêtes et les documents ; dans le cadre de SR, les données désignent la matrice de notation utilisateur-item.

dimension définie par variables latentes¹³⁴ grâce auxquelles les classes latentes peuvent être découvertes. Ainsi, les utilisateurs et les items sont classifiés. Vu que ces classes latentes sont plus denses et plus réduites que la matrice de notation, de nouvelles relations entre les utilisateurs et les items peuvent être découvertes mêmes si les utilisateurs n'ont noté aucun item en commun (Rosario, 2000; Ekstrand, Riedl et Konstan, 2011). Étant donné que la dimension définie par les variables latentes est plus réduite que l'espace original, ce type de technique est appelée la « réduction de la dimension » ou la « réduction dimensionnelle ». La technique de la décomposition en valeurs singulières (*Singular Value Decomposition*, SVD) et de l'analyse en composantes principales (*Principal Component Analysis*, PCA) sont deux méthodes bien connues pour réaliser la réduction de la dimension (Rosario, 2000; Jannach *et al.*, 2010; Aggarwal, 2016).

Dans le contexte de SR, ces deux techniques projettent les données dans une dimension plus réduite en factorisant la matrice de notation utilisateur-item. Ainsi, cette approche est appelée souvent l'approche de factorisation de la matrice (*Matrix Factorization*). Dans le domaine du FC basé sur le modèle, des méthodes diverses ont été proposées pour construire le modèle de facteurs latents (*Latent Factor Model*, LFM) : on peut citer l'analyse sémantique latente probabiliste (*probabilistic Latent Semantic Analysis*, pLSA), les réseaux de neurones (*neural networks*), l'allocation de Dirichlet latente (*Latent Dirichlet Allocation*) et surtout le modèle basé sur la factorisation des matrices de notation utilisateur-item, appelé également le modèle basé sur la décomposition en valeurs singulières (*Singular Value Decomposition*, SVD) (*SVD-based Model*).

Factorisation de la Matrice

Pendant le concours du prix Netflix, la méthode de factorisation de la matrice (Koren, Bell et Volinsky, 2009) a attiré beaucoup d'attention. Cette méthode est employée par de nombreuses équipes participantes car elle s'est avérée particulièrement utile pour améliorer la précision prédictive des SR grâce à sa flexibilité (Bennett et Lanning, 2007; Koren, Bell et

¹³⁴ Les variables latentes désignent les variables moins bien définies et des variables complètement ininterprétables. Par opposition aux variables qui peuvent être définies, comme le genre, le réalisateur du film, etc. (Koren, Bell et Volinsky, 2009).

Volinsky, 2009; Jannach *et al.*, 2010). Cette méthode est utilisée pour caractériser à la fois les utilisateurs et les items par les vecteurs des facteurs latents déduits selon les modèles d'évaluation (Koren, Bell et Volinsky, 2009; Jannach *et al.*, 2010). C'est l'une des approches du modèle de facteur latent. En effet, certains modèles de facteurs latents les plus fructueux sont basés sur la technique de factorisation de la matrice (Koren, Bell et Volinsky, 2009). La sophistication de cette approche est l'une des raisons pour lesquelles le modèle de facteur latent devient l'approche la plus avancée pour les SR (Aggarwal, 2016).

Cette méthode emploie les techniques de réduction dimensionnelle telle que l'analyse en composantes principales (*Principal Component Analysis*, PCA) (Goldberg *et al.*, 2001) ou la SVD (Bell, Koren et Volinsky, 2007) pour factoriser la matrice de notation utilisateur-item dans le but d'exploiter les facteurs latents. La méthode de factorisation de la matrice réduit la dimension en factorisant la matrice de notation utilisateur-item en plusieurs matrices plus simples. Plus précisément, cette méthode cherche à donner une approximation de la matrice de notes utilisateur-item R $m \times n$ du rang r par une matrice $\hat{R} = PQ^T$ de rang $k \leq r$, où P est la matrice unitaire représentant $m \times k$ facteurs d'utilisateurs dans l'espace réduite (ou l'espace latent) et la matrice Q dispose $n \times k$ facteurs d'items dans l'espace réduit (où l'espace latent). P et Q sont deux matrices orthonormales. Ces deux matrices, P et Q sont déterminées en minimisant l'erreur de reconstruction définie par la norme de Frobenius au carré :

$$\begin{aligned} err(P, Q) &= ||R - PQ^T||_F^2 \\ &= \sum_{u,i} (r_{ui} - p_u q_i^T)^2 \end{aligned}$$

où r_{ui} est la note de l'item $_i$ faite par l'utilisateur $_u$ et \hat{r}_{ui} est la note estimée. La u -ième ligne de P est $p_u \in \mathbb{R}^k$ qui est le vecteur représentant l'utilisateur $_u$ projeté dans l'espace latent en dimension k . Pareillement, la i -ième ligne de Q est $q_i \in \mathbb{R}^k$ qui est le vecteur représentant l'item $_i$ projeté dans cet espace latent. Le vecteur q_i mesure si l'item $_i$ possède les facteurs latents positivement ou négativement. Le vecteur p_u mesure la préférence de l'utilisateur $_u$ sur ces facteurs latents positivement ou négativement (Koren, Bell et Volinsky, 2009).

La technique utilisée pour minimiser cette erreur de reconstruction est SVD parce que cette minimisation équivaut à l'identification de la décomposition en valeur singulière (Van Loan et Golub, 1983; Desrosiers et Karypis, 2011). La théorie de SVD (Golub et Kahan, 1965)

affirme qu'une matrice R peut être décomposée en trois matrices plus simples (voir Figure 21) Alors, il existe une factorisation de la forme :

$$SVD(R) = U \times \Sigma \times V^T$$

où U est une matrice unitaire $m \times m$, Σ une matrice diagonale $m \times n$, et V est une matrice de $n \times n$. La SVD de R donne la meilleure approximation linéaire de R en choisissant les k premières colonnes d' U , les k premières valeurs singulières de Σ et les k premières lignes de V .

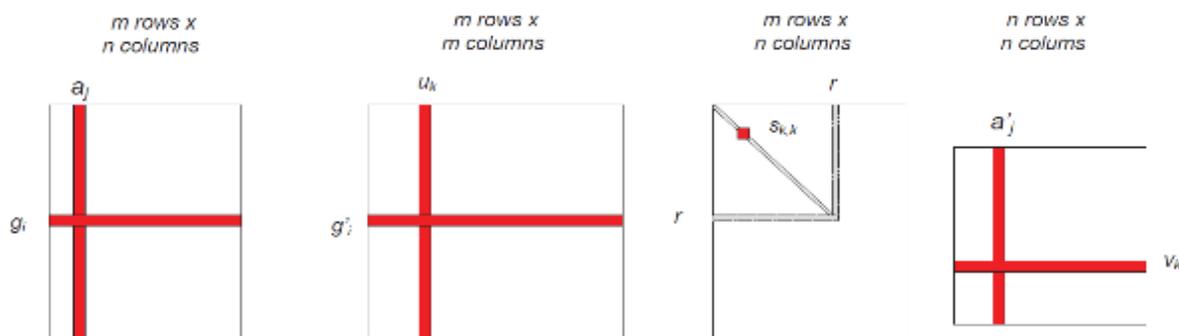


Figure 21 La décomposition en valeurs singulières

Dans le contexte de SR, la matrice de notation utilisateur-item, soit une matrice $m \times n$ (m utilisateurs et n items) peut être factorisée en trois matrices (Figure 22 La décomposition en valeurs singulières: une matrice items \times caractéristiques peut être décomposée en trois éléments différents: un item \times concepts, la force de chaque concept et un concept \times caractéristiques (traduction de (Ricci et al., 2011)). U est une matrice $m \times k$ (m utilisateurs et k facteurs latents), Σ est une matrice $k \times k$ et V est une matrice $n \times k$ (n items et k facteurs latents) où $k \leq m$ et $k \leq n$. L'enjeu de ce type de décomposition est que nous pouvons approximer la matrice complète en observant seulement les caractéristiques les plus importantes, à savoir les caractéristiques avec les valeurs singulières les plus larges. Ces caractéristiques les plus importantes sont les facteurs latents que nous chercherons à découvrir. Ces facteurs latents caractérisent les utilisateurs et les items. Autrement dit, ils classent les utilisateurs et les items dans les groupes latents qui ne sont pas toujours interprétables (Koren, Bell et Volinsky, 2009; Jannach *et al.*, 2010).

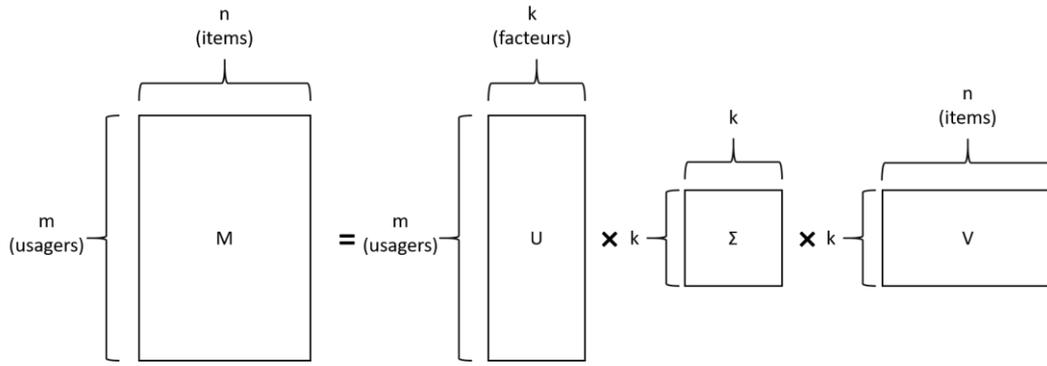


Figure 22 La décomposition en valeurs singulières: une matrice items \times caractéristiques peut être décomposée en trois éléments différents: un item \times concepts, la force de chaque concept et un concept \times caractéristiques (traduction de (Ricci et al., 2011)).

Dans ce cas, les matrices P et Q peuvent être obtenues : $P = U\Sigma^{\frac{1}{2}}$ et $Q = V\Sigma^{\frac{1}{2}}$. La prédiction de l'item_i pour l'utilisateur_u est calculée :

$$\hat{r}_{ui} = p_u q_i^T$$

La prédiction \hat{r}_{ui} décrit l'interaction entre l'utilisateur_u et l'item_i, soit la préférence générale de l'utilisateur_u sur les caractéristiques de l'item (Koren, Bell et Volinsky, 2009).

Néanmoins, cette application de SVD pose un problème à cause du manque de données. En réalité, la matrice de notation utilisateur-item est toujours si incomplète que la plupart des valeurs de r_{ui} sont indéfinies. Dans cas, il est impossible d'appliquer la SVD pour décomposer des matrices incomplètes. Une solution est simplement d'imputer des valeurs par défaut aux valeurs manquantes. Pourtant, l'imputation des valeurs peut être très coûteuse car elle augmente considérablement la quantité de données. En outre, une imputation inexacte peut considérablement déformer les données (Sarwar et al., 2000b; Koren, Bell et Volinsky, 2009; Desrosiers et Karypis, 2011).

L'autre solution proposée (Bell et Yehuda, 2007; Paterek, 2007; Takács et al., 2008; Funk, 2009) est d'utiliser uniquement les notes observées en évitant le surapprentissage et en utilisant un modèle régularisé pour définir les vecteurs p_u et q_i :

$$err(P, Q) = \sum_{r_{ui} \in R} (r_{ui} - p_u q_i^T)^2 + \lambda(|p_u|^2 |q_i|^2)$$

où λ est un paramètre qui contrôle le niveau de régularisation, $|p_u|^2 |q_i|^2$ est utilisé pour

éviter le surapprentissage. La formule peut être améliorée en ajoutant les biais de l'utilisateur, des items, la prédiction est calculée :

$$\hat{r}_{ui} = \mu + b_i + b_u + p_u q_i^T$$

où μ désigne la moyenne des notes, b_i et b_u sont respectivement le biais de l'item $_i$ et de l'utilisateur $_u$ qui indiquent les écarts observés. En tenant compte que la préférence d'un utilisateur évolue ainsi que la popularité d'un item, les biais du temps sur les utilisateurs et les items peuvent être ajoutés :

$$\hat{r}_{ui} = \mu + b_i(t) + b_u(t) + q_i^T p_u(t)$$

où $b_i(t)$, $b_u(t)$, $p_u(t)$ sont les fonctions pour mettre à jour respectivement le biais de l'item $_i$, le biais de l'utilisateur $_u$ et la préférence de l'utilisateur $_u$. Koren et Bell ont proposé la méthode qu'il appelle « SVD++ » pour perfectionner la prédiction en prenant compte les rétroactions implicites d'un utilisateur (Koren et Bell, 2015).

$$\hat{r}_{ui} = \mu + b_i(t) + b_u(t) + q_i^T (p_u(t) + |R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} y_j)$$

où $|R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} y_j$ représente les rétroactions implicites de l'utilisateur $_u$ qui complète ses notes explicites. $R(u)$ est la note moyenne globale faite par l'utilisateur $_u$ et $y_i \in \mathbb{R}^k$ est le vecteur représentant les facteurs de l'item $_i$ (Koren et Bell, 2015).

IV.3.3.2 Modèle de probabiliste

Une autre approche largement utilisée pour construire un modèle est le modèle probabiliste. L'idée de base est de modéliser les notes des items faites par les utilisateurs par un modèle probabiliste pour la prédiction de notes dans le futur. Le modèle probabiliste pense le problème de prédiction comme un problème de calcul de la valeur d'une note attendue (Breese, Heckerman et Kadie, 1998). Il consiste à représenter le calcul des prédictions sous forme de distribution de probabilité (Schafer *et al.*, 2007). Dans cette approche, la probabilité de chaque valeur possible de notes est calculée. La prédiction de l'item $_i$ pour l'utilisateur $_u$ est calculée :

$$\hat{r}_{ui} = E(r|u, i) = \sum_{r \in R} r \cdot P(r|u, i)$$

où R est l'ensemble des valeurs possibles pour une note ; $P(r|u, i)$ représente la probabilité que l'utilisateur u donne la note r à l'item i .

Classifieur bayésien utilisé dans l'approche basée sur le modèle

Les classifieurs bayésiens sont largement utilisés dans le SR basé sur le contenu comme nous avons présenté dans la section précédente. Ils sont également populaires pour les SR basé sur FC et notamment pour l'approche basée sur le modèle (Amatriain *et al.*, 2011).

Dans le cadre de SR basé sur le modèle, deux méthodes principales existent : le classifieur bayésien naïf et les réseaux bayésiens (Breese, Heckerman et Kadie, 1998).

Le modèle naïf bayésien est un modèle génératif qui est souvent utilisé pour la classification (Aggarwal, 2016). Breese Heckerman et Kadie ont proposé ce modèle pour le SR basé sur le modèle. L'idée est qu'il existe certains groupes ou types d'utilisateurs qui partagent un ensemble de préférences et de goûts en commun (Breese, Heckerman et Kadie, 1998). Ces groupes sont considérés comme des classes marquant C qui correspondent aux clusters des utilisateurs. La formule est :

$$P(C = c, v_1, \dots, v_n) = P(C = c) \prod_{i=1}^n P(v_i | C = c)$$

où $P(C=c, v_1, \dots, v_n)$ désigne la probabilité d'observer l'ensemble de valeurs d'un utilisateur de la classe C ; $P(C=c)$ est la probabilité d'observer les utilisateurs de la classe C et $P(v_i|C=c)$ est la probabilité conditionnelle d'observer certaine valeur de note v_i dans la classe C . $P(C=c)$ et $P(v_i|C=c)$ sont estimés en utilisant l'algorithme EM (*Expectation Maximization*) (Dempster, Laird et Rubin, 1977) qui permet d'apprendre et de définir le paramètre de ce modèle.

La prédiction de l'item i pour l'utilisateur u peut être faite en se basant sur la probabilité que l'utilisateur u appartienne à la classe C et sur la probabilité que la note soit donnée par les utilisateurs de classe C (Breese, Heckerman et Kadie, 1998; Jannach *et al.*, 2010).

Une autre méthode est le réseau bayésien (*Bayesian Network* ou *Bayesian belief network*) (Breese, Heckerman et Kadie, 1998). Un réseau bayésien est une méthode de classification largement utilisée dans le domaine d'informatique et de statistique (Su et Khoshgoftaar, 2009). C'est un modèle graphique probabiliste représentant des variables aléatoires sous la forme d'un graphe orienté acyclique (*directed acyclic graph*, DAG). Dans le réseau bayésien $G = (N, A)$, chaque variable aléatoire est représentée par un nœud $n \in N$ et chaque association probabiliste entre les variables est représentée par un arc $a \in A$ entre les nœuds. Breese et ses collègues proposent de construire un modèle à partir ces réseaux bayésiens en employant les arbres de décision pour représenter les probabilités (Breese, Heckerman et Kadie, 1998). Ces réseaux permettent d'explicitier les relations et les dépendances existées entre les variables (les items).

Probabilistique LSA

Probabilistique LSA (pLSA) est une méthode proposée pour découvrir les facteurs latents des communautés des utilisateurs et le modèle de préférence d'évaluation dans la base de données. Similaire à la technique de LSA, cette méthode consiste à identifier les relations « cachées ». Au lieu de se basant sur l'algèbre linéaire, elle repose sur la statistique (Hofmann, 2004; Jannach *et al.*, 2010). Cette méthode cherche à découvrir les classes latentes auxquelles les items et les utilisateurs peuvent être classifiés. La probabilité de l'utilisateur u donne une note r pour l'item i est calculée

$$P(r|u, i) = \sum_{z \in Z} P(r|i, z)P(z|u)$$

où Z est l'ensemble de classes latentes ; $P(z|u)$ représente la probabilité que l'utilisateur u appartienne à une classe latente z ; $P(r|i, z)$ représente la probabilité que l'item i appartenant à classe z ait une note r . Ainsi, la prédiction de l'item i pour l'utilisateur u peut être calculée :

$$\hat{r}_{ui} = \sum_{r \in R} (r \cdot \sum_{z \in Z} P(r|i, z)P(z|u))$$

IV.3.3.3 Règles d'association

Les règles d'association (*Association rules*) consistent à découvrir les règles d'association en reposant sur les bases de données transactionnelles (Agrawal et Srikant, 1994). Les règles sont des relations ayant un intérêt pour le statisticien entre deux ou plusieurs variables stockées dans de très importantes bases de données. Cette méthode est souvent utilisée dans les supermarchés sous la forme « si un client achète quelques produits, dans 70% de cas, il/elle achète aussi certains produits » dans le but de faire des promotions (Jannach *et al.*, 2010).

Dans cette approche, l'ensemble d'items est représenté par $I = \{i_1, \dots, i_m\}$ et l'ensemble de transactions est représenté par $T = \{t_1, \dots, t_n\}$, telles que t_i soit un sous-ensemble de I ($t_i \subseteq I$). Un itemset désigne une collection qui contient au moins un item. Un k -itemset désigne un itemset contenant k items. Un itemset est un sous-ensemble d'une transaction et une transaction contient au moins un itemset. Le compteur de support (*Support count*) d'un itemset X dans T est la fréquence de X , soit le nombre de transactions de $t_i \subseteq T$ qui contiennent X et qui est noté :

$$\sigma(X) = |\{t_i | X \subseteq t_i, t_i \in T\}|$$

où le symbol $|\cdot|$ dénote le nombre d'items dans l'un ensemble. Le support de l'itemset X est alors la proportion des transactions de T contenant X , soit :

$$support(X) = \frac{\sigma(X)}{|T|}$$

où $|T|$ est le nombre total de transactions. Le support de l'itemset X , $support(X)$ est une estimation de la probabilité $P(X)$. Un itemset fréquent est un itemset dont le support est supérieur ou égale à un *minsupport* (le seuil prédéfini).

La règle d'association est exprimée sous la forme $X \Rightarrow Y$, où X et Y sont les sous-ensembles de T et I , $X \cap Y = \emptyset$. La force d'une règle d'association est mesurée par son support et sa confiance. Le support de la règle $X \Rightarrow Y$ calcule la proportion de transactions de $t_i \subseteq T$ qui comprend à la fois X et Y (noté $X \cup Y$) dans la base de données, c'est-à-dire, la probabilité de la cooccurrence de X et Y dans une même transaction t_i , $P(X \cup Y)$ soit :

$$support(X \Rightarrow Y) = \frac{\sigma_r}{|T|} = \frac{\sigma(X \cup Y)}{|T|}$$

où σ_r représente le nombre de transactions de la règle $X \Rightarrow Y$, soit l'ensemble de t_i contenant à la fois l'itemset X et Y ($X \cup Y$). La « confiance » de la règle $X \Rightarrow Y$ la probabilité conditionnelle que Y apparaisse dans une transaction t_i du fait que X est compris dans cet itemset, $P(Y|X)$, soit la proportion de transactions de $t_i \subseteq T$ contenant X qui contiennent aussi Y :

$$conf(X \Rightarrow Y) = \frac{\sigma_r}{\sigma(X)} = \frac{\sigma(X \cup Y)}{\sigma(X)}$$

Afin de découvrir les règles d'association, un seuil du support c et un seuil de la confiance c doivent être définis. Une règle d'association dans T est alors toutes les règles dont le support et la confiance sont supérieurs ou égales aux seuils prédéfinis, soit support $(X \Rightarrow Y) \geq s$ et confiance $(X \Rightarrow Y) \geq c$. Une telle règle d'association $X \Rightarrow Y$ montre la relation entre deux itemsets X et Y : si les items de X apparaissent dans $t_i \subseteq T$, il est très probable que les items de Y sont des items d'une même transaction $t_i \subseteq T$ (Sarwar *et al.*, 2000a; Jannach *et al.*, 2010).

La découverte des règles d'association repose souvent sur deux étapes essentielles :

1. la découverte d'itemsets fréquents (*Frequent Itemset Generation*) consiste à identifier tous les ensemble d'items dont le support $\geq s$
2. la découverte des règles d'association (*Rule Generation*) consiste à découvrir, à partir des itemsets fréquents, les règles d'association dont la confiance $\geq c$.

Plusieurs techniques existent pour optimiser la découverte d'itemsets fréquents. L'approche la plus utilisée est l'algorithme APriori (Agrawal et Srikant, 1994).

Dans le contexte de FC, l'idée de base est de transformer la matrice de notation utilisateur-item dans la forme de transactions et d'appliquer les règles d'associations découvertes pour construire un modèle de recommandation (Jannach *et al.*, 2010; Aggarwal, 2016). Deux approches possibles peuvent être distinguées : « *item-wise* » et « *user-wise* » (Aggarwal, 2016). L'approche « *item-wise* » considère le rang de la matrice de notes comme une transaction qui indique l'ensemble de notes (positive) ou d'achats d'un utilisateur. Cette approche analyse à quel point il est probable qu'un utilisateur qui aime l'item 1 aime aussi l'item 2 (Item1 \Rightarrow Item2). L'approche « *user-wise* » considère la colonne de la matrice comme une transaction et transpose cette matrice. La question à analyser est : quelle est la probabilité qu'un item aimé par l'utilisateur A soit aussi aimé par l'utilisateur B (Utilisateur A \Rightarrow Utilisateur B).

Dans cette approche, on cherche à construire un modèle de recommandation avec l'ensemble des règles d'association découvertes avec de bonnes valeurs de support et de la confiance (Sarwar *et al.*, 2000a; Jannach *et al.*, 2010; Aggarwal, 2016). Le modèle est construit hors ligne et consiste à :

1. définir le seuil minimum du support et le seuil minimum de la confiance des règles
2. découvrir toutes les règles d'association dont le support et la confiance sont supérieurs aux seuils définis.

Une fois le modèle est construit, les recommandations pour l'utilisateur_u peuvent être réalisées dans 4 étapes suivantes :

1. de déterminer l'ensemble de règles d'association ($X \Rightarrow Y$) pertinentes pour l'utilisateur_u ;
2. de calculer l'union des items apparaissant dans le Y qui n'ont pas encore été achetés ou notés par l'utilisateur_u ;
3. de classer ces items par ordre en fonction la confiance de cette règle et de recommander la liste des Top-N items.
4. De renvoyer les N premiers items de cette liste ordonnée sous forme de recommandation.

Cette approche montre l'efficacité de la découverte de modèles de préférences en exploitant les règles d'association. Pourtant cette approche est similaire à l'approche FC basée sur l'item mais moins flexible puisqu'elle nécessite une notion explicite de transaction, *e.g.* la cooccurrence d'événements dans une session donnée (Amatriain *et al.*, 2011).

IV.3.3.4 Clustering

Un cluster est une collection d'objets (d'utilisateurs ou ' items) qui sont similaires entre eux et distincts des objets appartenant aux autres clusters (Han, Kamber et Tung, 2001). L'objectif d'un algorithme de clustering est de minimiser la distance entre les objets intra-cluster pour obtenir des clusters plus homogènes et de maximiser la distance entre les objets de

clusters différents pour obtenir des sous-ensembles bien différenciés. La similarité entre les objets est déterminée en mesurant la distance entre eux (Amatriain *et al.*, 2011). Deux mesures largement utilisées sont la distance de Minkowski et la corrélation de Pearson (Su et Khoshgoftaar, 2009).

Dans le contexte de SR, cette approche traite le problème de recommandation comme un problème de classification. Le but est de découvrir les clusters homogènes d'objets dans le système. Les clusters découverts sont des modèles qui peuvent être calculés et créés hors ligne.

Deux méthodes principales de clustering peuvent être distinguées : la méthode du partitionnement et la méthode hiérarchiques (Han, Kamber et Tung, 2001; Su et Khoshgoftaar, 2009; Amatriain *et al.*, 2011). La méthode du partitionnement divise les objets en clusters disjoints. Ainsi, chaque objet est attribué à un cluster défini. Les exemples typiques sont *k*-moyenne (*k-means*) (MacQueen, 1967) et *k*-medoids (Kaufman, Rousseeuw et Dodge, 1987) etc. La méthode hiérarchique regroupe successivement des objets dans des clusters trouvés et crée une décomposition hiérarchique représentée sous forme d'un arbre hiérarchique aussi appelé dendrogramme. Les applications typiques sont AGNES (Kaufman et Rousseeuw, 2009), DIANA (Kaufman et Rousseeuw, 2009), BIRCH (Zhang, Ramakrishnan et Livny, 1996) etc.

Les méthodes de clustering consistent à classifier N objets en k clusters. Ainsi est déterminé le nombre (k) de clusters à créer au début. Elles construisent les k clusters et les corrigent jusqu'à obtenir une similarité satisfaisante.

La méthode de *k*-moyenne (*k-means*) est l'une des méthodes du partitionnement les plus utilisées. Dans une base de données avec n objets, cette méthode consiste à créer k clusters disjoints C qui contient j objets. L'objectif est de classifier ces n objets de manière que les distances entre les objets de même cluster sont les plus proches que possible et les distances entre les objets de clusters différents sont les plus loin que possible. C'est-à-dire, les objets de même cluster sont les plus similaires entre eux et sont les plus dissimilaires aux objets d'autres clusters. Cette méthode est réalisée en 4 étapes :

1. de choisir k objets O_c (centroïdes) qui sont des points situés au centre de chaque cluster ;
2. d'affecter chaque objet O au cluster C_j dont la distance entre O et le O_c est minimale ;

3. de recalculer la position du centroïde O_c de chaque cluster en tenant compte l'influence des nouveaux objets affectée dans le cluster ;
4. de réaffecter chaque objet O au cluster C_j .

Cette opération itérée a pour l'objectif de la somme de distance entre l'ensemble des objets O et le centroïde O_c :

$$E = \sum_1^k \sum_{n \in C_j} d(O_n, O_c)$$

où O_n est le vecteur représentant le n-ième objet, O_c est le vecteur représentant le centroïde de cluster C_j et d est la distance mesurée. Cette opération continue jusqu'à E ne peut plus être diminué, à savoir, les centroïdes deviennent stables et ne changent plus. Cette opération peut être effectuée hors ligne en utilisant les données historiques. Les clusters découverts sont le modèle selon lequel les nouveaux objets peuvent être affectés au cluster propre.

La méthode de k-moyenne est efficace et facile (Su et Khoshgoftaar, 2009; Amatriain *et al.*, 2011). Cependant, elle montre certaines limites comme : le choix des centroïdes reste problématique qui suppose une connaissance préalable des données afin de choisir ces k centroïdes appropriés ; les clusters sont très sensibles à la sélection des centroïdes initiaux et aux données aberrantes ; risque de créer les cluster vides (Amatriain *et al.*, 2011).

Les méthodes hiérarchiques

La méthode hiérarchique utilise la matrice de distance comme critère de regroupement et produit un l'ensemble de clusters organisés hiérarchiques. Ce type de méthode ne nécessite pas de détermination du nombre de clusters au début, c'est-à-dire, le k n'est pas défini. Mais une condition d'arrêt est exigée.

La méthode hiérarchique peut être réalisée par l'agglomération de groupes similaires ou par la division de groupes dissimilaires. AGNES et DIANA représentent respectivement les deux applications typiques de ces deux méthodes (Figure 23).

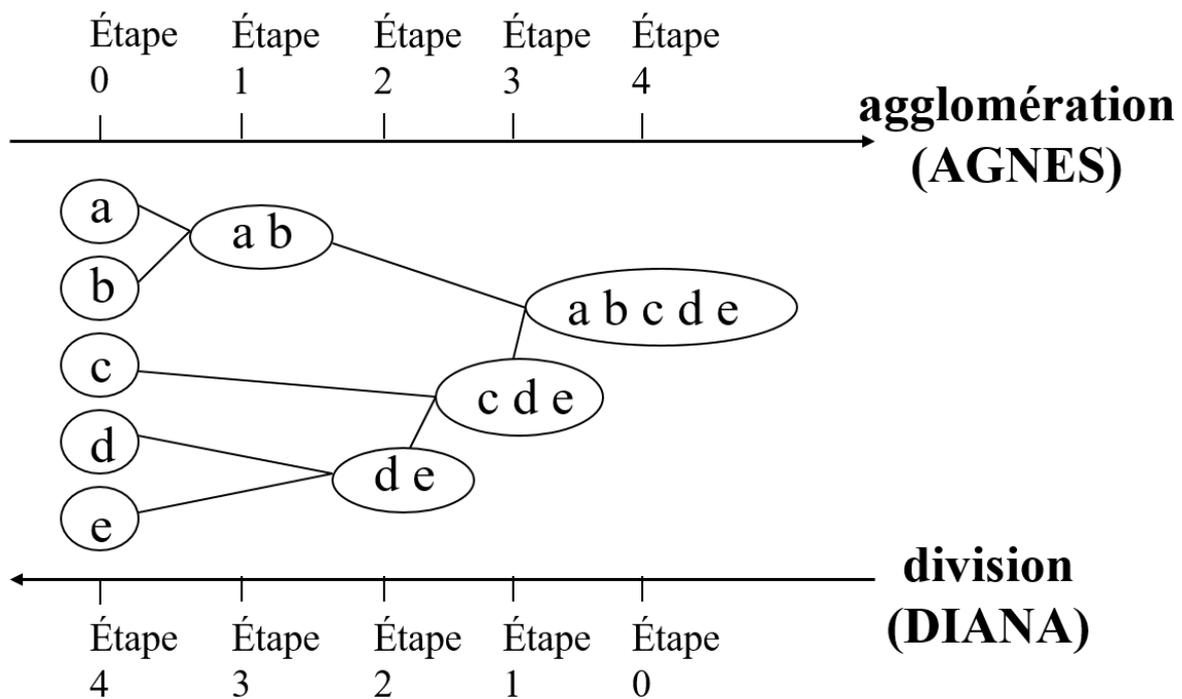


Figure 23 Deux applications de méthodes hiérarchiques

Le clustering hiérarchique par l'agglomération considère chaque objet comme un cluster au début. Il ensuite identifie les deux clusters les plus proches et groupe ces deux clusters en un nouveau cluster. Cette opération itère jusqu'il n'y a qu'un cluster (ou n clusters). Au contraire, le clustering hiérarchique par division commence à considérer l'ensemble d'objets comme un cluster. Il divise de manière hiérarchique les clusters selon un critère de dispersion des objets jusqu'au nombre de clusters est atteint (n clusters) ou chaque cluster ne contient qu'un objet.

Les modèles de clustering montrent une meilleure scalabilité aux méthodes du FC typiques comme ils font des prédictions dans des clusters beaucoup plus petits que dans l'ensemble de la base de données (Ng et Jordan, 2000; Sarwar *et al.*, 2000a; Zhao, Ye et Zhu, 2008; Su et Khoshgoftaar, 2009). Le calcul de clustering peut être effectuée hors ligne. Cette approche améliore l'efficacité de recommandation. Néanmoins, la qualité et l'exactitude de prédiction ne sont pas idéales (Su et Khoshgoftaar, 2009; Amatriain *et al.*, 2011).

IV.3.4 Avantages et limites

Un SR basé sur le FC résout certains problèmes du SR basé sur le contenu. Il est capable

de traiter n'importe quelle forme de contenu et de diffuser des ressources non nécessairement similaires à celles déjà reçues. C'est-à-dire qu'il peut recommander tous les types d'items y compris des items qui ne sont pas similaires à ceux que l'utilisateur a aimés dans le passé. Il a également comme avantages :

- d'utiliser les notes d'autres utilisateurs pour évaluer l'utilité des items ;
- de trouver des utilisateurs ou groupes d'utilisateurs dont les intérêts correspondent à l'utilisateur courant ;
- plus il y a d'utilisateurs plus il y a de notes : meilleurs sont alors les résultats.

Mais il montre quelques limitations. Premièrement, Il souffre du problème du démarrage à froid. Les nouveaux utilisateurs commencent par un profil vide et doivent le constituer à partir de zéro » (Berrut et Denos, 2003). Deuxièmement, l'item ne peut être recommandé qu'après avoir été évalué par un nombre considérable d'utilisateurs (Adomavicius et Tuzhilin, 2005). Cela signifie que le FC a des difficultés à recommander l'item qui n'est pas évalué et qui est évalué par peu d'utilisateurs. Dans le but de traiter ces problèmes, les SR actuels sont souvent hybrides et combinent les types de filtrage.

PARTIE 3 LES THEORIES DE L'IDENTITE SOCIALE ET DE LA DESINDIVIDUATION A L'EPREUVE DES PLATEFORMES DE CINEPHILIE

Le développement des RSN aggrave le problème de la surcharge d'informations et nécessite les SR. Les SR qui sont confrontés au problème de la manque de données peuvent bénéficier également de riches données sur les RSN. Il n'est pas surprenant que les SR soient largement appliqués dans les RSN pour aider les utilisateurs à découvrir des produits qui correspondent les mieux à leurs préférences personnelles. Néanmoins, la crédibilité et la fiabilité des évaluations fournies par les utilisateurs dans les RSN restent encore douteuses. En effet, le SR dans le RSN a deux faces complètement contradictoires : la personnalisation et la désindividuation. D'une part, l'objectif principal du SR est de fournir à chaque utilisateur des recommandations personnalisées qui correspondent à ses préférences personnelles. Il est donc essentiel que le SR obtienne des rétroactions de l'utilisateur qui ne sont pas influencés par d'autrui et qui reflètent ses opinions personnelles sur des items. D'autre part, les utilisateurs des RSN sont très susceptibles d'être influencés par les autres. Alors, il est nécessaire de vérifier si les pensées et les comportements des utilisateurs des RSN sont influencés par les autres et quels sont les facteurs qui les influencent. Par conséquent, nous abordons les théories psychologiques qui traitent les influences ayant l'impact sur le comportement de l'individu dans le groupe.

Cette partie expose les théories de la désindividuation sur laquelle s'adosse la recherche menée dans cette thèse. Le chapitre V présente la théorie de l'identité sociale et les approches principales. Le chapitre VI est consacré aux théories de la désindividuation, les approches principales et le modèle de l'identité sociale des effets de la désindividuation. Dans le chapitre VII, nous présentons les 4 plateformes qui forment le terrain de cette recherche : Douban, Allociné, Maoyan et SensCritique qui concernent notre recherche.

Chapitre V La théorie de l'identité sociale

La théorie de l'identité sociale est l'une des théories les plus influentes de la psychologie

sociale. Elle fournit de nombreuses explications sur le comportement intergroupes en se focalisant sur des relations intergroupes, les relations entre groupes et le concept du soi (Turner *et al.*, 1987; Tajfel et Turner, 2001; Hogg, 2003, 2005, 2006; Abrams et Hogg, 2006). Le principe fondamental est que les individus tirent une partie de leur concept du soi des groupes et des catégories sociales auxquels ils appartiennent, c'est-à-dire, de leur identité sociale. La théorie de l'identité sociale, ainsi que l'approche ou la perspective qu'elle implique, englobe un certain nombre de sous-théories différentes portant sur les aspects sociaux-cognitifs, motivationnels, sociaux interactifs et macro-sociaux de la vie du groupe (Turner *et al.*, 1987). Les deux composantes fondamentales sont la « théorie de l'identité sociale des relations entre les groupes » de Tajfel et Turner (Tajfel et Turner, 1979) et la « théorie de l'identité sociale du groupe » de Turner (Turner, 1985), qui est ensuite devenue une théorie de la catégorisation autonome. Plus précisément, la théorie de l'identité sociale se concentre sur le rôle de l'identité dans les conflits intergroupes, tandis que la théorie de l'auto-catégorisation se focalise sur l'architecture socio-cognitive des processus d'identité sociale. Celles-ci constituent le fondement des théories de l'identité sociale et expliquent, par exemple, l'action collective, la cohésion du groupe, le leadership, l'influence sociale *etc.*

Reliant la structure et le contexte sociaux aux processus de catégorisation et d'identité, la métathéorie globale de l'approche l'identité sociale donne la priorité à la relation réciproque entre les relations intergroupes et la conception de soi, ainsi qu'au développement et à l'articulation de constructions théoriques aux niveaux social, cognitif, interactif, intergroupe et sociétal (Abrams et Hogg, 2004).

Fondé sur une critique systématique des conceptions individualistes du moi et des processus de groupe (Turner et Oakes, 1986), l'accent conceptuel est mis sur le soi collectif défini en termes de groupe et lié aux autres membres du groupe (Markus et Kitayama, 1991; Brewer et Gardner, 1996; Oyserman, Coon et Kimmelmeier, 2002). Les processus cognitifs et sociaux associés à l'identité sociale produisent des comportements intragroupes et intergroupes.

V.1 L'étude des relations intergroupes

Les relations intergroupes désignent la façon dont les membres (individus) d'un groupe perçoivent, pensent, ressentent et agissent envers les membres (individus) d'autres groupes.

La définition des relations intergroupes et des comportements intergroupes la plus classique est proposé par Sherif :

« Les relations intergroupes désignent les relations entre deux groupes ou plus et leurs membres respectifs. Chaque fois que des individus appartenant à un groupe interagissent, collectivement ou individuellement, avec un autre groupe ou ses membres en termes d'identification de groupe, nous avons une instance de comportement intergroupe » (Sherif, 1966).

Tajfel et Turner définissent un « groupe » comme « un ensemble d'individus qui se perçoivent comme membres d'une même catégorie, qui attachent une certaine valeur émotionnelle à cette définition d'eux-mêmes et qui ont atteint un certain degré de consensus concernant l'évaluation de leur groupe et de leur appartenance à celui-ci » (Tajfel et Turner, 1986, 2001; Autin, 2010).

La définition plus précise peut être formulée comme : tous les aspects de l'interaction humaine qui impliquent des individus se percevant comme membres d'une catégorie sociale (d'un groupe), ou étant perçus par autrui comme appartenant à une catégorie sociale (d'un groupe) (Taylor et Moghaddam, 1994).

« Les relations entre les groupes peuvent varier considérablement dans leur extrémité - des images généralisées inoffensives, de la tolérance et la rivalité amicale à la haine profondément ancrée, l'intolérance et les conflits violents...cette dernière forme de comportement intergroupe est responsable des injustices et inhumanités les plus effroyables, l'étude des relations intergroupes a eu tendance à se concentrer sur cette forme extrême de relations intergroupes - préjugés, discrimination, fanatisme, agression intergroupe, etc. » (Hogg, 2013).

Les études des relations intergroupes définissent non seulement la relation intergroupe mais également la relation entre l'individu et le groupe auquel il appartient.

Ces études montrent que le comportement des individus peut être influencé par le groupe ainsi que par les membres du groupe. En fait, cette influence n'est pas seulement présente dans les groupes réels, mais également dans les groupes virtuels en ligne tels que les RSN qui concernent notre recherche.

V.2 Les approches principales

De nombreuses études dans ce domaine ont été menées à partir des années 1950 par les psychologues. Plusieurs approches diverses peuvent être identifiées. Une grande partie des travaux sur des relations intergroupes se focalise sur les *patterns* des préjugés et des discriminations individuelles en mettant l'accent sur les différences d'individus et de personnalité. Cette approche se concentre sur les processus psychologiques de l'intra-individuel ou interpersonnels conduisant à des attitudes préjudiciables ou à un comportement discriminatoire (Tajfel et Turner, 1986).

Les psychologues tentent ici de développer des explications sur les comportements extrêmes et pathologiques et sur les préjugés. Ils argumentent que le comportement de préjugé et de discrimination est causé par une dysfonction de personnalité liée aux expériences d'enfance qui les prédisposent à être extrêmement ethnocentriques et intolérants.

Adorno et ses collègues proposent la théorie de la personnalité autoritaire en adoptant le cadre de la psychanalyse (Adorno *et al.*, 1950). La personnalité autoritaire est considérée comme un facteur clé qui est responsable de la tendance à l'ethnocentrisme : attitudes positives à l'égard de l'intra-groupe (favoritisme de l'intra-groupe), attitudes négatives à l'égard des out-groupes (préjugé de l'out-group). Cette personnalité autoritaire se formerait à la suite de la répression, lors de la petite enfance, de tendances agressives envers des parents trop sévères et s'actualiserait dans la projection de ces tendances sur les groupes minoritaires (Licata, 2007). Les individus ont une personnalité autoritaire qui les prédispose à des formes extrêmes de comportement intergroupe : l'obsédé du statut et de l'autorité, l'hostilité et l'agressivité envers les plus faibles (Hogg, 2003). Certaines études expliquent le préjugé est la manifestation irrationnelle d'une force causée par la frustration qui réside dans l'individu (Dollard *et al.*, 1939; Berkowitz, 1962, 1969, 1974).

La notion d'autoritarisme continue d'être populaire, mais en mettant l'accent sur la tendance des gens à se soumettre aux conventions et à l'autorité du groupe et à punir les déviants du groupe (Altemeyer, 1998). Une autre explication des préjugés est l'opinion de Rokeach selon laquelle certaines personnes ont une personnalité dogmatique et à l'esprit fermé qui les prédispose à l'ethnocentrisme, à l'intolérance entre groupes et aux préjugés (Rokeach, 1960). Sidanius et Pratto décrivent une analyse relativement sophistiquée, mais néanmoins

« différenciée » des relations intergroupes fondées sur le pouvoir exploitant, appelée théorie de la dominance sociale (Sidanius et Pratto, 2019).

Néanmoins, cette approche est critiquée car elle accentue les différences individuelles sans centrer leur attention sur « l'imbrication complexe du comportement individuel ou interpersonnel avec les processus sociaux contextuels des conflits intergroupes et leurs effets psychologiques » (« *The complex interweaving of individual or interpersonal behavior with the contextual social processes of intergroup conflict and their psychological effects* ») (Tajfel, 1981; Turner et Giles, 1981; Tajfel et Turner, 1986). Les psychologues sociaux ont tendance à préférer une description des préjugés axée sur la socialisation et les processus cognitifs (Hogg, 2013). L'explication de la personnalité et des différences individuelles négligent l'impact du contexte et de la situation. Les perspectives interactionnistes sont axées sur la relation entre la personnalité et le contexte social.

Les études expérimentales de Muzafer Sherif et ses collègues représentent une autre approche principale que Campbell qualifie la « théorie des conflits réels (*Realistic group Conflict Theory*, RCT) » (Campbell, 1965). Au lieu d'effectuer une analyse ascendante en mettant l'accent sur la personnalité et des différences individuelles comme une explication des comportements intergroupes (Hogg, 2013). Sherif *et al.* adoptent une perspective réellement « intergroupe » qui met l'accent sur les relations entre groupes en faisant une analyse descendante (Sherif, 1958, 1962, 1966; Sherif *et al.*, 1961).

Son premier objectif est d'établir les conditions qui sont à l'origine des conflits entre groupes sociaux et de dégager ainsi les éléments qui peuvent résoudre ces conflits. Son point de départ pour expliquer le comportement intergroupe réside dans ce que Sherif a appelé les relations fonctionnelles entre groupes sociaux (Sherif, 2017).

L'hypothèse centrale de cette théorie est qu' « un véritable conflit d'intérêts du groupe provoque un conflit intergroupe (*real conflict of group interests causes intergroup conflict*) » (Tajfel et Turner, 1986). Selon cette théorie, les relations intergroupes peuvent être soit compétitives, soit coopératives. Des conflits émergent lorsque l'objectif des groupes est incompatible, c'est-à-dire, dans des situations où des groupes sont en compétition pour l'obtention de ressources rares et valorisées, qu'elles soient de nature concrète (biens matériels ou territoires) ou abstraite (pouvoir, biens symboliques). Par contre, des conflits s'atténuent lorsque l'objectif des groupes est commun (le but supra-ordonné), c'est-à-dire, dans des

situations où l'obtention de ressources rares et valorisées nécessite la coopération intergroupe. La compétition intergroupe améliore le moral, la cohésion et la coopération au sein des groupes (Vinacke, 1964; Fiedler, 1967; Kalin et Marlowe, 1968). C'est-à-dire que les conflits réels d'intérêts des groupes augmentent l'identification et l'attachement positif à l'intérieur du groupe (Tajfel et Turner, 1986).

V.3 Approche de l'identité sociale

La théorie de l'identité sociale découle de l'ethnocentrisme qui explique le comportement intergroupe. Il fait référence au favoritisme au sein du groupe et au rejet de ceux qui sont en dehors du groupe. Les comportements intragroupes et intergroupes présentent des caractéristiques évidentes d'ethnocentrisme.

V.3.1 Théorie de l'identité sociale

Au début des années 1970, Henri Tajfel et son équipe mènent une série d'études sur les groupes minimaux dont l'objectif principal était de comprendre les conflits intergroupes et plus généralement les relations entre groupes (Tajfel, 1970, 1978; Tajfel *et al.*, 1971; Tajfel et Turner, 1979). Le paradigme du groupe minimal est largement adopté comme condition de base pour la recherche sur les relations intergroupes.

Dans ces études, les participants ont été assignés à des groupes conçus sur la base de critères arbitraires. Ces groupes sont « minimaux » car ils sont purement cognitifs, les participants sont divisés en deux groupes sur base d'un critère aléatoire ou peu impliquant. Ils ne connaissent pas les affectations des autres participants dans les deux groupes. L'idée est de créer une situation où il n'y a aucun conflit d'intérêts. Les groupes sont anonymes et il n'existe aucune hostilité intergroupe avant le début de l'expérience. Il n'y a aucune interaction intragroupe ou intergroupe. Néanmoins, des comportements de compétition entre les groupes et des comportements discriminatoires envers l'exogroupe sont observés car lorsqu'on demandait aux participants d'allouer des points à d'autres participants dont il ne connaissait que l'appartenance au groupe, ils allouaient systématiquement plus de points aux membres de leur groupe qu'aux

membres des autres groupes.

Ces études montrent que le fait simple d'une catégorisation arbitraire des individus, en l'absence de tout enjeu objectif ou de relation particuliers entre groupe ou entre individus, peut suffire à faire apparaître des comportements tels que des compétitions et des discriminations envers l'exogroupe. Les résultats de ces études permettent à Tajfel de distinguer l'identité sociale et l'identité personnelle qui corresponde aux processus du groupe ou l'intergroupe et aux processus individuels interpersonnels. Tajfel définit l'identité sociale comme « cette partie du concept de soi qui provient de la conscience qu'a l'individu d'appartenir à un groupe social (ou à des groupes sociaux), ainsi que la valeur et la signification émotionnelle qu'il attache à cette appartenance » (Tajfel, 1978). Le concept de soi peut changer selon les contextes lorsque la situation met en évidence différentes identités sociales (Turner *et al.*, 1987).

La théorie de l'identité sociale est développée dans le but d'expliquer comment les individus créent et définissent leur place dans la société. Tajfel intègre dans sa théorie trois processus fondamentaux : la catégorisation sociale ; l'auto-évaluation à travers l'identité sociale ; la comparaison sociale intergroupe. Ceux-ci permettent d'expliquer différentes formes de comportements du groupe, notamment les conflits intergroupes.

V.3.2 Théorie de l'auto-catégorisation

John Turner et ses collègues développent la théorie de l'auto-catégorisation qui complète et améliore la théorie de l'identité sociale (Tajfel et Turner, 1979; Turner, 1985, 1991; Turner et Oakes, 1986; Turner *et al.*, 1987). Ils étudient les facteurs cognitifs liés à l'identité sociale en précisant comment les gens interprètent leur propre position dans différents contextes sociaux et comment cela influence leur perception des autres (stéréotypes, par exemple), ainsi que leur propre comportement dans des groupes (influence sociale). Ensemble, la théorie de l'auto-catégorisation et la théorie de l'identité sociale peuvent être appelées l'approche de l'identité sociale.

La catégorisation sociale fait référence à la tendance des individus à percevoir eux-mêmes et les autres en tant que membres relativement interchangeables d'un groupe au lieu d'individus séparés et uniques. La comparaison sociale est le processus par lequel les individus déterminent

la valeur relative ou le statut social d'un groupe particulier et de ses membres. Tajfel et Turner adoptent une définition « subjectiviste » du groupe social : « une collection d'individus qui se perçoivent comme membres de la même catégorie sociale, partagent quelque engagement émotionnel dans cette définition commune d'eux-mêmes, et atteignent un certain degré de consensus à propos de l'évaluation de leur groupe et de leur appartenance à celui-ci » (Tajfel et Turner, 1986). Dans ce système théorique, le processus de catégorisation sociale joue un rôle primordial. Tajfel définit la catégorisation sociale comme « un système d'orientation qui crée et définit la place particulière d'un individu dans la société » (Tajfel, 1972). À la suite de ses travaux sur le processus de catégorisation avec Wilkes, Tajfel met en évidence les effets de la différenciation catégorielle : une accentuation des différences entre les éléments appartenant à des catégories différentes et l'accentuation des ressemblances entre les éléments appartenant à une même catégorie. Les objets catégorisés semblent plus similaires entre eux qu'ils le sont réellement ; et plus différents des objets classés dans d'autres catégories qu'ils le sont réellement. La fonction principale de ce processus est d'organiser et de réduire la complexité du réel. La catégorisation sociale permet à l'individu de reconnaître la manière dont la réalité sociale est divisée en catégories et, surtout, de savoir quelles sont les catégories auxquelles il appartient et quelles sont celles auxquelles il n'appartient pas.

John Turner et son équipe poursuivent les travaux entrepris dans le cadre de la théorie de l'identité sociale. C'est dans son prolongement qu'ils proposent une théorie dont l'objet central n'est plus les relations entre groupes sociaux, mais bien le groupe social. L'identité sociale est conçue comme un facteur explicatif des comportements intergroupes – particulièrement la discrimination et le conflit – dans la théorie de l'identité sociale ; elle devient l'objet même d'étude dans la théorie de l'auto-catégorisation. Cette théorie ne cherche pas à expliquer un type particulier de comportement, mais bien à comprendre dans quelles conditions des individus sont capables de se comporter comme un groupe. Elle a donc une portée plus générale puisqu'elle peut s'appliquer à toute situation où l'identité sociale est impliquée. Il s'agit, selon ses auteurs (Turner *et al.*, 1987), d'une théorie générale du groupe social.

La théorie de l'identité sociale et la théorie de l'auto-catégorisation s'inscrivent dans la perspective de l'étude des relations intergroupes et exercent une influence prépondérante sur les approches psychosociales du groupe. Ce sont les théories de base à partir desquelles les psychologues développent le modèle SIDE, qui fournit le cadre de nos recherches.

Chapitre VI La théorie de la « désindividuation »

La théorie de la désindividuation est une théorie des psychologues sociaux décrivant la psychologie individuelle et le comportement dans un groupe, inspirée des idées des sociologiques du début du 20ème siècle comme Gustave Le Bon et Gabriel Tarde (Tarde, 1890; Le Bon, 1900). Des études antérieures ont suggéré que la désindividuation est un état psychologique de la réduction de la conscience de soi et de perte d'individualité des individus dans le groupe. La désindividuation réduit la prise de conscience et le contrôle du comportement individuel. Elle entraîne des comportements qui ne se produisent pas lorsque les individus sont seuls, c'est-à-dire des comportements atypiques, qui ne sont pas l'agrégation de comportements individuel mais un comportement anti normatif (Zimbardo, 1969; Diener, 1980; Prentice-Dunn et Rogers, 1989; Postmes, Spears et Lea, 1999).

Du point de vue de l'identité sociale et de la catégorisation sociale, Reicher, Spears et Postmes estiment que les individus ne perdent pas la conscience de soi dans un état de la désindividuation : leur identité du groupe devient saillante (par rapport à leur identité personnelle). Les individus utilisent cette identité du groupe (au lieu de l'identité personnelle) de façon stratégique pour expliquer leur comportement qui est en conformité avec les normes du groupe (Reicher, Spears et Postmes, 1995). Les normes du groupe sont l'ensemble des règles de conduite qu'il convient de suivre au sein d'un groupe. Elle peut être formelle et écrite (lois, règlements) ou bien informelle et a pour but de garantir le bien vivre ensemble et la survie du groupe. Reicher et ses collègues ont proposé le « modèle d'identité sociale des effets de désindividualisation (*the Social Identity model of Deindividuation Effects (SIDE)* »), qui non seulement élucide les résultats de recherche incohérents de la théorie traditionnelle de la désindividualisation, mais aussi qui est un modèle étendu et appliqué dans d'autres domaines. Ce modèle est initialement proposé par Lea et Spears et développé dans une série d'études.

VI.1 La théorie de la désindividuation

Comme nous venons de l'indiquer, la théorie de la désindividualisation trouve son origine dans l'étude des foules par le sociologue français Gustave Le Bon en 1895. Dans son livre

Psychologie des foules, il décrit en détail la transformation des comportements des individus après la submersion dans des groupes. Le Bon croit que « dans certaines circonstances données, et seulement dans ces circonstances, une agglomération d'hommes possède des caractères nouveaux fort différents de ceux des individus composant cette agglomération. La personnalité consciente s'évanouit, les sentiments et les idées de toutes les unités sont orientés dans une même direction » et « une âme collective apparaît » (Le Bon, 1900). Cette âme collective a un caractère instinctif et est régie par une « loi d'unité mentale des foules ». De cette manière, l'agrégat se transforme en une foule organisée que Le Bon qualifie de « foule psychologique ». Les individus en foule acquiert trois caractères que l'on ne trouve que dans l'état de foule : l'anonymat (qui permet l'irresponsabilité), la « contagion » et la suggestibilité. « Évanouissement de la personnalité consciente, prédominance de la personnalité inconsciente, orientation par voie de suggestion et de contagion des sentiments et des idées dans un même sens, tendance à transformer immédiatement en actes les idées suggérées, tels sont les principaux caractères de l'individu en foule. Il n'est plus lui-même, il est devenu un automate que sa volonté ne guide plus » (Le Bon, 1900).

En 1952, Leo Festinger, Albert Pepitone et Theodore Newcomb, en reprenant les idées de Le Bon, introduisent pour la première fois le concept de « désindividuation » dans la psychologie sociale sous la forme de la théorie de la désindividuation (Festinger, Pepitone et Newcomb, 1952). Dans leurs études, ils constatent que la désindividuation se produit lorsque les individus sont rendus anonymes au sein d'un groupe. Cet anonymat entraîne la perte du sentiment de soi, appelé désindividuation, et favorise ainsi un comportement anti-normatif et désinhibé (Festinger, Pepitone et Newcomb, 1952). Festinger et ses collègues partagent l'opinion de Le Bon sur le comportement de la foule en soulignant que les individus sont effectivement submergés dans la foule, ce qui entraîne une diminution de leurs responsabilités. Cependant, ils distinguent la désindividuation de la psychologie des foules en réformant l'idée que la perte d'individualité au sein d'une foule est remplacée par la mentalité du groupe. Festinger et ses collègues argumentent que la perte de l'individualité entraîne la perte de contrôle sur les contraintes internes ou morales (Postmes et Spears, 1998).

Les recherches de Festinger et ses collègues n'ont pas précisé les causes qui pourraient provoquer le phénomène de désindividuation. Zimbardo étudie plus en détail les variables qui conduisent à la désindividuation et aux comportements connexes, dont les plus importantes sont : l'anonymat, la perte de responsabilité individuelle, l'excitation, la surcharge sensorielle,

les situations nouvelles ou non structurées, l'implication comportementale et la consommation de substances psychotropes. Philip Zimbardo signale que la désindividuation est influencée par le groupe. Il souligne que les variables internes et externes à l'individu peuvent toutes provoquer une désindividuation, mais qu'il s'agit essentiellement d'un processus intra-individuel (Zimbardo, 1969). Ses études suggèrent que les variables sont : l'anonymat, le sens des responsabilités partagées ou diffuses, de nombreux groupes, une perspective temporelle altérée, l'excitation, la surcharge d'informations sensorielles, une implication physique dans des actions du groupe, ou des états de conscience altérés. Ces variables accompagnent un état de désindividuation. Cet état est caractérisé par des changements dans la perception de soi et des autres tels que l'observation de soi et le souci de l'évaluation sociale sont réduits. Ainsi, le jugement des autres sur soi suscite peu d'inquiétude, ce qui crée une tendance à adopter des comportements normalement inhibés parce qu'ils transgressent les normes sociales. La plupart de ses recherches cherchent à modéliser les relations entre la désindividuation et le comportement antisocial.

L'étude de la désindividuation par Zimbardo a favorisé des recherches sur les aspects intrapsychiques de la désindividuation. En effet, les chercheurs remarquent que l'étude devrait se concentrer davantage sur les effets intrapsychiques de la désindividuation plutôt que sur des variables externes dans le but de donner une plus grande validité à la lumière des modèles cognitifs du traitement de l'information (Ziller, 1964; Singer, Brush et Lublin, 1965; Diener, 1979). De nombreuses études (Duval et Wicklund, 1972; Gergen, Gergen et Barton, 1973; Westford, Diener et Diener, 1973; Diener et al, 1975; Wicklund, 1975; Jorgenson et Dukes, 1976; Ickes, Layden et Barnes, 1978; Diener, 1979; Prentice-Dunn et Rogers, 1980) passent au niveau intrapsychique et analysent les phénomènes intrapsychiques de désindividuation par rapport à la conscience de soi. Ces études montrent que la concentration de l'attention consciente détermine le niveau de conscience de soi. Lorsque l'attention consciente d'un individu est concentrée sur lui-même, le niveau de conscience de soi est élevé, alors que lorsque l'attention consciente est concentrée sur l'extérieur, la conscience de soi est faible. Les individus qui affichent un niveau élevé de conscience de soi ont davantage tendance à se comporter selon des normes sociales que celles qui affichent un faible niveau. L'étude de Diener et Wallbom suggèrent que la réduction de la conscience de soi est cruciale pour l'apparition de la désindividuation. Ils indiquent que les restrictions de la conscience de soi sont réduites et les comportements anti-normatifs seraient induits lorsque l'attention consciente n'est pas centrée sur soi-même (Diener et Wallbom, 1976).

Prentice-Dunn et Rogers proposent un autre modèle de désindividuation qui distingue la conscience de soi publique de la conscience de soi privée. La conscience de soi publique désigne l'attention de l'individu sur lui-même en tant qu'objet social, alors que la conscience de soi privée concerne un objectif interne et personnel. Les auteurs déclarent qu'il y a deux types de variables qui affectent ces deux types de conscience de soi : les indices de responsabilité et les indices d'attention. Les indices de responsabilité fournissent des informations sur la mesure dans laquelle les individus seront tenus responsables de leurs comportements. Les indices d'attention mesurent l'attention de l'individu, par exemple en demandant à l'individu de se concentrer sur ce qu'il ressent au moment présent ou en se concentrant sur les réactions de quelqu'un d'autre que lui-même. Prentice-Dunn et Rogers croient que la réduction du niveau de conscience de soi publique n'est pas liée à la désindividuation, car les individus peuvent percevoir clairement leur comportement dans cette situation. Au contraire, la réduction du niveau de conscience de soi privée est liée à la désindividuation car, à ce moment-là, l'individu ne perçoit pas ses propres pensées, émotions ni autres processus internes (Prentice-Dunn et Rogers, 1980, 1982, 1989).

Johnson et Downing proposent une explication différente de la désindividuation et un modèle qui tient compte de l'interprétation par un individu de la norme perçue d'une situation (Johnson et Downing, 1979). Les auteurs affirment que l'anonymat induit la désindividuation et que la personne désindividué(e) est davantage susceptible de suivre les normes saillantes dans une situation particulière. En conséquence, si un individu se trouve dans une situation de désindividuation où il existe une tendance au comportement prosocial, l'individu a tendance à réagir de manière prosociale. En revanche, si un individu est exposé dans une situation où les normes antisociales sont saillantes, le comportement antisocial sera stimulé. Cette étude de Johnson et Downing montre l'importance de la norme situationnelle qui est cruciale pour la création du modèle d'identité sociale des effets de la désindividuation (Johnson et Downing, 1979; Spears et Lea, 1992, 1994).

Reicher aborde l'influence des normes de situation (Reicher, 1984, 1987). Selon lui, le fait que le comportement anti-normatif peut se produire dans un groupe est une représentation de ce qui est normatif dans le groupe et dans la situation. Selon lui, il est essentiel de distinguer la norme sociale de la norme du groupe. La norme sociale concerne le contexte général de la vie quotidienne, alors que la norme du groupe est spécifique à chaque groupe dans certaines situations. Si un individu se trouve dans une situation de désindividuation, les normes de son

groupe se superposent aux normes sociales générales, même si elles sont incompatibles. Cette focalisation sur l'importance des normes situationnelles est à l'origine du modèle SIDE (Spears et Lea, 1992, 1994).

Reicher, Spears et Postmes procèdent à une évaluation globale des théories de la désindividuation en affirmant qu'elles contiennent toutes des facteurs fondamentaux en commun (Reicher, Spears et Postmes, 1995). Premièrement, elles incluent toutes le concept de « la perte de soi ou l'isolement ». Selon ces études, l'individu désindividualisé perd l'autorégulation, c'est-à-dire que l'individu perd de contrôle sur son propre comportement. La deuxième caractéristique commune est qu'elles considèrent le groupe comme l'une des variables antécédentes de la désindividuation et le groupe subvertit l'individualité, l'intellect et le contrôle du comportement. Troisièmement, la désindividuation entraîne la désinhibition du comportement.

VI.2 Le modèle SIDE

À partir des études de Reicher (Reicher, 1984, 1987) qui focalisent sur l'importance des règles de situation, le modèle de désindividuation appelé SIDE a été créé (Spears & Lea, 1992, 1994).

VI.2.1 La création du modèle de SIDE

Reicher, Spears et Postmes estiment que l'incohérence de nombreux résultats de recherche de la désindividuation est probablement due au fait que la théorie précédente repose sur deux hypothèses : (1) le concept de soi est la base de tout comportement rationnel ; (2) le groupe empêche l'individu de contrôler le soi et donc le comportement du groupe est toujours irrationnel (Reicher, Spears et Postmes, 1995). Les auteurs proposent le « modèle d'identité sociale de l'effet de désindividuation » (*Social Identity Deindividuation Effect, SIDE*), réinterprétant le phénomène de désindividuation (Reicher, Spears et Postmes, 1995) en se basant sur la théorie de l'identité sociale (Tajfel et Turner, 1979) et sur celle de l'auto-

catégorisation (Turner *et al.*, 1987). Leur point de vue fondamental est que la désindividuation n'est pas une perte de soi, mais la transformation d'une identité personnelle en une identité sociale. Il en résulte le respect des normes du groupe. Ils soulignent également que la désindividuation influence sur l'expression stratégique du comportement identitaire des membres du groupe, particulièrement lorsque les membres sont face à l'opposition de groupes extérieurs. La désindividuation donnera aux membres du groupe le « pouvoir » d'exprimer leur identité. Ces deux aspects constituent la dimension cognitive et la dimension stratégique de SIDE.

Postmes and Spears (Postmes et Spears, 1998) ont mené une méta-analyse de 60 études sur la désindividuation en constatant que :

1. L'effet de l'anonymat, de la taille du groupe et de la dispersion des responsabilités sur le comportement anti-normatif sont faibles.
2. En tant que médiateur entre la désindividuation et le comportement anti-normatif, l'effet de la conscience de soi n'est pas significatif. Autrement dit, il ne peut pas être prouvé que la désindividuation est provoquée par la réduction de la conscience de soi.
3. La corrélation la plus significative entre la désindividuation et la norme situationnelle est que, dans l'état de désindividuation, l'individu présente davantage de comportements conformes aux normes situationnelles. Cela contredit les théories précédentes de la désindividuation.
4. La corrélation entre la désindividuation et le comportement anti-normatif est plus élevée dans la condition du groupe que dans la condition individuelle ; plus la taille du groupe est grande, plus le comportement anti-normatif est fort.

Ce résultat confirme que la désindividuation est un phénomène de groupe. Les variables de désindividuation telles que l'anonymat, la taille du groupe et la conscience de soi ont peu d'impact sur les comportements qui violent les normes sociales générales (telles que les comportements antisociaux), mais augmentent le respect individuel des normes contextuelles.

Dans cette étude (Postmes et Spears, 1998), Postmes and Spears estiment que les études de la désindividuation peuvent être divisées en trois catégories. **La première** est appelée « la théorie classique de la désindividuation » qui déclarent que la désindividuation est associée à

une perte de contrôle ou de responsabilité lorsqu'un individu se trouve dans un groupe. Cette première catégorie est composée des études de Festinger et ses collègues (Festinger, Pepitone et Newcomb, 1952), de Zimbardo (Zimbardo, 1969) et de Diener (Diener, 2006). **La deuxième** catégorie est appelée « la théorie contemporaine de la désindividuation ». Elle correspond aux études de Prentice-Dunn et Rogers (Prentice-Dunn et Rogers, 1982, 1989) qui proposent la réduction de la conscience de soi privée comme principale variable inductrice de la désindividuation. **La troisième** catégorie est constituée par le modèle SIDE car c'est la seule théorie qui différencie les normes du groupe des normes sociales.

VI.2.2 La dimension cognitive de SIDE

La dimension cognitive de SIDE développe directement les points principaux de la théorie de l'identité sociale et de la catégorisation sociale. Tajfel propose la théorie de l'identité sociale lors de l'étude des relations entre groupes dans les années 1970 en soulignant que le soi est un système complexe qui comprend l'identité individuelle liée aux caractères personnels et l'identité sociale liée au groupe (Tajfel, 1970, 1978; Tajfel *et al.*, 1971; Tajfel et Turner, 1979). La théorie de l'auto-catégorisation de Turner (Turner, 1982, 1991; Turner *et al.*, 1987) indique que le soi se catégorise à trois niveaux, à savoir, bas niveau (tel que moi, toi), niveau intermédiaire (tel que nous, vous) et haut niveau (tel que l'humain). L'individu ne perd pas tout son soi dans le groupe, mais passe d'un niveau d'identité inférieur à un niveau d'identité supérieur. Ce changement d'identité rend l'identité sociale de l'individu plus saillante et rend son comportement plus conforme aux normes sociales.

Le modèle SIDE intègre ces idées et les applique à la désindividuation. Selon ce modèle, lorsque l'individu est dans le groupe ou manque d'indices individuels, la désindividuation favorise la transformation de l'identité individuelle en l'identité sociale. Ainsi, ses comportements seront en conformité avec les normes du groupe. SIDE fait la distinction entre les normes sociales générales et les normes situationnelles (ou du groupe) spécifiques. Dans un état de désindividuation, l'individu respecte les normes situationnelles (ou du groupe) qui pourraient déroger aux normes sociales générales. Par exemple, dans une situation expérimentale l'attaque des autres est une exigence d'une norme situationnelle qui n'est pas en conformité avec les normes sociales. Selon Reicher et de ses collègues, l'anonymat a un impact plus important sur l'individu lorsque son identité de groupe est plus saillante que son identité

individuelle. Cela suggère que l'individu immergé dans le groupe est contrôlé par des normes du groupe et que les variables de la désindividuation (anonymat) sont liées à l'auto-catégorisation individuelle et à la saillance de l'identité.

La théorie traditionnelle de désindividuation estime qu'un fonctionnement désindividué affaiblit la conscience de soi et le contrôle de soi, ce qui conduit à un comportement anti-normatif. En revanche, SIDE estime que la désindividuation n'affaiblit pas la conscience de soi, mais amène l'individu à respecter davantage des normes situationnelles. Les études montrent que les membres du groupe acceptent les normes agressives ou les normes d'assistance mutuelle. Plus le groupe est grand, plus le comportement des membres du groupe est cohérent avec les normes du groupe. Dans le processus de formation des normes de groupe, les membres d'un groupe n'acceptent pas immédiatement les normes disponibles. Au contraire, c'est sous l'interaction qu'ils organisent les indices de normes et affichent enfin des comportements conformes à ces normes.

VI.2.3 La dimension stratégique de SIDE

La dimension cognitive du SIDE suggère que les individus désindividualisés respectent davantage les normes lorsque leur identité sociale est saillante. Cependant, lorsque les membres d'un groupe font face aux autres groupes, une identité sociale saillante ne suffit pas à aider les membres à se comporter en conformité avec les normes de leur groupe. Par conséquent, les membres du groupe ont besoin d'une sorte de « pouvoir » pour permettre l'expression d'un comportement conforme aux normes du groupe. C'est-à-dire que lorsqu'un individu a une identité sociale saillante, la désindividuation affectera l'expression du comportement conforme à son identité saillante. C'est la dimension stratégique de SIDE, selon laquelle, lorsque les membres d'un groupe sont identifiables à l'externe, ils réduisent les comportements conformes aux normes de leur groupe telles qu'elles sont opposées aux autres groupes. Au contraire, lorsque les membres d'un groupe sont identifiables dans leur propre groupe, ils se comportent davantage en conformité avec les normes de leur groupe telles qu'elles se distinguent des autres groupes.

Une série d'études expérimentales de Reicher et ses collègues montrent que l'identifiabilité des individus par les groupes externes empêche les comportements qui ne sont pas acceptés par

des groupes externes. Cela confirme l'hypothèse de la dimension stratégique (Reicher, Spears et Postmes, 1995). Des études récentes montrent que l'identifiabilité des individus par les membres dans le groupe entraîne également un comportement stratégique. Barreto et Ellemers suggèrent que l'identifiabilité de l'individu par les membres du même groupe augmente les comportements des individus qui se conforment aux normes du groupe (Barreto et Ellemers, 2000, 2002).

En général, Le modèle SIDE comprend la dimension cognitive et la dimension stratégique liées respectivement à l'auto-catégorisation et à l'expression de soi. Le SIDE analyse les conditions qui facilitent ou entravent les deux, l'anonymat étant la condition la plus étudiée. Le SIDE explique non seulement raisonnablement les résultats incohérents des recherches précédentes, mais éclaire également la psychologie et le comportement de groupe en analysant les relations entre les individus et les groupes dans des conditions spécifiques, de manière à ce qu'elles puissent être vérifiées et appliquées dans un plus grand nombre de domaines.

VI.3 La désindividuation dans l'environnement CMC

Le SIDE est largement appliqué dans les domaines différents depuis la publication de la méta-analyse par Postmes et Spears (Postmes et Spears, 1998). De nombreuses études ont été menées dans des environnements de communication assistée par ordinateur (*Computer Mediated Communication, CMC*) (Vilanova *et al.*, 2017). Le CMC fait référence à la communication via des supports technologiques tels que des ordinateurs. Par rapport à la communication en face à face, la plus grande caractéristique de CMC est l'anonymat et la distance physique qui sont susceptibles d'induire la désindividuation (McKenna et Bargh, 2000).

SIDE s'oppose au point de vue des études précédentes qui estimaient que l'anonymat du CMC libérait l'individu des contraintes de la communication sociale en face à face, de sorte qu'il n'est pas contrôlé par les normes sociales, tandis que la distance physique éloigne l'individu des autres et crée des barrières entre eux (Christopherson, 2007). Les premières recherches de Spears et ses collègues révèlent que l'anonymat du CMC améliore le respect des normes du groupe par l'individu lorsqu'il dispose une identité du groupe élevée ou d'une identité personnelle faible (Spears, Lea et Lee, 1990). Dans une étude de Postmes et ses collègues, les auteurs constatent que les membres de groupe anonymes du CMC avaient des

comportements plus cohérents aux les normes du groupe (Postmes *et al.*, 2001a). D'autres études montrent également que, dans les CMC, la restriction de l'individualité et l'anonymat font passer la communication du niveau individuel au niveau du groupe, et que cette communication intergroupe augmente les stéréotypes et les préjugés et entraîne une attraction interne et un rejet des groupes externes. Cela renforce ainsi la frontière intergroupe (Lea, Spears et de Groot, 2001). Ces résultats montrent que l'anonymat au sein de la CMC ne sépare pas les individus du groupe, mais renforce l'influence du groupe sur les individus, ce qui entraîne la conformité aux normes du groupe, l'attraction du groupe ainsi que les stéréotypes et les différences intergroupes.

Chapitre VII Désindividuation et identité sociale à l'épreuve de quatre plateformes numériques de notations de films.

Méthodologie de la recherche

Dans ce chapitre, nous allons présenter les 4 plateformes où se déroule notre recherche. Nous comparerons les comportements des utilisateurs de ces quatre plateformes en matière de notations des films. Nous avons choisi 4 plateformes : un RSN pour les cinéphiles chinois : Douban ; deux RSN pour les cinéphiles français : Allociné et SensCritique ; une plateforme qui n'est pas un RSN (non-RSN) pour les cinéphiles chinois : Maoyan. Ces 4 plateformes partagent un même objectif principal : aider leurs utilisateurs à évaluer la qualité d'un film.

Les utilisateurs sont encouragés à noter des films plutôt en fonction de la qualité du film. Cela pourrait être l'une des normes principales qui définissent le comportement de notation des utilisateurs dans ces 4 plateformes. En d'autres termes, il est très probable que les notations des films sur ces 4 plateformes soient similaires. Pourtant, la théorie de SIDE indique que l'influence normative peut parfois exercer un fort impact sur le comportement de l'utilisateur dans les groupes de CMC (Spears et Lea, 1992, 1994; Postmes et Spears, 1998). Cela signifie que les utilisateurs de 3 RSN (Douban, Allociné et SensCritique) seraient plus influencés par les normes du groupe que les utilisateurs de non-RSN (Maoyan). Donc, nous examinons de ces 4 plateformes pour vérifier s'ils disposent l'environnement favorisant la formation et le renforcement de la norme du groupe. Nous présentons également les SR dont ces plateformes disposent pour vérifier si le comportement des utilisateurs influence la fonction de leur SR.

VII.1 Les 4 plateformes pour les cinéphiles de notre recherche

VII.1.1 Un RSN pour les cinéphiles chinois : Douban et Douban

Film

Lancé en 2005, Douban est l'un des premiers RSN chinois et il est encore très populaire. Il est basé sur les centres intérêts des utilisateurs. Il peut être considéré comme une combinaison de Goodreads, IMDb, Blogger, Facebook, Pandora, Spotify, Ello et Fancy. Douban est connu notamment pour ses commentaires générés par les utilisateurs sur les livres, les films ainsi que les musiques. Les utilisateurs de Douban sont très actifs sur le site et il continue à attirer de plus en plus de nouveaux utilisateurs 15 ans après sa fondation. En 2019, il y avait 82 millions d'utilisateurs enregistrés et 200 millions de visites chaque mois¹³⁵.

Douban est avant tout un RSN qui permet à l'utilisateur de créer et de gérer son profil public (Figure 24). Ce profil est accessible à tout le monde y compris aux visiteurs qui ne sont pas membres de Douban. L'utilisateur peut également abonner les autres utilisateurs et être abonné par autrui. Bien entendu, la liste de ses abonnements et de ses abonnées n'est accessible qu'aux autres utilisateurs enregistrés. À part cela, le profil montre aussi les informations sur ses centres d'intérêt : les films, les livres, les musiques qu'il aime ; les événements auxquels il participe, les groupes qu'il fréquente etc.

En fait, Douban est un RSN hybride : il est à la fois un RSN de réseautage et un RSN de navigation (voir II.3 Typologie des réseaux sociaux numériques). Non seulement il aide l'utilisateur à en rencontrer d'autres qui partagent les mêmes centres d'intérêt et des goûts proches, mais il lui offre aussi l'opportunité de découvrir des informations (dont la plupart concernent des produits culturels). Effectivement, Douban est équipé d'un SR hybride (le SR basé sur le contenu et basé sur le FC). Le SR facilite la découverte de ce qui est susceptible d'intéresser un utilisateur.

Étant donné que l'un des objectifs principaux de Douban est d'aider à la découverte de tout ce qui est intéressant, il est évident que le « centre d'intérêt » est non seulement l'élément

¹³⁵Source : <https://www.douban.com/partner/qa/28764296/>

central de Douban mais également le pivot des relations entre les utilisateurs. En principe, les utilisateurs de Douban partagent notamment des centres d'intérêt liés à « la vie et la culture ». Par conséquent, Douban n'est pas une plateforme seulement pour les cinéphiles. Il est en fait une plateforme multiproduits orientés « vie et culture » qui fonctionne sur le principe de l'UGC¹³⁶. Plus précisément, les trois catégories de centres d'intérêt les plus populaires de Douban sont le livre, le film et la musique qui ont chacun une section spécifiquement dédiée. Les utilisateurs enregistrés peuvent laisser leurs commentaires sur des livres, des films et de la musique dans ces trois sections : « Douban Livre ¹³⁷ », « Douban Film ¹³⁸ » et « Douban Musique¹³⁹ ». Les informations des produits culturels sur Douban sont considérées comme crédibles en raison de sa grande base d'utilisateurs, ce qui induit une grande quantité de commentaires et d'avis.

Toutefois, Douban ne se limite pas à aider l'utilisateur à découvrir des produits culturels correspondant à ses centres d'intérêt. Douban cherche également à faciliter la rencontre des utilisateurs partageant des mêmes centres d'intérêt et des goûts proches. Outre les trois sections dédiées au produit culturel, la section « Douban Groupe » leur permet de créer des groupes autour de leurs centres d'intérêt. En réalité, « Douban Musique » et « Douban Film » étaient à l'origine les deux groupes les plus populaires de « Douban Groupe ».

En ce qui concerne notre recherche, « Douban Film » est l'une des sections les plus populaires qui dispose plus de 130 milles films et plus de 4,42 millions de critiques d'utilisateurs jusqu'en août 2019¹⁴⁰. Les informations, les critiques, les notes des films de Douban sont citées par 80% des médias chinois¹⁴¹. Elle coopère avec les cinémas, les sites de billetterie en ligne et les sites web d'hébergement de vidéo. Ainsi les utilisateurs peuvent acheter des billets ou accéder à des films en ligne via Douban.

Depuis 2012, Douban a mis à disposition l'application mobile dont la structure principale est presque pareille que celle de la version desktop mais avec certaines nuances.

¹³⁶ UGC : *Le contenu généré par les utilisateurs (User-Generated Content)*

¹³⁷ <https://book.douban.com/>

¹³⁸ <https://movie.douban.com/>

¹³⁹ <https://music.douban.com/>

¹⁴⁰ Source : <https://www.csuldw.com/2019/09/08/2019-09-08-moviedata-10m/>

¹⁴¹ Source : <https://blog.douban.com/>

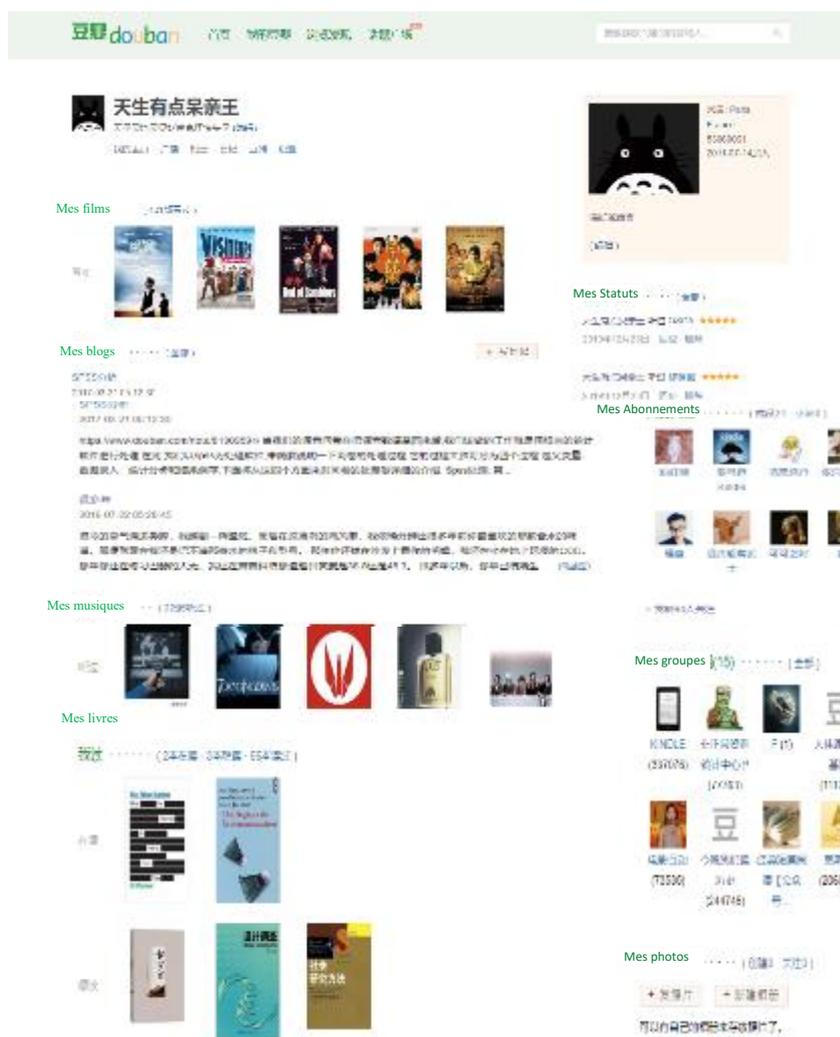


Figure 24 Profile d'un utilisateur sur Douban

La page d'accueil de Douban Film est divisée dans 3 parties principales, l'espace de recherche, la zone des films populaires et la zone des critiques populaires (Figure 25)¹⁴² :

- L'espace de recherche offre deux manières à découvrir et trouver les produits culturels :

- 1). par mots-clés : les utilisateurs peuvent entrer les mots-clés (le titre, les artistes *etc.*) dans le champ de recherche qui situé au début de la page pour chercher les films qu'ils les ont déjà connus.

¹⁴² <https://movie.douban.com/subject/1292063/?from=showing>

2). par catégorie : les utilisateurs peuvent découvrir les produits culturels de certains thèmes ou dans les genres qui les intéressent par catégorie ou par tags.

- La zone de produits culturels populaires ;

- La zone des critiques populaires offrent les critiques les plus populaires.

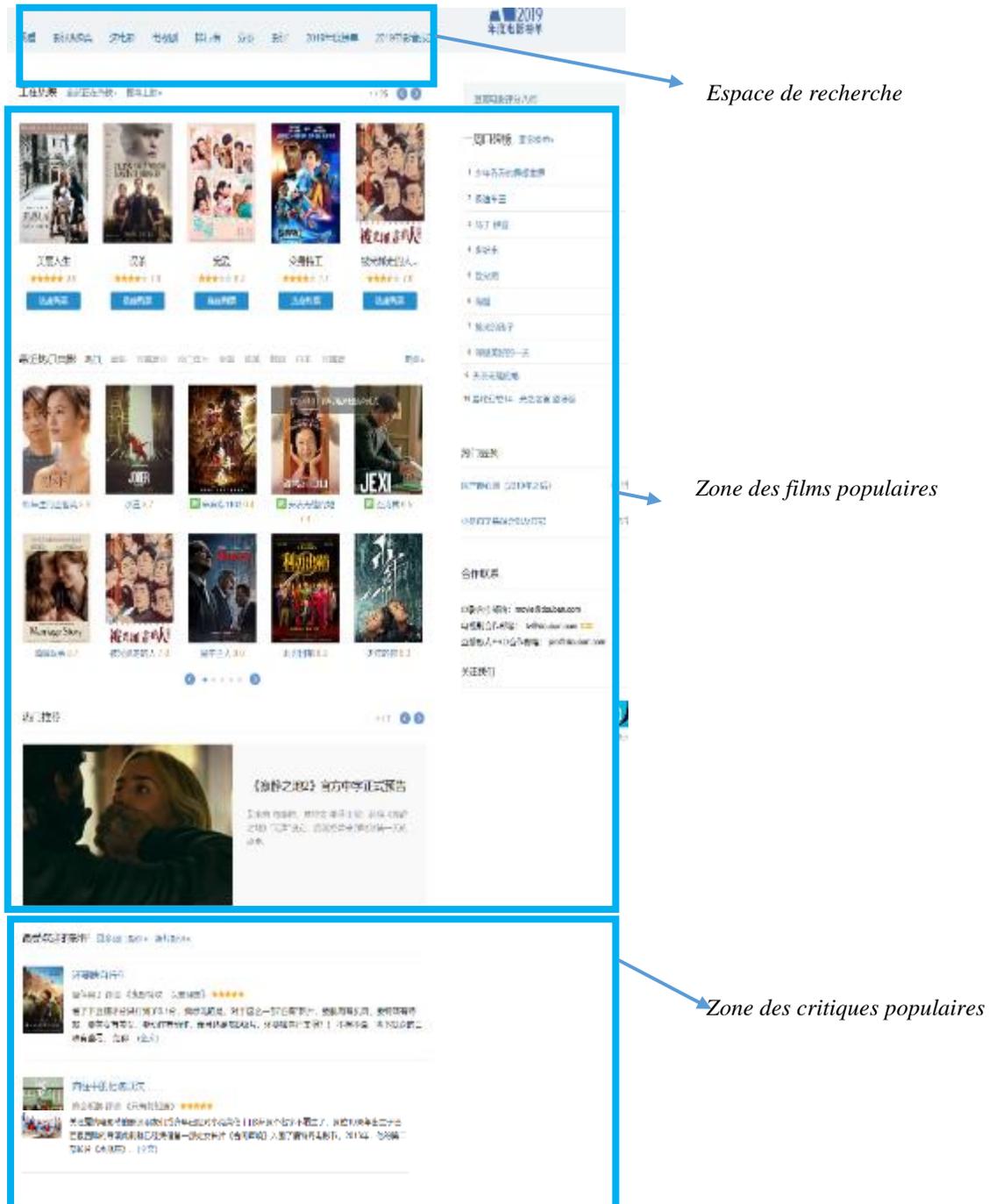


Figure 25 Page d'accueil de Douban Film (version desktop)

Dans la version mobile, la zone des critiques populaires est remplacée par la zone recommandation appelée « recommandation pour vous » où il y a des recommandations de films (basées sur les activités des utilisateurs dans le passé) et les raisons de la recommandation (e.g., les films de même réalisateur) (Figure 26).

La page d'accueil est une entrée pour accéder à la page d'un film particulier. Sa fonction principale est de faire connaître aux utilisateurs les films et de les aider à trouver les films qui les intéressent. Ils permettent aussi aux utilisateurs de découvrir des critiques et des commentaires des films.

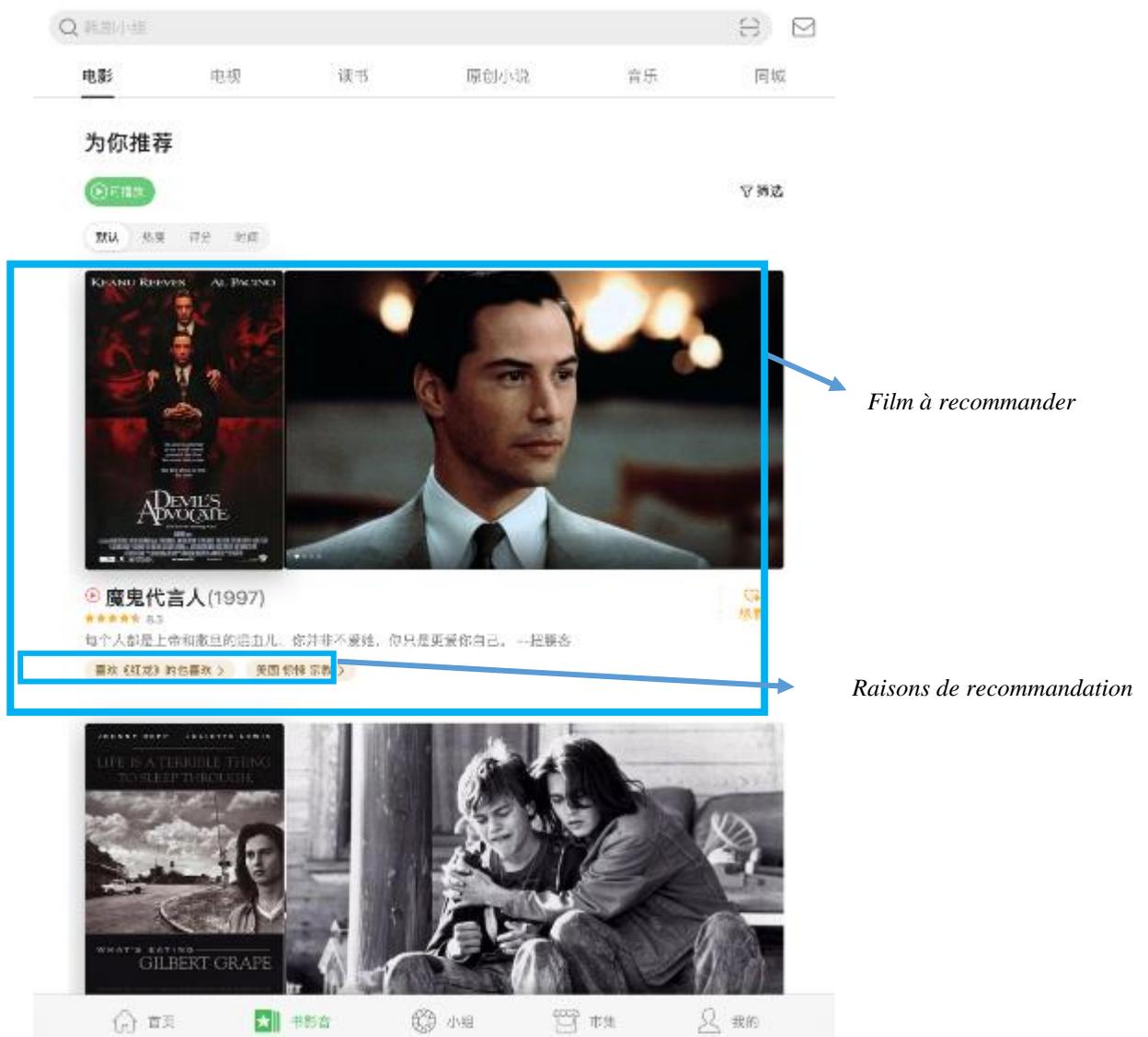


Figure 26 Zone de recommandation dans la version mobile

Chaque film a une page dédiée sur Douban. Les utilisateurs peuvent créer une page pour

un film qu'ils ne trouvent pas sur Douban. Dans chaque page d'un film, 4 types d'information sont proposés (Figure 27 et Figure 28):

- **L'information sur le film** telle que la date de sortie, le réalisateur, les acteurs/actrices, la synopsis, la bande-annonce, les photos *etc.*;
- **L'information pratique pour regarder le film** : les liens vers des billetteries en ligne et/ou les liens vers des sites de la vidéo à la demande (e.g, Youku, Tudou) ;
- **L'information pour découvrir d'autres film** : les recommandation (e.g., des films similaires) et des tags ajoutés au film;
- **L'information d'évaluation** : la notation, des critiques, des commentaires court et des discussions (lancées par les utilisateurs concernant le film).



Figure 27 Page d'un film particulier de la version desktop

Une seule nuance entre la version desktop et la version mobile est que si l'utilisateur

accède la page d'un film via son appareil mobile, d'autres œuvres des principaux artistes (les réalisateurs/réalisatrice, les acteurs/actrices et les scénaristes) de ce film sont recommandées. En effet, la version mobile propose davantage de recommandation que la version desktop.

Les utilisateurs peuvent également évaluer eux-mêmes le film (en notant le film sur une échelle de 1 à 5 étoiles, en laissant des commentaires courts et en rédigeant des critiques) et communiquer avec les autres (en lançant des discussions et en laissant des commentaires sur les critiques d'autrui).

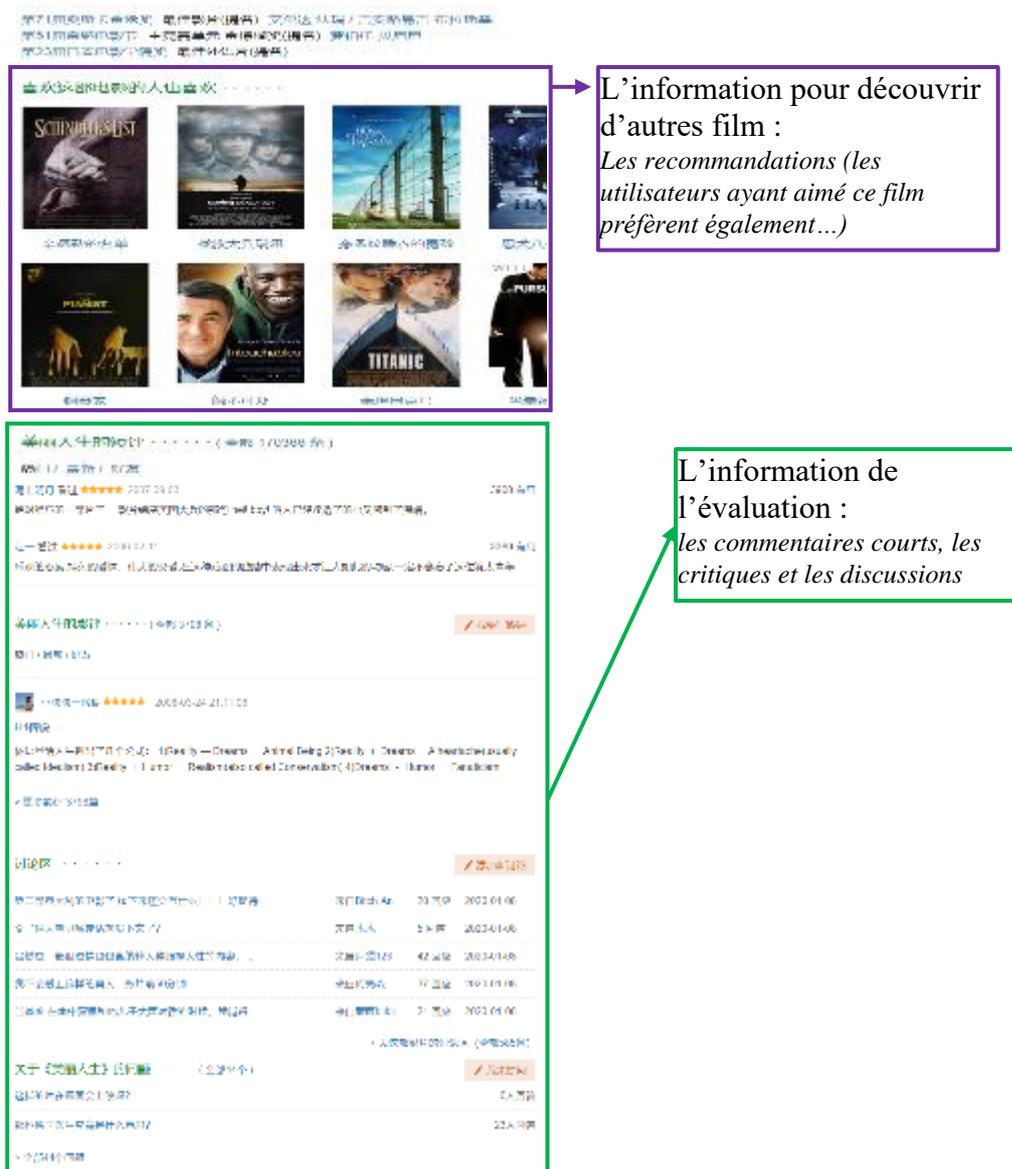


Figure 28 Page d'un film particulier de la version desktop (suite)

En résumé, Douban Film est avant tout une plateforme de RSN pour les cinéphiles. Elle

aide l'utilisateur à découvrir les films qui l'intéressent en fournissant des informations sur le film ainsi qu'à faciliter la rencontre avec d'autres utilisateurs qui partageraient les mêmes préférences avec lui. En plus d'être un RSN centré sur le film, Douban Film fonctionne comme une plateforme d'avis sur les films. En consultant la note moyenne et en feuilletant des critiques du film, l'utilisateur pourrait mieux juger si un film répond à ses préférences.

VII.1.2 Un RSN pour les cinéphiles français : Allociné

Allociné est un service français fournissant des informations cinématographiques en ligne. Fondé en 1992, il était initialement, un service d'informations téléphoniques sur les programmes de cinéma. Depuis le lancement de son site web en 1997, l'entreprise se diversifie ensuite pour s'imposer comme le portail web de référence en France dans ce domaine. Allociné se concentre principalement sur la promotion des nouveautés en salles et en DVD et sur l'information concernant la programmation. Vingt-huit ans après sa fondation, Allociné est plus qu'un portail web : il est devenu également une plateforme du Web 2.0. Allociné propose à l'internaute de s'inscrire sur le site et de créer son propre profil public (Figure 29).

En réalité, Allociné permet à l'utilisateur non seulement de construire un profil mais également d'abonner des profils d'autrui (et/ou d'être abonné), c'est-à-dire qu'il possède une liste d'autres utilisateurs avec lesquels ils partagent une connexion. La liste de ses abonnements est complètement publique ainsi que la liste de ses abonnés (Figure 30). À travers de ces listes, on peut consulter des profils des autres et éventuellement établir lien avec eux. Cela rend Allociné conforme parfaitement à la définition de RSN proposée par Boyd et Ellison (Boyd et Ellison, 2007). En outre, Allociné ouvre sa plateforme éditoriale aux contributions des internautes sur la rubrique sur les films dès janvier 2009. Un ensemble de fonctionnalités permet aux utilisateurs de participer directement au contenu éditorial en créant des nouvelles fiches films par exemple. Nous pouvons ainsi caractériser Allociné comme une RSN typique.

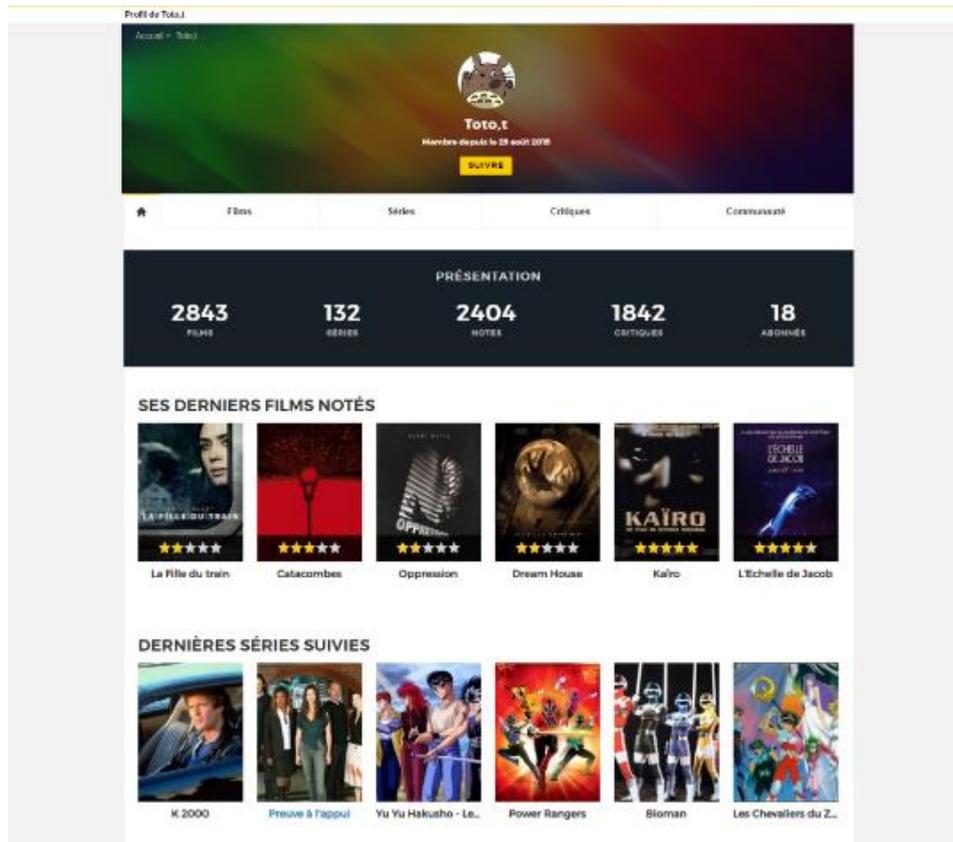


Figure 29 Profil de l'utilisateur sur Allociné

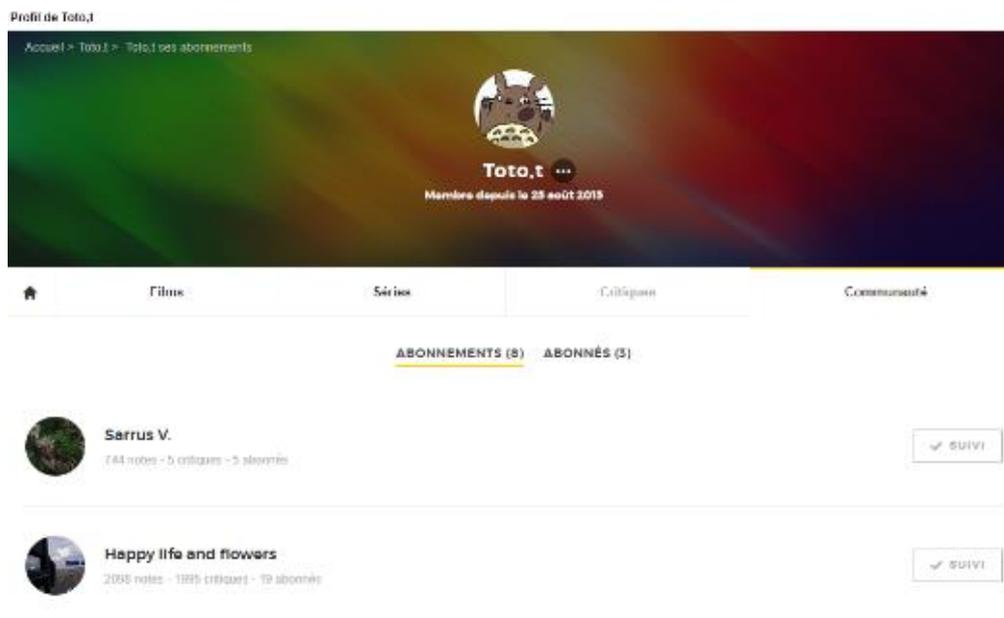


Figure 30 Liste d'abonnements et d'abonnés de l'utilisateur

Allociné offre également une base de données couvrant tous les films qui ont été diffusés en France. Les données comprennent de nombreuses évaluations du film. Comme pour Douban, chaque film a sa page de profil où 4 types d'information sont proposés (Figure 31 et Figure 32):

- **L'information sur le film** telle que la date de sortie, le réalisateur, les acteurs/actrices, synopsis, bande-annonce, les photos, des interviews et des nouvelles concernant du film *etc.*;
- **L'information pratique pour regarder le film** : les liens vers des site de cinéma, les liens vers des sites de VOD (e.g., Netflix, Orange) et les liens vers des sites de vente pour acheter DVD (e.g. Fnac, Amazon)
- **L'information pour découvrir d'autres films** : des recommandations (les films similaires) et des tags ajoutés au film;
- **L'information d'évaluations** : la notation, les critiques, les commentaires courts.

L'information du film

L'information pratique pour regarder le film

L'information de l'évaluation

L'information pour découvrir d'autres films

Figure 31 Page profil d'un film sur Allociné

La page d'un film dans l'application mobile d'Allociné dispose une structure pareille que celle de la version desktop. Une seule nuance est que la version mobile ne propose pas d'information pour découvrir d'autres films.

The image shows a screenshot of a movie profile page on Allociné. It is divided into three main sections:

- FILMS RECOMMANDÉS:** A row of four movie posters: 'La Vie est belle', 'Le Dictateur', 'Million Dollar Baby', and 'Edward aux mains...'. A purple box with an arrow points to this section, containing the text 'l'information pour découvrir d'autres films'.
- SI VOUS AIMEZ CE FILM, VOUS POURRIEZ AIMER ...:** A row of four movie posters: 'La Liste de Schindler...', 'Le Garçon au pyjama...', 'Le Pianiste', and 'Au revoir les enfants...'. Below this is a link 'Voir plus de films similaires' and a list of 'Pour découvrir d'autres films'.
- COMMENTAIRES:** A section with 116 comments. It includes a search bar, a 'Recommander' button, and social media sharing options. Three comments are visible, each with a user profile picture and text. A blue box with an arrow points to this section, containing the text 'L'information de l'évaluation'.

Figure 32 Page profil d'un film sur Allociné (suite)

VII.1.3 Un autre RSN pour les cinéphiles français : SensCritique

SensCritique est un autre RSN francophone culturel qui propose de découvrir, de noter et d'écrire des critiques d'œuvres culturelles appartenant aux domaines des films, séries télévisées, livres, bandes dessinées, albums et morceaux musicaux, et jeux vidéo. Depuis sa fondation en 2011, le site connaît un succès grandissant et propose régulièrement des nouveautés. Clairement positionné pour favoriser la bouche à oreille culturel, SensCritique évolue avec son temps.

Tout comme Douban et Allociné, SensCritique est avant tout un RSN qui permet à l'utilisateur de créer et de gérer son profil public (Figure 11). Son profil peut être rendu publique

à tout le monde ou seulement aux membres inscrits sur le site selon le choix de l'utilisateur. L'utilisateur peut également abonner d'autres utilisateurs et être abonné par autrui. Bien entendu, la liste d'abonnements et des abonnés peut être consultée par les autres membres. À part cela, le profil montre aussi les informations de ses centres d'intérêt : les films, les livres, les musiques et les jeux préférés ainsi que toutes ses activités sur le site.

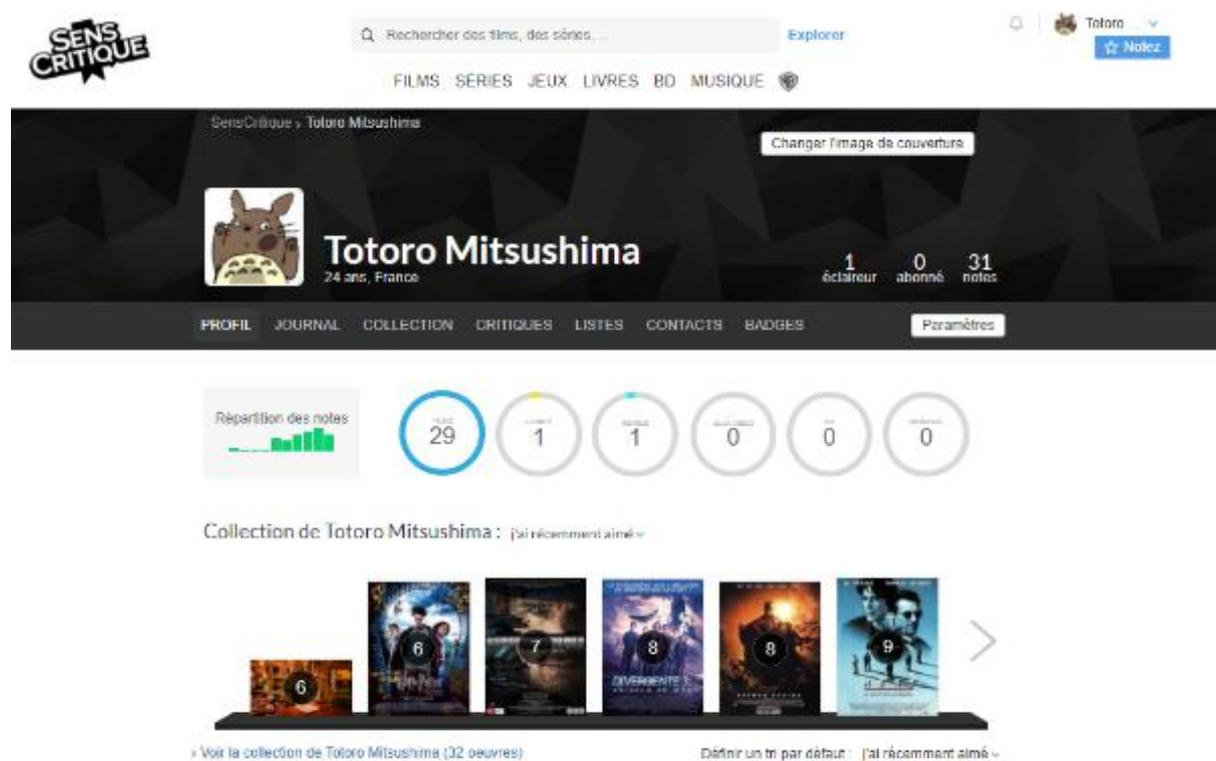


Figure 33 Profil de l'utilisateur sur SensCritique

Ayant l'objectif de permettre aux utilisateurs de découvrir des films, des séries, des livres, des bandes dessinées (BD), des jeux vidéo et de la musique, il offre également une base de données de produits culturels. Sur SensCritique, chaque film a sa page de profil où 4 types d'information sont proposés (Figure 31 et Figure 32):

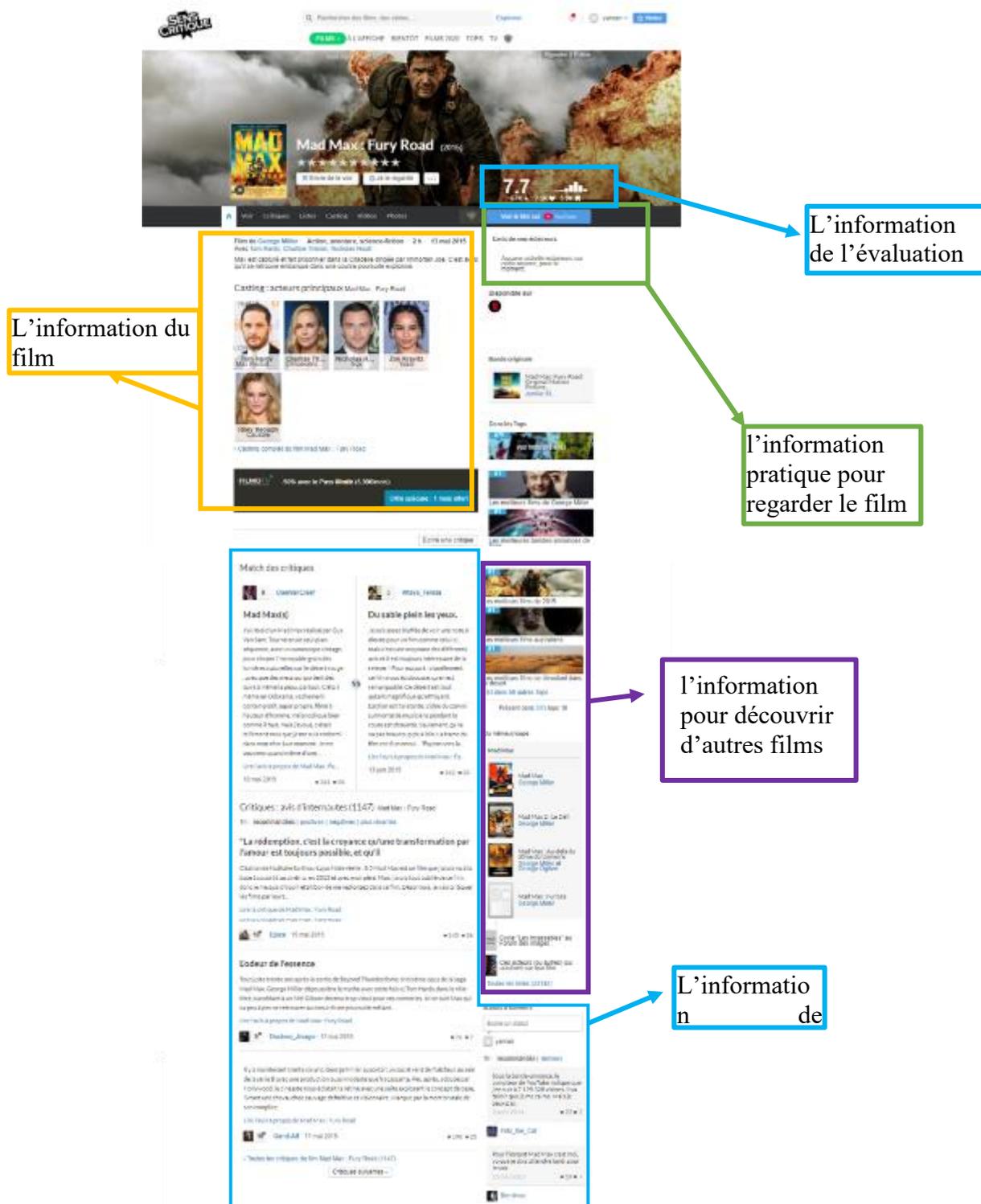


Figure 34 Page profil d'un film sur SensCritique

- **L'information sur le film** telle que la date de sortie, le réalisateur, les acteurs/actrices, synopsis, bande-annonce, les photos, des interviews et des nouvelles concernant du film *etc.*;

- **l'information pratique pour regarder le film** : les liens vers des site de cinéma, les liens vers des sites de VOD (e.g., Netflix, Orange) et les liens vers des sites de vente pour acheter DVD (e.g. Fnac, Amazon)
- **l'information pour découvrir d'autres films** : des recommandations (les films similaires) et des tags ajoutés au film;
- **l'information d'évaluations** : la notation, les critiques, les commentaires courts.

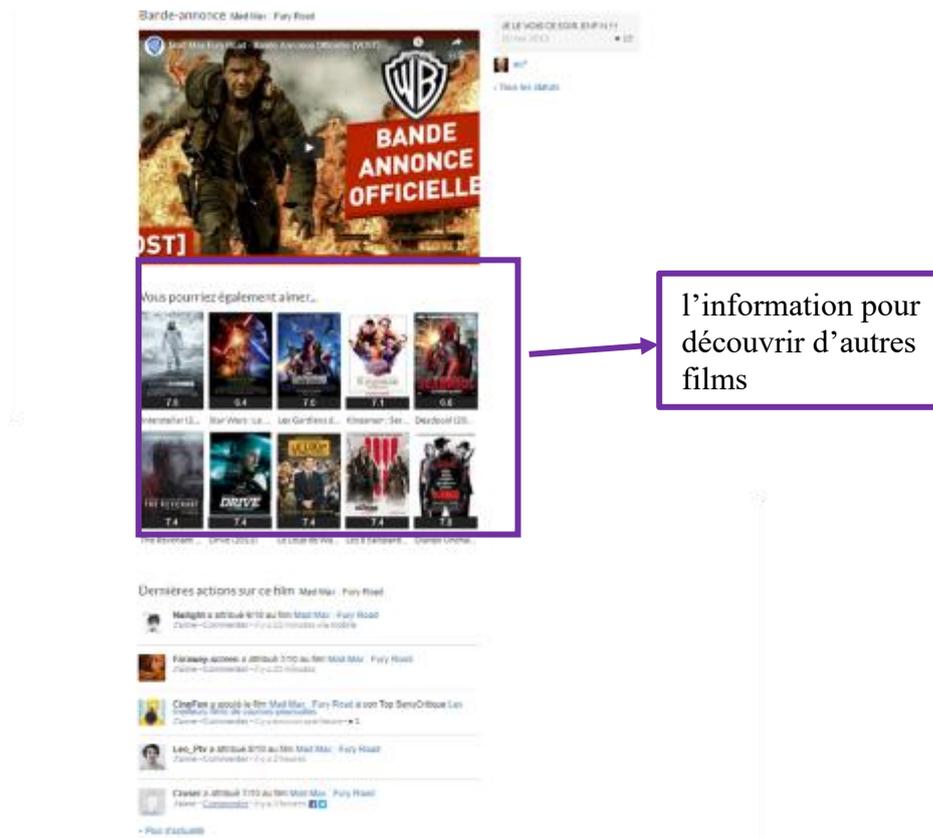


Figure 35 Page profil d'un film sur SensCritique (suite)

VII.1.4 Une plateforme non-RSN pour les cinéphiles chinois

Lancé en 2012, Maoyan Film est un site web de billetterie de films en ligne en Chine. Il se concentre principalement sur la vente de billets de cinéma et la promotion concernant la programmation. Maoyan offre une base de données couvrant tous les films qui ont été diffusés en Chine, de manière similaire à Allociné. Sa page d'accueil offre principalement des informations sur les films actuellement en salle, les films qui vont sortir, le box-office en Chine, les films les plus attendus (marqués par les plus d'utilisateurs comme « envie de regarder ») et les films les mieux notés (Figure 37 Page d'accueil de Maoyan).

The screenshot displays the Maoyan website interface with several key sections:

- ACTUELLEMENT EN SALLES**: A grid of 8 movie posters currently in theaters.
- BOX OFFICE CHINE**: A list of movies with their respective box office earnings in China.
- SORTIES BIENTOT (95)**: A grid of 8 movie posters for upcoming releases.
- FILMS LES PLUS ATTENDUS**: A section highlighting the most anticipated films, featuring movie posters and titles.
- MEILLEURS**: A section for top-rated or best movies, showing movie posters.
- TOP 100**: A section for the top 100 movies, including a poster for 'The Great Wall'.

Movie Title (Chinese)	Box Office (RMB)
《战狼2》	215.78亿
《流浪地球》	199.83亿
《哪吒之魔童降世》	159.55亿
《你好，李焕英》	75.82亿
《唐人街探案3》	85.41亿
《刺杀小说家》	43.75亿
《误杀》	42.51亿
《速度与激情9》	34.48亿
《长津湖》	56.99亿

Figure 37 Page d'accueil de Maoyan



Figure 38 Page d'un film sur Maoyan

Chaque film possède une page de profil dont l'organisation est pareille que celle de Douban Film et d'Allociné. Dans une page de film (Figure 38), 4 types d'information sont mises à disposition :

- **L'information sur le film** telle que la date de sortie, la durée du film, le réalisateur, les acteurs/ actrices, le synopsis, la bande-annonce, des photos, des actualités *etc.*;
- **L'information pour découvrir d'autres films** : des films connexes;
- **L'information d'évaluations** : la notation et les commentaires courts
- **L'information pratique** : le lien pour acheter des billets.

VII.3 SR dans les plateformes pour les cinéphiles

L'application des SR au domaine du cinéma intéresse beaucoup les chercheurs depuis de nombreuses années (Alspector, Koicz et Karunanithi, 1997). Au cours des 20 dernières années, les SR ont été largement appliqués dans les plateformes et les RSN pour les cinéphiles tels que IMDb, Flixster, Rotten Tomatoes, Allociné, SensCritique et Douban. Le SR cherche à prédire automatiquement les préférences de l'utilisateur pour les films et lui fournit des recommandations. Les évaluations des films faites par l'utilisateur dans le passé sont des données primordiales pour que le SR puisse apprendre les préférences de l'utilisateur. Pourtant, dans la plupart des plateformes et des RSN pour les cinéphiles, les utilisateurs semblent avoir tendance à évaluer des films en fonction de la qualité de films au lieu de leurs préférences. Cela implique que les évaluations des utilisateurs ne reflètent pas toujours leurs préférences et que des recommandations risquent d'être imprécises. Dans cette section, nous présenterons les SR dans Douban, Allociné et SensCritique pour illustrer ce problème réel dans l'application de SR aux plateformes pour les cinéphiles.

VII.3.1 Les recommandations fournies dans Douban et Allociné

En tant que RSN pour les cinéphiles, Douban Film et Allociné non seulement fournissent de l'information sur le cinéma et facilitent la rencontre des cinéphiles mais ils cherchent également à aider l'utilisateur à découvrir des films en lui fournissant des recommandations : « Des interfaces de bases de données filmographiques sont mises à disposition des utilisateurs pour une appréciation. Cette méthode, utilisée dans MovieLens, est exactement celle présentée dans la partie d'analyse du filtrage collaboratif (Schafer *et al.*, 2007). À partir des votes de chaque utilisateur, il est possible de lui fournir des recommandations » (Saleh, Kembellec et Chartron, 2014).

Sur Douban Film et Allociné, chaque film dispose d'une page dédiée où l'utilisateur peut trouver des recommandations proposées par le système. Plus précisément, Allociné propose des films (Figure 39) et une liste de films (Figure 40) similaires à celui-ci en proposant des

suggestions comme « si vous aimez ce film, vous pourriez aimer... » (Figure 39).



Figure 39 Les recommandations fournies par Allociné

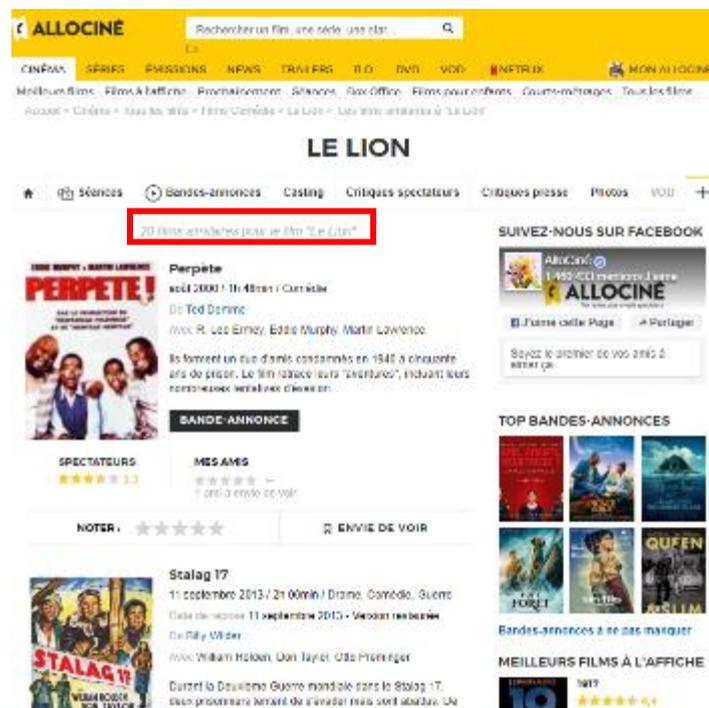


Figure 40 Liste des films similaires pour un film proposée par Allociné

Douban Film fournit l'utilisateur de recommandation des films qui sont « appréciés par

des utilisateurs qui aiment ce film » (Figure 41). D'ailleurs, dans l'application mobile de Douban Film, des recommandations plus personnalisées sont proposées avec les raisons de a recommandation : des films similaires à ceux que l'utilisateur « a aimés » dans le passé, des films appréciés par des utilisateurs (ses amis, ses abonnements et abonnés) avec qui l'utilisateur partage un lien sur Douban (Figure 42).

小妇人的获奖情况 ····· (全部)

第92届奥斯卡金像奖 最佳影片(提名) 艾米·帕斯卡尔
 第77届金球奖 电影类 剧情片最佳女主角(提名) 西尔莎·罗南
 第73届英国电影学院奖 电影类 最佳女主角(提名) 西尔莎·罗南

Les utilisateurs qui aiment ce film apprécient également...

婚姻故事	伯德小姐	傲慢与偏见	纽约的一个雨天	布鲁克林
过春天	爱在黎明破晓前	82年生的金智英	请以你的名字呼唤我	怦然心动

Figure 41 Les recommandations fournies par Douban

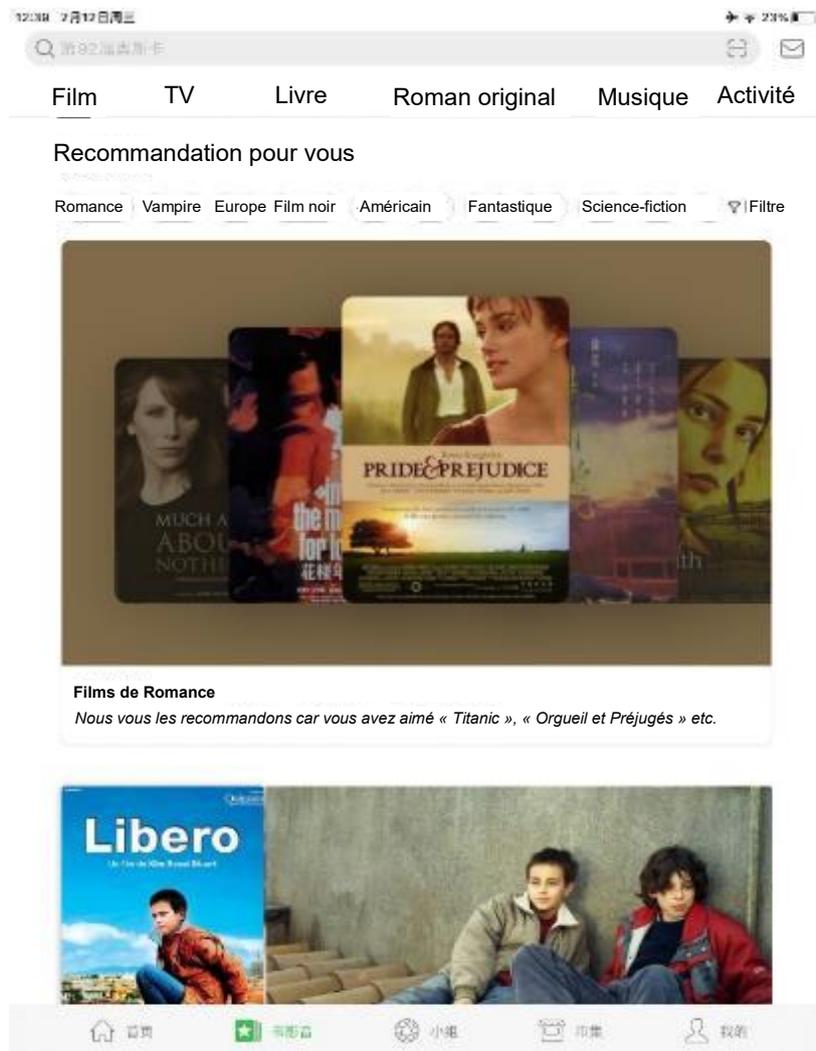


Figure 42 Les recommandations personnalisées dans l'application mobile de Douban Film

VII.3.2 Les deux critères de notation du film

La méthode de FC est principalement utilisée par les SR de Douban Film et d'Allociné. Ils fournissent les recommandations reposant sur les préférences de l'utilisateur comme nous avons présenté dans la section précédente. Les notations de films attribuées par l'utilisateur sont la donnée de base pour inférer ses préférences. Cependant, nous remarquons *l'ambiguïté* de l'interprétation de la notation sur ces deux plateformes. Tandis que Douban Film et Allociné infèrent les préférences de l'utilisateur en fonction de ses notations du film dans le passé, il apparaît qu'ils encouragent l'utilisateur à noter des films en fonction de ce qu'il pense être la

qualité du film au lieu de le faire en fonction de ses propres préférences (ce qu'il aime regarder comme film).

En fait, les plateformes pour les cinéphiles tels que IMDb, SensCritique, Allociné et Douban visent principalement à aider les utilisateurs à inférer la qualité du film (Koh, Hu et Clemons, 2010). Ils permettent aux utilisateurs d'évaluer des films et de consulter les évaluations d'autrui. Cela signifie que d'évaluer le film en fonction de la qualité du film pourrait être une norme très importante pour les membres de ces plateformes.

Dans le but d'aider les utilisateurs à sélectionner les films, les 4 plateformes que nous avons choisies permettent aux utilisateurs d'évaluer les films en les notant et en rédigeant des critiques ou des commentaires courts. L'échelle de notation de 5 étoiles allant de 0,5 étoile à 5 étoiles est axée sur la qualité du film (Tableau 3). SensCritique ne dispose pas une explication pour l'échelle de notation, mais il cherche également à être plus objectif dans les notations comme ce qu'indique une des leurs articles : « une de nos ambitions était de tendre vers une relative objectivité de l'appréciation d'une œuvre... Globalement, on a réussi à créer une échelle de valeur qui tend vers cette objectivité. »¹⁴³. En fait, les utilisateurs de ces 4 plateformes partagent une norme de groupe en commun qui est « de noter des films en fonction de la qualité ».

Etoile \ Plateforme	Douban	Allociné	Maoyan
0,5		Nul	Nul
1	Très mauvais	Très mauvais	Nul
1,5		Mauvais	Mauvais
2	Mauvais	Pas terrible	Mauvais
2,5		Moyen	Pas mal
3	Pas mal	Pas mal	Pas mal
3,5		Bien	Bien
4	Recom mandé	Très bien	Bien
4,5		Excellent	Parfait
5	Très recommandé	Chef-œuvre	Parfait

Tableau 3 Les explication de l'échelle de notation sur 3 plateformes

¹⁴³ <https://medium.com/@SensCritique.com/des-nouvelles-de-la-note-senscritique-6d4d10699139>

On peut donc dire que la notation sur ces plateformes indique plutôt la qualité du film. Tandis que l'échelle de notation de MovieLens mesure la préférence de l'utilisateur pour un film (« 1 étoile » est interprétée comme « Je n'aime pas du tout ce film » et « 5 étoiles » comme « J'adore vraiment ce film »), l'échelle de notation d'Allociné indique la qualité du film (« 0,5 étoile » indique qu'un film est « nul » et « 5 étoiles » signifient que ce film est un « chef-d'œuvre »). L'échelle de notation de Douban est également axée sur la qualité du film (Tableau 4). Ces deux interprétations de l'échelle de notation impliquent qu'il y a deux différents critères pour évaluer du film.

Critère Etoile	Critère objectif <i>(Douban et Allociné)</i>	Critère subjectif <i>(MovieLens)</i>
1	Très mauvais	Je déteste
2	Mauvais	Je n'aime pas
3	Pas mal	Neutre
4	Recommandé	J'aime
5	Très recommandé	J'aime beaucoup

Tableau 4 Deux explications de l'échelle de notation selon le critère objectif et le critère subjectif

En outre, l'explication de l'échelle de notation est présentée à l'utilisateur chaque fois qu'il note un film sur Allociné et Douban (Figure 43 et Figure 44). L'idée d'évaluer des films en fonction de leur qualité est constamment renforcée. Il semble que les utilisateurs ont tendance à noter des films d'une manière plus objective (selon la qualité du film plutôt selon leurs préférences).

8 janvier 2020 / 1h 38min / Comédie
De Jézabel Marques
Avec Chantal Lauby, Camille Chamoux, Giovanni Pucci
Nationalité Français

BANDE-ANNONCE **SÉANCES (317)**

PRESSE ★★★★★ 3,1
10 critiques

SPECTATEURS ★★★★★ 3,5
10 critiques dont 35 critiques

MES AMIS ★★★★★ --

NOTER: ★★★★★ **5 Chef-d'oeuvre**

ENVIE DE VOIR RÉDIGER MA CRITIQUE ...

Figure 43 Sur Allociné, l'explication de notation se présente lorsque l'utilisateur note un film

理查德·朱维尔的哀歌 Richard Jewell (2019)



更新描述或海报

导演: 克林特·伊斯特伍德
编剧: 玛丽·布伦纳 / 比利·雷
主演: 保罗·沃尔特·豪泽 / 山姆·洛克威尔 / 奥利维亚·王尔德 / 乔恩·哈姆 / 凯西·贝茨 / 更多...

类型: 剧情
制片国家/地区: 美国
语言: 英语
上映日期: 2020-01-10(中国大陆) / 2019-11-20(AFI Fest) / 2019-12-13(美国)
片长: 129分钟
又名: 李察朱维尔: 惊世疑案(港) / 李察朱威尔事件(台) / 美国噩梦 / 理查德·朱厄尔 / 理查德·朱厄尔的悲歌 / 理查德·朱厄尔的歌谣 / 理查德·朱维尔的歌谣 / 理查德·杰威尔的歌谣 / The Ballad of Richard Jewell / American Nightmare / American Nightmare: The Ballad of Richard Jewell
IMDb链接: tt3513548

豆瓣评分

8.3 ★★★★★
11486人评价

5星 27.7%
4星 58.0%
3星 13.4%
2星 0.7%
1星 0.1%

好友评分 ?

10.0 1人评价

好于 83% 剧情片

想看 看过 评价: ★☆☆☆☆很差 → **Très mauvais**

写短评 写影评 + 添加到豆列 分享到 推荐

Figure 44 Sur Douban, l'explication de notation se présente lorsque l'utilisateur note un film

Ainsi nous constatons que l'utilisateur dispose de deux critères différents pour noter des films : soit il note des films en fonction de ses préférences, soit il note des films en fonction de la qualité du film. Cette coexistence de deux critères cause une ambiguïté dans l'interprétation de la notation pour peu qu'ils ne soient pas convergents : l'utilisateur pourrait ne pas aimer des films qu'il juge de bonne qualité, et vice-versa. Il est nécessaire de vérifier si ces deux critères s'accordent toujours. Au cas où ils ne le seraient pas, il faut examiner quel critère est employé par l'utilisateur lorsqu'il note des films. En outre, l'influence sociale joue un rôle important dans les RSN comme Douban Film et Allociné. Les opinions d'autrui et les normes du groupe ont la possibilité d'influencer fortement l'utilisateur quand il note des films.

VII.3 Les trois RSN et un non-RSN pour les cinéphiles

Comme nous l'avons indiqué dans les sections précédentes, les 4 plateformes ont pour objectif d'aider les utilisateurs à inférer la qualité du film. Cela est l'une des propriétés principales de ces 4 plateformes. Nous montrons qu'ils encouragent leurs utilisateurs à évaluer des films en fonction de la qualité du film. Cela signifie que noter du film en fonction de la qualité est l'une des normes principales pour leurs utilisateurs.

Les études montrent que les normes du groupe pourraient être renforcées par les interactions entre les utilisateurs. (T. Postmes, Spears et Lea, 2000; Postmes *et al.*, 2001b; Hogg et Reid, 2006). L'interaction entre les utilisateurs développe un sens d'identité de groupe significatif et fort (Finholt et Sproull, 1990; Bouas et Arrow, 1995; Postmes, Spears et Lea, 1999; Lea, Spears et de Groot, 2001). Ce sens d'identité de groupe est lié aux normes qui définissent la façon dont les membres du groupe doivent penser, ressentir ou se comporter (Postmes, Spears et Lea, 2000). Les normes du groupe sont déduites des propriétés prototypiques du groupe. Le prototype informe un membre du groupe des comportements typiques et, par conséquent, appropriés, souhaitables ou attendus dans le groupe (Turner, 1982).

Selon les études de la désindividuation, les normes du groupe seraient renforcées sur les 3 RSN (Douban, Allociné et SensCritique) qui facilitent les interactions entre leurs utilisateurs. Nous avons constaté que les utilisateurs de ces 3 RSN peuvent communiquer et interagir : 1) en

envoyant des messages par messagerie privée, 2) en postant dans des forums ou des groupes, et 3) en exprimant leur avis sur des critiques de films rédigées par d'autres membres.

Douban dispose de sa propre messagerie privée qui permet à l'utilisateur de communiquer directement avec les autres (Figure 45). L'utilisateur peut aussi créer des groupes dans la section « Douban Groupe » autour d'un sujet spécifique. En ce qui concerne le film, il existe de nombreux groupes différents dont le nombre de membres est considérable. Certains films populaires ont leur propre groupe dans la section « Douban Groupe ». D'autre part, l'utilisateur peut laisser des commentaires sur les critiques d'autrui. Il y a la zone de discussion dans la page du film (Figure 47). Dans la page d'un film, les utilisateurs peuvent également exprimer leur opinion à propos des critiques de films rédigées par les autres membres en laissant des commentaires et en votant « utile » ou « inutile » (Figure 46).



Figure 45 La messagerie privée de Douban

Quant à Allociné, il permet à l'utilisateur de laisser des commentaires sur des critiques d'autrui (Figure 49). L'utilisateur peuvent également se communiquer avec les autres via la messagerie privée et dans le forum d'Allociné (Figure 48).



Figure 48 les forums et la messagerie privée d'Allociné



Figure 49 On peut laisser des commentaires sur des critiques d'autrui

Sur SensCritique, pareillement, les membres peuvent échanger des opinions par la messagerie (Figure 50). Bien entendu, les utilisateurs peuvent également exprimer leur opinion sur les critiques d'autrui (Figure 51).

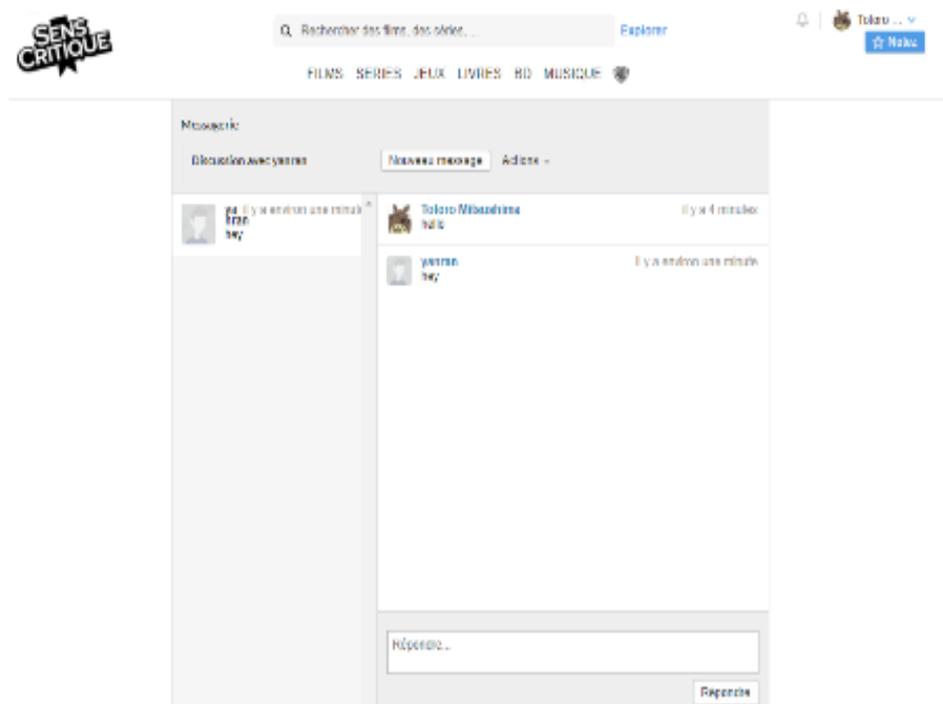


Figure 50 La messagerie sur SensCritique

Outre l'interaction entre les membres, l'anonymat améliore également le respect des normes du groupe par les individus lorsqu'ils ont une identité du groupe forte ou une identité personnelle faible selon une série d'études de Spears, Postmes et Lea (Spears, Lea et Lee, 1990; Lea et Spears, 1991; Spears et Lea, 1994; Reicher, Spears et Postmes, 1995; Postmes et Spears, 1998; Postmes, Spears et Lea, 1999; Tom Postmes, Spears et Lea, 2000). Les individus se comportent de manière plus conforme aux les normes du groupe dans un CMC anonyme. Étant donné que Douban, Allociné et SensCritique sont des RSN anonymes où les utilisateurs emploient des pseudonymes au lieu de leur nom et prénom réel comme sur Facebook, cela signifie que les utilisateurs de ces trois plateformes se conformeraient davantage à certaines normes prototypiques telles que de noter des films en fonction de la qualité.



Figure 51 Une critique de film sur SensCritique

Il est nécessaire de les comparer avec une autre plateforme qui n'est pas un RSN et où l'influence normative est moins puissante. De ce fait, nous introduisons la plateforme « Maoyan ».



Figure 52 Profil de l'utilisateur de Maoyan

Bien que Maoyan repose sur la même norme prototypique que Douban Film, Allociné et SensCritique, cette norme n'aurait pas la même influence faute d'interactions suffisantes. Effectivement, Maoyan ne dispose pas de système de messagerie privée et n'est pas un RSN comme Douban Film et Allociné. Quoique l'utilisateur puisse créer et gérer son propre profil, il ne peut pas abonner les autres ou ajouter des amis sur Maoyan. Les utilisateurs ont le droit de voter pour ou contre une critique mais ils ne peuvent pas laisser de commentaire sur les critiques d'autrui (Figure 53). Il n'existe pas de liste d'autres utilisateurs avec qui l'un d'entre eux partage un lien et l'utilisateur ne peut pas consulter la liste des autres non plus. En conséquence, nous ne pouvons pas qualifier Maoyan de RSN conformément à la définition de RSN (Boyd et Ellison, 2007).

热评

5.0



xiao 小可
00-21 ★★★★★

1852

丁晓岚说：青少年都应该看看一部这样的影片，增强我们的爱国主义观念，民族凝聚力。我们青年比父老，更奋进！！



Boogqlqia
08-21 ★★★★★

1615

我国革命从军可当悔不已，她感恩祖国革命先烈，我生在社会主义革命发源地。



晓雷雷
19小时前 ★★★★★

254

一部热血澎湃的红色影片，让亿万观众热血沸腾且动容的泪眼！保护党的机密，刘胡兰英勇不屈的牺牲精神在上海外滩一决死地，女英雄守军人数400人左右，对外宣布300勇士，震撼人心，牢记历史，勿忘国耻！致敬英雄！西女不知亡国恨，她在硝烟中绽放，一江才气，千秋，大节，战火纷飞，在人们心中铸就的国魂，铭记上海，铭记革命先辈们！！我们的旗帜，永远飘扬在富士山头！长死了，作你英——英雄儿女！大力神在硝烟中绽放，鲜血，再一次染红了大地，因为我们的无产英雄……14年抗战期间，中国军民伤亡人数3500万人以上，财产损失达1000亿美元！中国人，没有一个人有背叛祖先人愿被日本人！

Figure 53 L'espace de critique sur la page d'un film sur Maoyan

Il semble que les utilisateurs de Maoyan ne développeraient pas un sens d'identité du groupe fort et que l'influence normative entraînerait des impacts très limités sur ses utilisateurs. Nous considérons donc que les notations des films sur Maoyan présentent une différence relativement importante par rapport aux trois autres plateformes.

VII. 4 Allociné et Douban : 2 RSN pour les cinéphiles issus de 2 cultures différents

À travers notre observation présentée dans les sections précédentes, nous argumentons que Douban Film et Allociné sont deux RNS très similaires à plusieurs égards.

Douban Film et Allociné sont tous les deux des plateformes centrées sur le film. Ils disposent des bases de données assez complètes et détaillées. Les données comprennent des informations de films, des informations pour regarder des films et des évaluations de films. Les utilisateurs sont permis d'y évaluer des films en notant des films et en rédigeant leurs critiques et commentaires. Les utilisateurs profitent de ces évaluations qui leur aident à faire décision avant d'aller au cinéma.

Douban Film et Allociné disposent aussi le SR qui propose aux utilisateurs des films qui les intéresseraient. Douban Film et Allociné sont également des plateformes du Web 2.0. Ils hébergent des blogs de leurs utilisateurs ainsi que de nombreux contenus générés par les utilisateurs : les notations, les critiques et les commentaires, et éventuellement des nouvelles fiches films. La folksonomie est également employée en permettant aux utilisateurs d'ajouter les tags aux films. Ils fournissent des flux RSS que les utilisateurs peuvent abonner. En résumé, ces deux plateformes permettent aux utilisateurs de publier et d'exprimer leurs opinions. Elles facilitent la recherche et la collection d'informations aux utilisateurs. Elles sont des médias sociaux où les utilisateurs peuvent partager et se communiquer avec les autres.

Douban Film et Allociné sont des RSN typiques qui favorisent la rencontre des utilisateurs. Du fait qu'ils conforment parfaitement à la définition de RSN proposée par Boyd et Ellison (Boyd et Ellison, 2007). Ils permettent, tous les deux, à l'utilisateur de construire son propre profil public, de posséder une liste d'autres utilisateurs avec qui il partage un lien, de consulter les profils des autres à travers la liste d'autrui et éventuellement d'établir des connexions avec eux.

Nous distinguons également certaines différences entre ces deux plateformes :

1. Douban fournit seulement des évaluations des spectateurs (ses utilisateurs) alors qu'Allociné fournit aussi des évaluations des presses principales à part des

opinions des spectateurs.

2. Les deux plateformes disposent, tous les deux, leur application mobile. Douban propose davantage de recommandations dans l'application mobile que sur son site desktop. Allociné, au contraire, n'offre des guère de recommandation dans son application mobile.

Malgré ces deux nuances, Douban Film et Allociné sont quasi équivalents. La similitude entre Douban Film et Allociné signifie que les utilisateurs de ces deux plateformes seraient comparables ainsi que leurs comportements. Cela nous permet à examiner et à vérifier si les comportements de notations sont variés dans les cultures différentes.

PARTIE 4 SYNTHÈSE DES RESULTATS, CONCLUSION ET PERSPECTIVES

Nous menons deux études pour cette recherche. Le **Chapitre VIII** Enquête par questionnaire auprès des utilisateurs de Douban présente les résultats de notre première étude. Dans cette étude, nous avons mené une enquête par questionnaire auprès des utilisateurs de Douban dans le but de comprendre leurs habitudes de notation et de découvrir les facteurs qui pourraient influencer leur comportement de notation. Dans cette étude, nous cherchons à vérifier si les utilisateurs perçoivent deux critères différents pour noter les films (le critère objectif et le critère subjectif) et si évaluer les films en fonction du critère objectif est l'une des normes explicites du groupe sur les plateformes pour les cinéphiles. Outre les normes explicites, nous essayons également de vérifier s'il existe des normes implicites (par exemple, les opinions d'autrui) qui peuvent influencer le comportement de notation des utilisateurs.

La première étude nous permet de découvrir les facteurs qui influenceraient les comportements de notation des utilisateurs. Afin de vérifier leurs déclarations, nous menons une deuxième étude comparative dont les résultats sont présentés dans le **Chapitre IX** Études comparatives. Dans cette deuxième étude, nous essayons d'étudier l'influence des normes du groupe sur le comportement de notation des utilisateurs en comparant les notations d'un même film sur différentes plateformes, par exemple, les notations sur les plateformes de RSN ayant les mêmes normes de groupe montrent une corrélation plus élevée, leurs utilisateurs ont tendance à donner des notations plus élevées à certains films, etc. Dans cette étude, nous faisons trois comparaisons. La première comparaison est réalisée entre Allociné et SensCritique, deux RSN français qui sont issus de la même culture et qui encouragent, tous les deux, leurs utilisateurs à évaluer les films selon le critère objectif. Le résultat sert à faire comparer avec les résultats des deux études comparatives suivantes. La deuxième comparaison est celle entre Douban et Maoyan, qui est une comparaison entre un RSN et un non RSN chinois qui sont issus de la même culture et qui encouragent, tous les deux, leurs utilisateurs à évaluer les films selon le critère objectif. Étant donné que l'influence normative est moins puissante sur la plateforme non RSN, la corrélation entre les notations de Douban et de Maoyan serait moins élevée que la corrélation entre les notations d'Allociné et de SensCritique. Ces deux comparaisons nous permettent de vérifier l'influence normative sur le comportement de notation des utilisateurs. Nous effectuons la troisième comparaison entre Douban et Allociné, deux RSN qui sont issus

de deux cultures différentes (chinois et français) et qui encouragent, tous les deux, leurs utilisateurs à évaluer les films selon le critère objectif. Cette comparaison nous permet de vérifier si la différence culturelle serait un facteur qui influence le comportement de notation des utilisateurs.

Dans le **Chapitre X** Interprétation des résultats, nous présentons l'analyse de nos résultats, la discussion et les solutions proposées.

Chapitre VIII Enquête par questionnaire auprès des utilisateurs de Douban

Notre objectif principal consiste à examiner si la notation du film reflète les préférences réelles des utilisateurs dans les RSN pour les cinéphiles. Nous considérons trois facteurs principaux qui sont les plus susceptibles d'influencer le comportement de notation : **le critère de notation du film, l'influence de la norme de groupe et la différence culturelle**. La plateforme principale que nous avons choisie pour cette étude est un RSN culturel chinois : Douban Film. Il offre à l'utilisateur des services d'évaluation et de recommandation pour le film. L'utilisateur peut attribuer la note au film sur une échelle de 1 à 5 étoiles (points). Etant un RSN, il permet à l'utilisateur de retrouver des amis qui partagent les mêmes centres d'intérêt et des goûts proches. Il fournit également un service de recommandation qui propose des films susceptibles d'intéresser l'utilisateur. Par conséquent, Douban semble être une plateforme idéale pour notre recherche. Nous avons mené une enquête par questionnaire auprès des utilisateurs de Douban pour examiner leur habitude de notation et découvrir les facteurs qui influencent leur comportement de notation.

VIII.1 Problèmes de recherche et les hypothèses

Selon Belkin et Croft (voir notre Partie II), le FI est l'expression utilisée pour décrire une variété de processus ayant pour but de fournir des informations à des personnes, informations en adéquation avec leurs centres d'intérêt (Belkin et Croft, 1992). Les SFI sont conçus pour faciliter aux utilisateurs l'extraction des informations pertinentes parmi une masse d'informations entrantes. Plus précisément, leurs fonctions sont : 1) d'éliminer les informations impertinentes ; 2) de faire parvenir aux utilisateurs les informations pertinentes.

En tant que type de FI, le SR consiste à faire parvenir à l'utilisateur des informations pertinentes parmi un grand volume d'information entrant (Jannach, 2010). Le SR se repose notamment sur les rétroactions de l'utilisateur et utilise les techniques statistiques afin de

réaliser les prédictions. Les rétroactions collectées sont analysées par le système en employant des algorithmes variés. La collection et l'analyse des rétroactions sont deux tâches essentielles pour le SR. Les rétroactions de l'utilisateur peuvent être divisées en deux parties :

1. Les rétroactions explicites : les évaluations (*ratings*) telles que les notes, les commentaires et les critiques sont des rétroactions explicites que le système collecte. Les informations démographiques sont éventuellement utilisées comme une stratégie pour résoudre les problèmes de démarrage à froid ou par les systèmes dont les recommandations sont basées sur la démographie ;

2. Les rétroactions implicites : les comportements de l'utilisateur tels que les items qu'il a vus, a marqués ou a ajoutés dans sa liste d'envies sur les sites de vente ; la fréquence de consultation d'un item ; les produits qu'il achète en ligne *etc.*

Les rétroactions permettent au SR d'apprendre les préférences de l'utilisateur. Ensuite, le SR est capable de proposer les nouveaux items qui correspondent le mieux aux préférences de l'utilisateur. Il implique que la collection de rétroactions, notamment les rétroactions qui reflètent bien les préférences de l'utilisateur, est indispensable pour le SR. Dans le domaine de recommandation du film, la notation du film attribuée par l'utilisateur est la rétroaction la plus importante pour apprendre ses préférences de film. L'idée de base est que les notes reflètent directement les préférences de l'utilisateur : l'utilisateur attribue les notes élevées aux films qu'il aime et les notes plus faibles aux films qu'il n'aime pas. Le SR souhaite que l'utilisateur note des films en fonction de ses préférences personnelles. Ce sont les notes qui aident à améliorer leurs recommandations prochaines.

Dans la pratique, le SR offre aux utilisateurs la possibilité de noter le film sur une échelle de 0 à 5 (ou 0 à 10) (Aggarwal, 2016 ; Jannach, Zanker, Felfernig, & Friedrich, 2010). Il exige que les utilisateurs notent des films en fonction de leurs préférences. En d'autres termes, ils doivent juger s'ils aiment un film ou pas. Dans ce cas, leurs évaluations semblent plutôt subjectives.

Le SR est largement utilisé dans les RSN pour les cinéphiles tels que IMDb, Allociné, Douban, SensCritique. Cependant, sur ces RSN, les utilisateurs semblent être invités à évaluer les films de manière plus objective. La note d'un film est souvent considérée comme un

indicateur de la qualité du film. Cela nous amène à la première question de notre étude :

Q1 : Est-ce que le film pourrait être évalué par deux critères différents ?

Effectivement, nous constatons que Douban, Allociné et Maoyan offrent une autre interprétation de l'échelle de notation. Au lieu de révéler les préférences de l'utilisateur, la notation du film semble plutôt indiquer la qualité du film. Cela implique qu'il existe deux critères différents pour évaluer le film. Ainsi, nous formulons notre première hypothèse de recherche :

H1 : Il y aurait deux critères différents pour évaluer le film : le critère subjectif qui se base sur les préférences de l'utilisateur et le critère objectif qui se repose sur la qualité du film.

Nous n'avons pas l'intention de discuter si le film peut complètement être évalué objectivement ou subjectivement en réalité. Ce qui nous intéresse est de savoir si les utilisateurs **perçoivent** l'existence de ces deux critères différents. Évidemment, ces deux critères ne doivent pas être toujours convergents. Ainsi, nous formulons les deux hypothèses suivantes :

H1a : les utilisateurs n'aimeraient pas toujours des films qu'ils jugent d'excellente qualité.

H1b : les utilisateurs aimeraient parfois des films qu'ils jugent de moindre qualité.

Si la coexistence de ces deux critères différents peut être prouvée, cela signifie qu'un film peut recevoir deux notations différentes en fonction de ces deux critères. Alors, nous nous intéressons à savoir si la coexistence de ces deux critères influence les comportements de notation des utilisateurs. En fait, il s'agit la deuxième question de notre recherche :

Q2 : Dans les RSN pour les cinéphiles, est-ce qu'il existe certaines normes de groupe qui influencent le comportement de notation ?

Vu que l'objectif des RSN pour les cinéphiles est d'aider les utilisateurs à déduire la qualité du film des évaluations, il semble que « de noter le film en fonction de la qualité du film » est l'une des normes importantes **pour les utilisateurs**.

Les études montrent que dans les groupes anonymes de CMC, les membres développent un sens d'identité de groupe significatif et font suite aux interactions entre eux (Finholt et

Sproull, 1990; Bouas et Arrow, 1995; Postmes, Spears et Lea, 1999; Lea, Spears et de Groot, 2001). Ce sens d'identité du groupe est lié aux normes qui définissent la façon dont les membres du groupe doivent penser, ressentir ou se comporter (Postmes, Spears et Lea, 2000).

Comme nous avons présenté dans la partie précédente, Douban est un RSN anonyme qui favorise les interactions entre les utilisateurs. Il signifie que l'influence normative pourrait avoir des impacts considérables sur le comportement des utilisateurs. Conséquemment, nous proposons notre deuxième hypothèse :

H2 : La norme explicite du groupe de « noter le film en fonction de la qualité du film » influencerait les utilisateurs de Douban.

Étant donné que la norme du groupe définit la façon dont les membres du groupe doivent penser, ressentir ou se comporter, nous proposons nos troisièmes et quatrièmes hypothèses :

H3 : Pour les utilisateurs, la note du film représenterait la qualité du film.

H3a : Les films d'excellente qualité seraient bien notés sur Douban.

H3b : Les films de moindre qualité seraient mal notés sur Douban.

H4 : les utilisateurs auraient tendance à noter le film en fonction de la qualité perçue (critère objectif) plutôt que selon leurs propres préférences (critère subjectif).

H4a : les utilisateurs auraient tendance à donner une note élevée à un film qu'ils jugent d'excellente qualité même s'ils ne l'aiment pas personnellement.

H4b : les utilisateurs auraient tendance à donner une note faible à un film qu'ils jugent de moindre qualité même s'ils l'aiment personnellement.

Sauf cette norme explicite, certaines normes implicites pourraient exister qui influenceraient l'utilisateur. En fait, de nombreuses d'études se focalisent sur l'impact des évaluations (les notes et les commentaires) sur la décision d'achat de l'utilisateur (Chevalier & Mayzlin, 2006 ; Dellarocas, Zhang, & Awad, 2007; Godes & Mayzlin, 2004). Ces études montrent que la décision de l'utilisateur est influencée de manière significative par les évaluations précédentes, c'est-à-dire que, l'utilisateur peut être influencé avant d'utiliser un produit. Dans cette étude, nous cherchons à savoir si les évaluations précédentes ont de l'influence sur l'utilisateur quand il note un film après l'avoir regardé et surtout quand il n'est

pas d'accord avec les évaluations précédentes. Nous prenons en compte trois facteurs d'influence principaux : la note moyenne, les notes des amis et les notes des utilisateurs du film abonnés par l'utilisateur. Nous proposons donc :

H5 : Les utilisateurs seraient influencés par les autres quand ils ne sont pas d'accord avec l'opinion de la majorité (la note moyenne), avec l'opinion de leurs amis, et avec l'opinion des critiques du film qu'ils suivent.

Tenant compte du biais de sous-déclaration signalé par des recherches précédentes (Anderson, 1998), les utilisateurs sont davantage motivés pour exprimer leurs opinions à travers des critiques et des évaluations sur les produits qu'ils aiment ou ceux qu'ils détestent. Ainsi, nous examinons en particulier l'impact sur les utilisateurs lorsqu'ils ne sont pas d'accord avec les évaluations précédentes de films qu'ils aiment ou ceux qu'ils détestent.

H5a : Les utilisateurs auraient tendance à donner une meilleure note à un film qui est bien noté par les autres, bien qu'ils ne le trouvent pas très bon personnellement.

H5b : Les utilisateurs auraient tendance à donner une note moins bonne à un film qui est mal noté par les autres, bien qu'ils ne le trouvent pas très mauvais personnellement.

VIII.2 Méthode

L'enquête est menée en avril 2017. Au total, 2000 questionnaires sont envoyés et 310 réponses sont reçues. Sur l'ensemble des répondants, 68 % sont des hommes et 32 % des femmes. 1,9 % des utilisateurs ont moins de 18 ans, 85,8 % ont entre 18 et 29 ans et 12,2 % ont plus de 31 ans (*Figure 54*). La démographie de nos répondants est bien représentative de l'ensemble d'utilisateurs de Douban dont les répartitions d'utilisateurs par âge et par sexe sont présentées dans la *Figure 55*¹⁴⁴.

¹⁴⁴ Source de Baidu Index en mai 2017

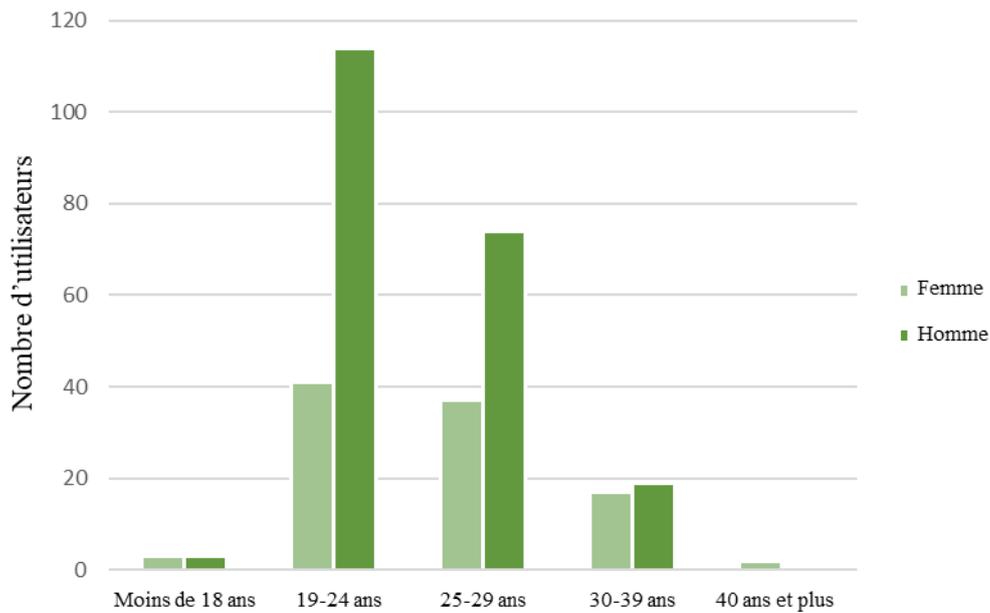


Figure 54 La répartition des répondants par âge et sexe



Figure 55 La répartition de l'ensemble utilisateurs de Douban par âge et sexe en 2017

Tous les répondants ont noté au moins 10 films. 31,6% des répondants ont noté plus de 500 films, 30,6% des répondants ont noté des films dont le nombre est entre 200-499 films, 37,7% des répondants ont noté moins de 200 films (Figure 56). 55,9% des répondants abonnaient certains critiques.

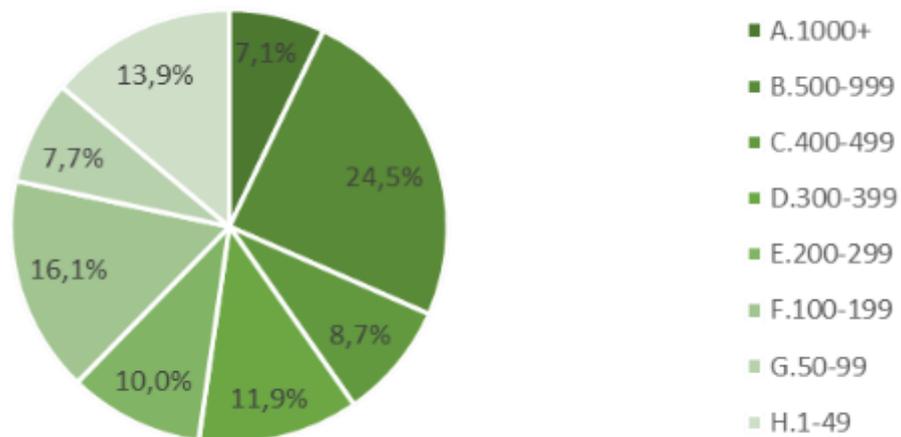


Figure 56 La répartition des répondants par le nombre de films qu'ils ont noté sur Douban

VIII.3 Résultats

Notre première question de recherche consiste à vérifier si les utilisateurs possèdent deux critères (subjectif et objectif) et si ces deux critères sont toujours cohérents.

57,4% des répondants sont d'accord avec l'**H1a** en déclarant qu'ils n'aiment pas toujours des films qu'ils jugent d'excellente qualité. 53,5% des répondants sont d'accord avec l'**H1b** en confirmant qu'ils aiment parfois des films qu'ils jugent de moindre qualité. Seulement 26,5% des répondants ne sont pas d'accord avec **H1a** ni **H1b** en affirmant qu'ils aiment tous les films de bonne qualité et n'aiment aucun film de moindre qualité. C'est-à-dire, pour ces 26,5% des répondants, ils n'ont qu'un critère pour évaluer le film car leur critère objectif et leur critère subjectif sont toujours pareils. Mais 73,5% des répondants, ils possèdent deux critères différents qui ne sont pas toujours d'accord. Ainsi, l'hypothèse 1 est confirmée par la majorité des répondants.

En ce qui concerne la deuxième question de recherche (**Q2** : dans les RSN pour les cinéphiles, est-ce qu'il existe certaines normes du groupe qui influence le comportement de notation ?), nous cherchons à vérifier si la norme du groupe « de noter le film en fonction de la qualité du film » influence les utilisateurs de Douban lorsqu'ils évaluent des films. Afin de vérifier notre deuxième hypothèse (**H2** : la norme du groupe « de noter le film en fonction de

la qualité du film » influence les utilisateurs de Douban), il faudrait, en premier lieu, examiner si cette norme influence les pensées et les comportements des utilisateurs de Douban. Ainsi, nous avons formulé notre troisième et quatrième hypothèse : **H3** : Pour les utilisateurs, la note du film représente la qualité du film et **H4** : les utilisateurs ont tendance à noter le film en fonction de la qualité perçue (le critère objectif) plutôt que de leurs propres préférences (le critère subjectif).

La troisième hypothèse consiste à examiner si la norme du groupe « de noter le film en fonction de la qualité du film » influence utilisateurs lorsqu'ils interprètent la notation du film. Plus précisément, pour eux, la note moyenne d'un film sur Douban représente la qualité du film ou les préférences de la communauté des utilisateurs de Douban. Le résultat montre que 68,3% des répondants estiment qu'une bonne note garantit une qualité supérieure et 75,2% des répondants pensent qu'une mauvaise note est due à la mauvaise qualité. Cela signifie que la plupart des utilisateurs estiment que la note d'un film sur Douban représente la qualité de ce film, en d'autres termes, qu'ils pensent que la plupart des utilisateurs notent un film en fonction de sa qualité. En fait, seulement 14,5% des répondants considèrent que la note d'un film ne représente que les préférences des utilisateurs pour le film. Il implique également que la norme du groupe « de noter du film en fonction de la qualité du film » influence les pensées de la plupart d'utilisateurs. L'hypothèse 3 est également confirmée par la plupart des répondants.

Ensuite, nous cherchons à savoir si cette norme du groupe influence les comportements de notation en interrogeant les utilisateurs comment ils notent le film en réel. Autrement dit, pour les utilisateurs ayant deux critères qui ne s'accordent pas toujours, selon quel critère notent-ils un film. Le résultat montre que 42,5 % des répondants déclarent qu'ils ne notent qu'en fonction du critère objectif, 40,3% des répondants affirment qu'ils notent seulement selon le critère subjectif et 17,1% des répondants indiquent qu'ils utilisent les deux critères.

Plus précisément, seulement 1,3% des répondants déclarent qu'ils ont donné des mauvaises notes (1 ou 2 étoiles) aux films qu'ils n'aimaient pas malgré la bonne qualité qu'ils avaient perçue. 57,5% des répondants confirment qu'ils ont donné des bonnes notes (4 ou 5 étoiles) aux films de bonne qualité même s'ils ne les aiment pas personnellement. L'hypothèse H4a peut être confirmée. En ce qui concerne l'hypothèse H4b, 6,1% des répondants déclarent qu'ils ont donné des mauvaises notes (1 ou 2 étoiles) aux films de mauvaise qualité même s'ils aimaient personnellement. 53,5% des répondants ont donné des bonnes notes (4 ou 5 étoiles) aux films qu'ils aiment malgré la moindre qualité. L'hypothèse H4b ne peut pas être confirmée.

Il suggère que la plupart des répondants tiennent compte des deux critères en même temps lorsqu'ils notent le film. Notre H4 ne peut pas être complètement confirmée.

En résumé, la plupart des répondants estiment qu'ils disposent de deux critères différents pour noter les films et qu'ils ne les notent pas qu'en fonction d'un seul critère. La norme « de noter le film en fonction de la qualité du film » influence les pensées et les de la majorité des utilisateurs. Cependant, elle semble avoir moins d'influence lorsqu'il s'agit des films de moindre qualité mais appréciés par les utilisateurs. Le résultat montre qu'à part les films qu'ils aiment, les utilisateurs ont également tendance à attribuer des bonnes notes aux films de bonne qualité quelles que soient leurs préférences. Lorsque l'utilisateur attribue une bonne note à un film, il demeure ignoré si cet utilisateur aime ce film ou s'il s'agit d'une bonne qualité perçue. Dans ce cas, notre deuxième hypothèse ne peut pas être complètement confirmée. Il serait difficile de déterminer selon quel critère les utilisateurs notent les films et ainsi cela nous empêche d'inférer précisément les préférences des utilisateurs.

Sauf la norme explicite « de noter le film en fonction de la qualité du film », certaines normes implicites et l'opinion d'autrui influencent également le comportement de notation. Les évaluations des autres empêcheraient les utilisateurs de s'exprimer leur propre opinion. Effectivement, le résultat montre que la majorité des répondants ne nient pas qu'ils hésitent à noter les films quand ils s'aperçoivent que leurs opinions ne sont pas d'accord avec celles des autres. Plus précisément, 73,8% des répondants hésitent lorsque leurs opinions sont en désaccord avec la majorité (la note moyenne), 70,1% des répondants sont hésitants si leurs opinions ne sont pas d'accord avec leurs amis sur Douban, et particulièrement 78,4% des répondants déclarent qu'ils sont hésitants quand ils ne sont pas d'accord avec les critiques qu'ils suivent. Cela suggère que les opinions d'autrui sont très susceptibles d'avoir une influence sur le comportement de notation.

73,8 % des répondants confirment qu'ils ont tendance à augmenter la note qu'ils avaient l'intention de donner quand ils constatent que le film est très bien noté par la majorité. 73,5% des répondants déclarent qu'ils finissent par donner une note plus élevée qu'ils avaient voulu aux films très bien notés par leurs amis. 77,8% des répondants confirment qu'ils augmentent les notes des films appréciés par les critiques qu'ils suivaient. Cela suggère que la majorité des utilisateurs sont influencés par ces trois types d'évaluations positives. Le résultat montre également que les influences de ces trois types d'évaluations sont positivement corrélées (Figure 57).

		Influence from average ratings	Influence from friends	Influence from film critics
Influence from average ratings	Pearson Correlation	1	.611**	.623**
	Sig. (2-tailed)		.000	.000
	N	297	297	166
Influence from friends	Pearson Correlation	.611**	1	.643**
	Sig. (2-tailed)	.000		.000
	N	297	297	166
Influence from film critics	Pearson Correlation	.623**	.643**	1
	Sig. (2-tailed)	.000	.000	
	N	166	166	166

Figure 57 Corrélations entre les influences d'évaluations positives des notes moyennes, des amis et des critiques de cinéma

Pour les films auxquels les répondants avaient l'intention d'attribuer une bonne note, 62,8 % des répondants affirment qu'ils donnent des notes plus faibles s'ils s'aperçoivent que ces films sont mal notés par la majorité des utilisateurs. 62,7% des répondants déclarent qu'ils attribuent des notes moins élevées qu'ils avaient l'intention de donner aux films très mal notés par leurs amis. Et 71,3% des répondants confirment qu'ils finissent par donner des notes faibles aux films très mal notés par les critiques qu'ils suivaient. Pareillement, les corrélations positives existent (Figure 58).

Figure 58 Corrélations entre les influences d'évaluations négatives des notes moyennes, des amis et des critiques de cinéma

		Influence from average ratings	Influence from friends	Influence from film critics
Influence from average ratings	Pearson Correlation	1	.688**	.562**
	Sig. (2-tailed)		.000	.000
	N	297	297	166
Influence from friends	Pearson Correlation	.688**	1	.638**
	Sig. (2-tailed)	.000		.000
	N	297	297	166
Influence from film critics	Pearson Correlation	.562**	.638**	1
	Sig. (2-tailed)	.000	.000	
	N	166	166	166

Figure 58 Corrélations entre les influences d'évaluations négatives des notes moyennes, des amis et des critiques de cinéma

Selon le résultat, les trois types de notations précédentes ont influencé la plupart des répondants. En particulier, les notes des critiques que les utilisateurs suivent ont influencé relativement plus de répondants que les notes moyennes tandis que les avis des amis ont relativement moins d'influence sur les répondants. Les notations positives ont un impact sur plus de répondants que les notations négatives. Cela suggère que les utilisateurs ne sont pas seulement influencés avant de regarder des films, mais que la plupart d'entre eux sont également influencés lorsqu'ils notent des films après les avoir regardés. C'est-à-dire que les notes des utilisateurs sont influencées par les notes précédentes (les notes moyennes, notes de leurs amis et notes des critiques qu'ils suivent). Le résultat montre également que les influences de ces notes précédentes sont positivement corrélées. Cela signifie que certains utilisateurs sont plus susceptibles d'être influencés par les autres. Ainsi, nous arguons que les notations de l'utilisateur peuvent ne pas refléter exactement ses opinions réelles.

VIII.4 Conclusion

Cette enquête par questionnaire auprès des utilisateurs de Douban a montré les facteurs qui sont susceptibles d'influencer le comportement des utilisateurs en matière de notation des films.

Le résultat confirme nos observations et hypothèses en suggérant que la majorité des utilisateurs ont deux critères (objectif et subjectif) pour noter le film et que ces deux critères ne sont pas toujours pareils. C'est-à-dire que l'utilisateur n'aime pas toujours les films de bonne qualité et que les films de moindre qualité peuvent être appréciés par lui. Cela prouve que le critère objectif et le critère subjectif coexistent pour la majorité des utilisateurs. Il est possible que la notation d'un film puisse être complètement différente en fonction de ces différents critères.

Nous constatons également que « de noter le film en fonction de la qualité du film » est une norme fondamentale sur Douban qui influence fortement sur les pensées et les comportements des utilisateurs. La plupart des utilisateurs considèrent la note moyenne d'un film peut représenter la qualité du film. Ils respectent également cette norme lors de la notation

des films, en tenant compte de la qualité du film. Ils affirment qu'ils ont donné des bonnes notes aux films d'excellente qualité bien qu'ils ne les aiment pas personnellement. Il y a cependant des exceptions, de nombreux utilisateurs déclarant qu'ils insistent pour donner une note élevée aux films qu'ils préfèrent malgré la mauvaise qualité.

Les opinions d'autrui influencent aussi le comportement de notation. La note moyenne (soit l'opinion de la majorité), les notes attribuées par des amis et les critiques ont des influences sur les utilisateurs. Lors de la notation, la plupart d'utilisateurs hésitent s'ils trouvent que leurs opinions ne sont pas en accord avec celles des autres. Ils confirment qu'ils donnaient souvent des notes plus élevées aux films bien notés par les autres et vice versa.

En résumé, cette étude prouve la coexistence de deux critères différents pour noter le film. Elle démontre également que l'influence normative et sociale ont un impact sur le comportement de notation dans un RSN pour les cinéphiles. Cependant, le résultat de cette étude se base sur les déclarations des utilisateurs de Douban. Dans le but de vérifier si leurs comportements de notation se conforment à leurs déclarations, nous avons effectué une série d'études comparatives qui comparent les notes des mêmes films sur Douban et sur les trois autres plateformes et dont les résultats sont présentés dans le chapitre suivant.

Chapitre IX Études comparatives

Nous avons réalisé une série d'études qui comparent les notes des mêmes films sur les 4 plateformes afin de vérifier si la norme « de noter le film en fonction de la qualité du film » influence la notation du film, s'il existe certaines normes implicites qui conduisent à favoriser des films avec certaines caractéristiques et si la différence culturelle influence le comportement de notation.

D'abord, nous avons comparé les notes de 3292 films sur Allociné et celles sur

SensCritique. Ce sont deux RSN pour les cinéphiles français. Ils partagent une même norme explicite qui encourage les utilisateurs à noter les films objectivement. Le résultat sert à comparer les résultats des deux études suivantes.

Nous avons comparé ensuite les notes de 3369 films sur Douban et sur Maoyan. Ce sont deux plateformes pour les cinéphiles chinois. Ils partagent également la même norme explicite qu'Allociné et SensCritique. La seule différence est que Douban est un RSN tandis que Maoyan ne peut pas être qualifié de RSN. Vu que l'influence normative est considérée comme moins puissante sur une plateforme " non RSN " que sur RSN. Le résultat de cette comparaison nous permet de vérifier et de mieux comprendre l'influence normative sur le comportement de notation dans les RSN.

La dernière comparaison s'est déroulée entre Douban et Allociné. Nous avons comparé les notes de 2728 films sur ces deux RSN. Ce sont deux RSN ayant la même norme explicite mais issus de deux cultures différentes. Cette étude nous permet d'examiner l'influence de la différence culturelle sur le comportement de notation.

IX.1 Question de recherche et hypothèses

Le résultat de l'enquête par questionnaire montre que « de noter le film en fonction de la qualité du film » est la norme explicite la plus importante pour les utilisateurs de Douban. Selon les déclarations des répondants, cette norme influence fortement les utilisateurs et leur comportement de notation.

Dans le but de vérifier l'impact de cette norme, nous avons comparé respectivement les notations des films sur Douban avec des mêmes films sur Allociné et sur Maoyan. Étant donné que ces 4 plateformes disposent tous de la même norme « de noter du film en fonction de la qualité du film », nous proposons notre sixième hypothèse :

H6a : Les notations des mêmes films sur Allociné et sur SensCritique seraient positivement corrélées

H6b : Les notations des mêmes films sur Douban et sur Maoyan seraient positivement corrélées.

H6c : Les notations des mêmes films sur Douban et sur Allociné seraient positivement corrélées.

Nous cherchons aussi à examiner s'il existe certaines normes implicites qui conduisent au favoritisme des films avec certaines caractéristiques. Donc, nous formulons notre septième hypothèse :

H7a : Il existe des films avec certaines caractéristiques qui sont plus appréciés sur Allociné que sur SensCritique.

H7b : Il existe des films avec certaines caractéristiques qui sont plus appréciés sur Douban que sur Maoyan.

H7c : Il existe des films avec certaines caractéristiques qui sont plus appréciés sur Douban que sur Allociné.

L'influence normative sur Maoyan semble moins puissante que celle sur Douban, sur SensCritique et sur Allociné car il n'est pas un RSN qui favorise les interactions entre les utilisateurs. Cela signifie que les utilisateurs de Maoyan seraient moins influencés par la norme « de noter du film en fonction de la qualité du film ». C'est la raison pour laquelle nous proposons la huitième hypothèse :

H8 : Par rapport à la corrélation entre les notations sur Douban et sur Maoyan, la corrélation entre les notations sur SensCritique et sur Allociné serait plus forte.

Nous considérons également que la différence culturelle pourrait être un autre facteur qui influence le comportement de notation du film. Cela concerne notre troisième question de recherche :

Q3 : Est-ce que la différence culturelle influence le comportement de notation du film ?

Dans le but de vérifier si la différence culturelle influence le comportement de notation, nous comparons les notations de mêmes films sur deux RSN pour les cinéphiles issus d'une même culture : Allociné sur SensCritique et les notations de mêmes films sur deux RSN issus de deux cultures différentes. Nous proposons notre dixième hypothèse :

H10 : Les notations sur Allociné et sur SensCritique seraient plus positivement corrélées par rapport aux notations sur Allociné et sur Douban.

IX.2 Chiffre statistique

Au total, 14560 films sortis pendant 2014-2018 ont été trouvés sur Douban parmi lesquels 5061 films ont été notés par au moins de 50 utilisateurs. Quant à Maoyan, il y avait 2617 films qui ont été notés par plus de 50 utilisateurs parmi les 10257 films trouvés en totalité. Il y a 4992 films sortis pendant 2014-2018 et 3345 films qui ont été notés par plus de 50 utilisateurs sur Allociné. Sur SensCritique, 4750 films ont été notés par au moins de 50 utilisateurs parmi 7983 films trouvés (Tableau 5).

Nombre de film Plateforme	Nombre d'utilisateurs qui notent				Total
	50-99	1 cent-1 mille	1 mille -10 milles	Plus de 10 milles	
Douban	363	2001	1663	1034	5061
Maoyan	1468	1144	867	846	4325
Allociné	603	1706	934	102	3345
SensCritique	986	2363	1143	258	4750

Tableau 5 La répartition de films par le nombre d'utilisateurs qui les notent

Nous avons effectué 3 comparaisons entre : 1) les notations des mêmes films sur SensCritique et sur Allociné ; 2) les notations des mêmes films sur Douban et sur Maoyan ; 3) les notations des mêmes films sur Douban et sur Allociné.

En ayant recours à un robot d'indexation, nous avons recueilli les notes de tous les films sortis pendant 2014-2018 sur les 4 plateformes pour comparer les notes des mêmes films sur les plateformes différentes. Nous avons identifié 3292 films qui sont notés par au moins de 50 utilisateurs d'Allociné et de SensCritique, 3369 films qui sont notés par au moins de 50 utilisateurs de Maoyan et de Douban en commun, et 2728 films qui sont notés par plus de 50 utilisateurs à la fois sur Allociné et sur Douban.

IX.3 Étude comparative: la comparaison entre Allociné et SensCritique

La première étude comparative que nous avons effectuée est la comparaison entre Allociné et SensCritique. Ce sont deux RSN pour les cinéphiles français qui encouragent leurs utilisateurs à évaluer les films de manière objective. Ayant la même origine culturelle et la même norme de groupe, les comportements de notation de leurs utilisateurs sont très similaires. Le résultat de cette étude sert de point de comparaison avec les résultats des deux études suivantes.

IX.3.1 Chiffres statistiques

Nous avons identifié 3292 films sortis entre 2014 et 2018 qui ont tous été notés par plus de 50 utilisateurs sur Allociné et sur SensCritique. En général, la note moyenne des films sur SensCritique est 5,70 et celle sur Allociné est 6,29 (soit par 0,49 point plus haute) (Figure 59 et Figure 60). Ces 3292 films proviennent de 78 pays et régions. Les films américains (1359 films, soit 41,3%) et les films français (934 films, soit 28,4%) sont les plus nombreux (Figure 61).

Au total, nous comptons 30 genres sur Allociné et Douban : action, animation, arts martiaux, aventure, biopic, Bollywood, catastrophe, comédie, comédie dramatique, comédie musicale, court-métrage, documentaire, drame, épouvante-horreur, érotique, espionnage, expérimental, famille, fantastique, guerre, historique, judiciaire, musical, péplum, policier, romance, science-fiction, sport, thriller et western.

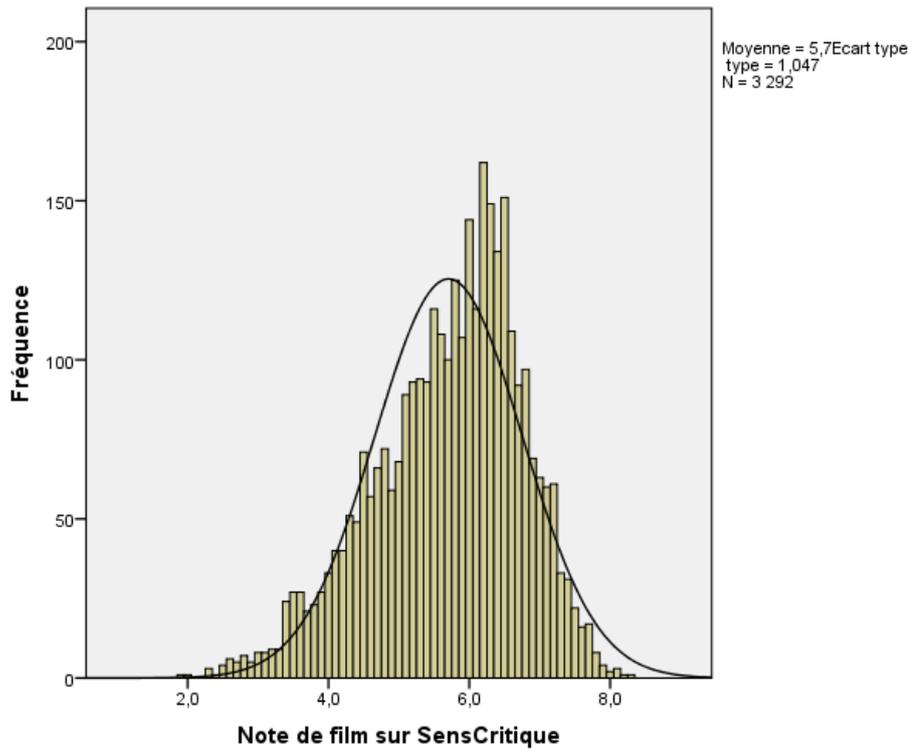


Figure 59 La réparation des notes sur SensCritique

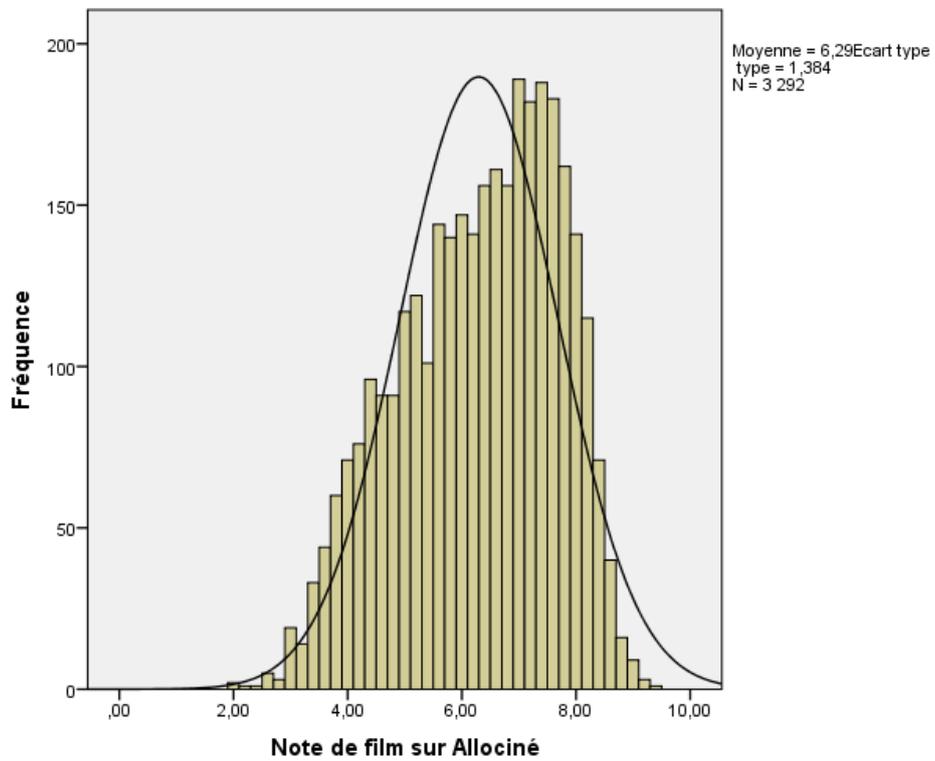


Figure 60 La réparation des notes sur Allociné

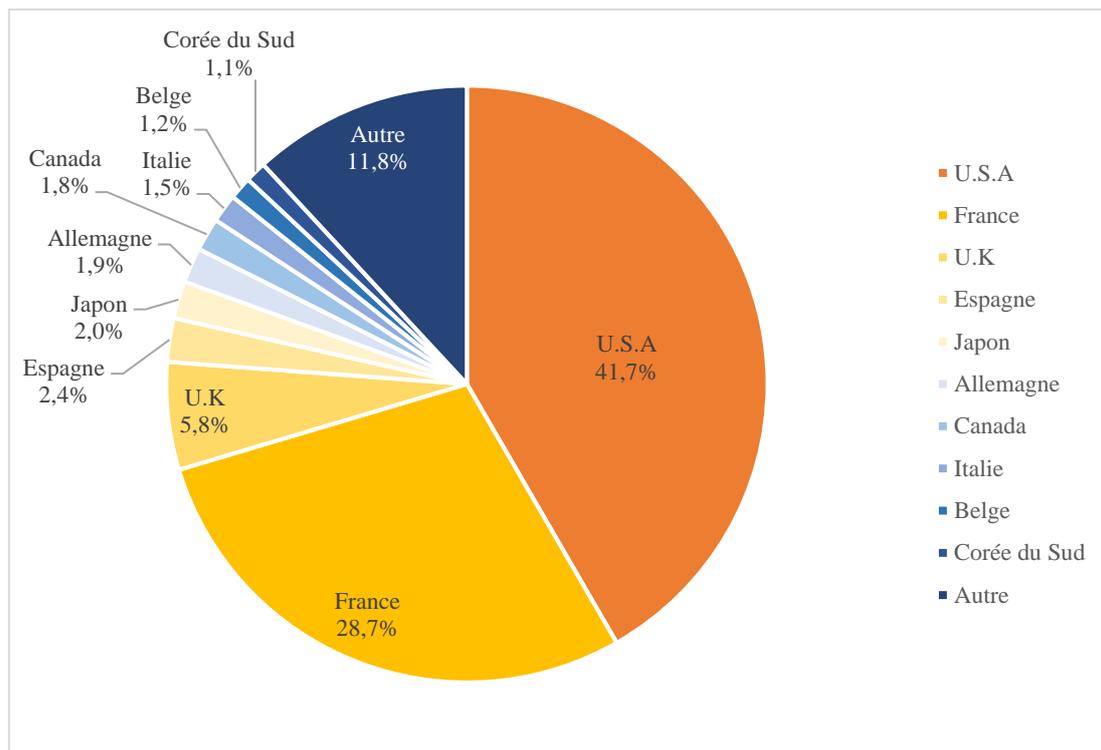


Figure 61 La répartition des films par nationalité

IX.3.2 Les films ayant les notations différentes sur les deux plateformes

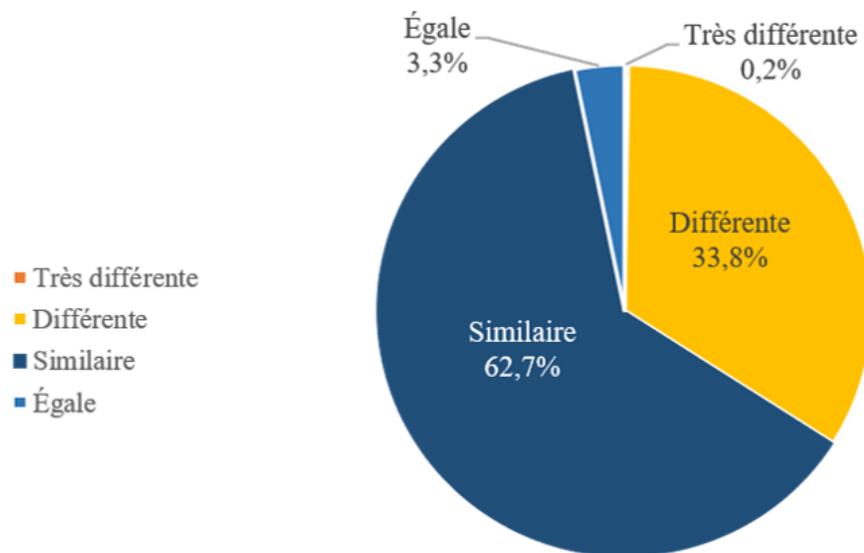
Nous constatons une corrélation positive très forte entre les notes de SensCritique et Allociné ($r=0,821$) (Tableau 6). Nous pouvons maintenant valider notre hypothèse **H6a**, soit les notations de mêmes films sur Allociné et sur SensCritique sont plus positivement corrélées. Ayant la même origine culturelle et la même norme du groupe, les utilisateurs de deux RSN se comportent de manière très similaire lorsqu'ils évaluent des films.

		Notes sur Allociné	Notes sur SensCritique
Notes sur Allociné	Corrélation de Pearson	1	,821**
	Sig. (bilatérale)		,000
	N	3292	3292
Notes sur SensCritique	Corrélation de Pearson	,821**	1
	Sig. (bilatérale)	,000	
	N	3292	3292

** . La Corrélacion est significative au niveau 0,01 (bilatéral).

Tableau 6 La corrélation entre les notes des mêmes films sur Allociné et SensCritique

Malgré la forte corrélation entre les notations des films sur Allociné et sur SensCritique, des différences existent. En effet, nous constatons qu'il y a 34% des films qui ont reçu des notations différentes (33,8%) et très différentes (0,2%)¹⁴⁵ (Figure 63).



	Très différente	Différente	Similaire	Égale
Différence entre la note moyenne sur Allociné et sur SensCritique	>3	2,9-1	0,9-0,1	0
Nombre de films	8	1113	2064	107

Figure 62 Répartition des films par la différence de notes des mêmes films sur Allociné et sur SensCritique

¹⁴⁵ Étant donné que les notes moyennes sont représentées sur une échelle de 1 à 10 points, nous considérons que les notes d'un film sur deux plateformes sont similaires si la différence entre les deux notes d'un film est inférieure à 1 point, les notes sont différentes si la différence entre les deux notes est entre 1-2,9 points et les notes sont très différentes si la différence entre les deux notes est supérieure à 3 points.

Bien que globalement, les notes des mêmes films sur les deux plateformes soient très similaires et que la plupart des films soient mieux notés sur Allociné, nous constatons que les films expérimentaux et d'épouvante-horreur sont moins appréciés par les utilisateurs d'Allociné que ceux de SensCritique. En revanche, aucun film de Bollywood, de court-métrage et d'espionnage obtient une meilleure note sur SensCritique que sur Allociné (Figure 63). Nous pouvons ainsi confirmer notre hypothèse H7a : Il existe des films avec certaines caractéristiques qui sont plus appréciés sur Allociné que sur SensCritique.

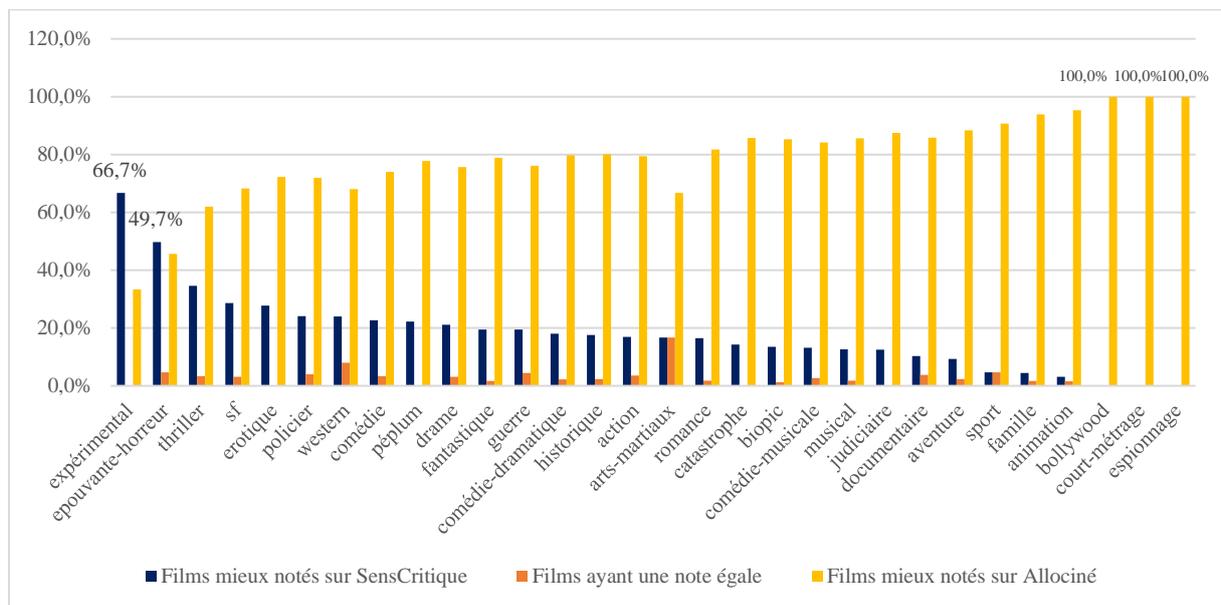


Figure 63 Comparaison des notes des films par genre sur les deux plateformes

En résumé, les résultats montrent que, par rapport aux utilisateurs d'Allociné, la majorité des utilisateurs de SensCritique disposent de critères plus sévères pour évaluer des films. En même temps, cette étude comparative montre que les utilisateurs de ces deux plateformes sont influencés par certains normes en commun lorsqu'ils notent les films et que chaque plateforme a ses propres critères et préférences pour évaluer les films.

IX.4 Étude comparative entre RSN et non RSN : Douban et Maoyan

La comparaison entre Allociné et SensCritique montre que les comportements des

utilisateurs en matière de notation des films sur deux RSN différentes sont très similaires lorsqu'ils ont la même culture et la même norme du groupe. Selon la théorie de la désindividuation, l'influence normative sera renforcée suite aux interactions entre les membres du groupe. Nous avons alors comparé les notations des mêmes films sur Douban et sur Maoyan dans la deuxième étude comparative. Cette comparaison entre un RSN et un non RSN issus de la même culture et disposant de la même norme du groupe nous permet de vérifier si le modèle SIDE est également applicable dans les RSN, c'est-à-dire, si les interactions entre les utilisateurs de RSN renforcent l'influence normative qui affecte finalement le comportement de notation.

IX.4.1 Résultats

Les 3369 films identifiés sont tous notés par au moins de 50 utilisateurs à la fois sur Douban et sur Maoyan.

Les 3369 films proviennent respectivement de 59 pays et régions (*Figure 64*). Parmi lesquels les films de la Chine (1161 films), des États-Unis (868 films), du Japon (361 films), de la Grande-Bretagne (166 films), de Corée du Sud (206 films), de la France (110 films), de Taïwan (73 films), de Hongkong (61 films), de l'Inde (46 films), et de l'Allemagne (44 films) représentent 91,9% des films au total.

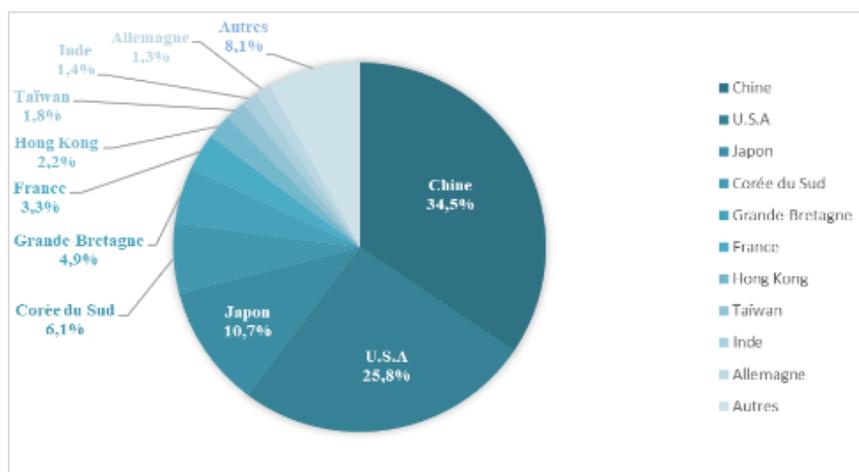


Figure 64 La répartition des films par nationalité

Chaque film est classifié dans au moins un genre. Au total, nous comptons 24 genres :

drame, comédie, action, romance, thriller, aventure, animation, policier, fantastique, science-fiction, épouvante-horreur, famille, biopic, historique, guerre, musical, documentaire, péplum, sport, catastrophe, court métrage, érotique, arts-martiaux et western.

En général, la note moyenne de tous les films sur Douban est 6,08 et celle sur Maoyan est 7,13 (soit par 1,05 point plus haute) (voir *Figure 65* et *Figure 66*).

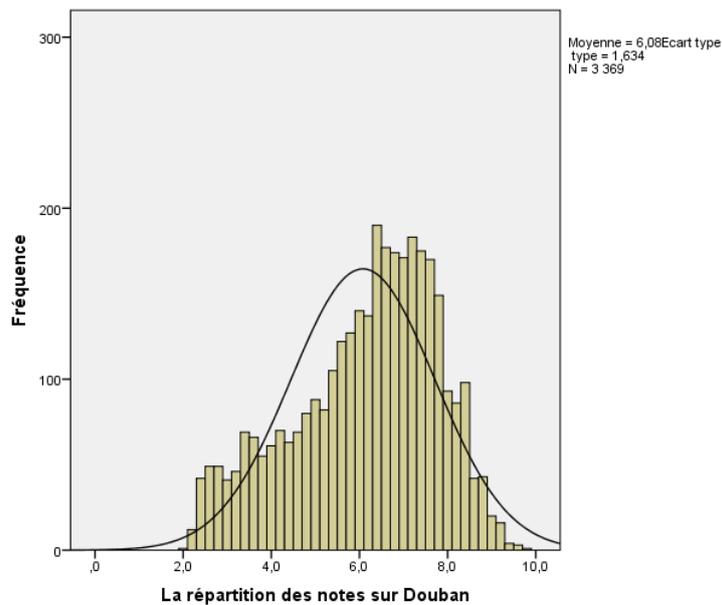


Figure 65 La répartition des notes sur Douban

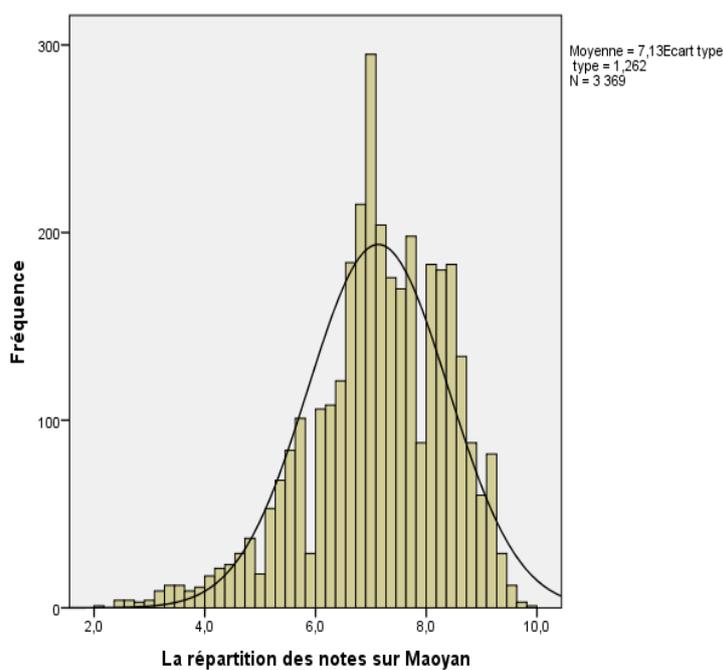


Figure 66 La répartition des notes sur Maoyan

IX.4.2 Les films ayant les évaluations différentes sur les deux plateformes

Nous constatons qu'il existe une relation modérément positive entre les notes de Douban et celles de Maoyan ($r=0,552$) (Tableau 7). L'hypothèse 6b est confirmée.

	DBnote	MYnote
DBnote	1	,522**
Corrélation de Pearson		
Sig. (bilatérale)		,000
N	3369	3369
MYnote	,522**	1
Corrélation de Pearson		
Sig. (bilatérale)		,000
N	3369	3369

** . La Corrélation est significative au niveau 0,01 (bilatéral).

Tableau 7 La corrélation entre les notes des mêmes films sur Douban et sur Maoyan

En comparant la corrélation positive très forte ($r=0,821$) entre les notations sur Allociné et sur SensCritique, cette corrélation positive entre les notations sur Douban et sur Maoyan est très faible. Ayant la même culture et la même norme du groupe, le comportement de notation des utilisateurs d'un RSN et d'un non RSN montre une très grande différence que celui des utilisateurs de deux RSN. Par conséquent, nous arguons que Le modèle SIDE est également applicable dans les RSN pour les cinéphiles et que les interactions entre les utilisateurs renforcent l'influence normative qui affecte largement le comportement de notation. Ainsi, notre huitième hypothèse (H8 : Par rapport à la corrélation entre les notations sur Douban et sur Maoyan, la corrélation entre les notations sur SensCritique et sur Allociné serait plus forte) peut être confirmée.

Ensuite, nous avons identifié les films qui ont reçu des évaluations différentes sur Douban et sur Maoyan. Nous constatons que la plupart des films partagent certains points en commun. Nous avons identifié 1619 films (soit 48,1% des films) qui ont reçu des évaluations différentes ou très différentes sur les deux plateformes. Seulement 0,2% des films ont reçu des notations très différentes. Globalement, 12,6% des films ont reçu des évaluations très différentes sur Douban et sur Maoyan, 35,3% des films ont reçu des évaluations différentes sur les deux

plateformes et 51,9% des films ont reçu des évaluations similaires¹⁴⁶ sur les deux plateformes (Figure 67).

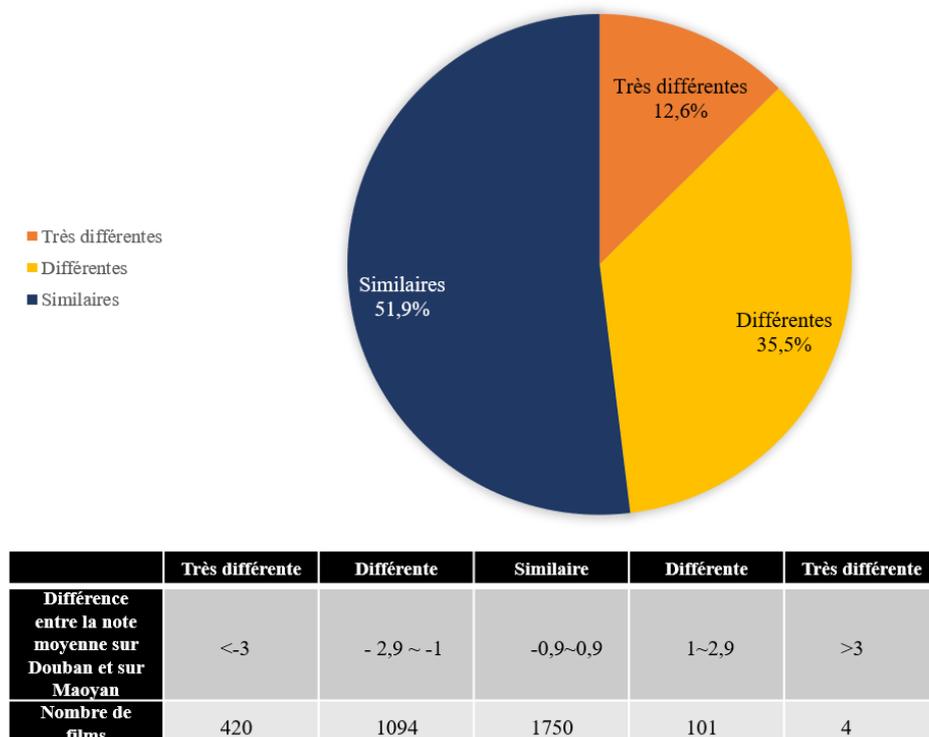


Figure 67 Répartition des films par la différence de notes des mêmes films sur Douban et sur Maoyan

Nous constatons que les utilisateurs de Douban restent en désaccord avec les utilisateurs de Maoyan sur les évaluations des films chinois. En effet, 390 films chinois (soit 33,6% des films chinois au total) représentent 92% des films dont la note sur Douban est très différente à la note sur Maoyan. 575 films chinois (soit 49,5% des films chinois au total) représentent 48,1% des films qui ont reçu les évaluations différentes sur Douban et sur Maoyan. Seulement 16,9% des films chinois ont obtenu les notes moyennes similaires sur les deux plateformes. Par contraste avec les films chinois, la plupart des films d'autres pays ont reçu des évaluations similaires sur les deux plateformes (Figure 68).

¹⁴⁶ Similaire : la différence entre deux notes sur les deux plateformes est inférieure à 1 point

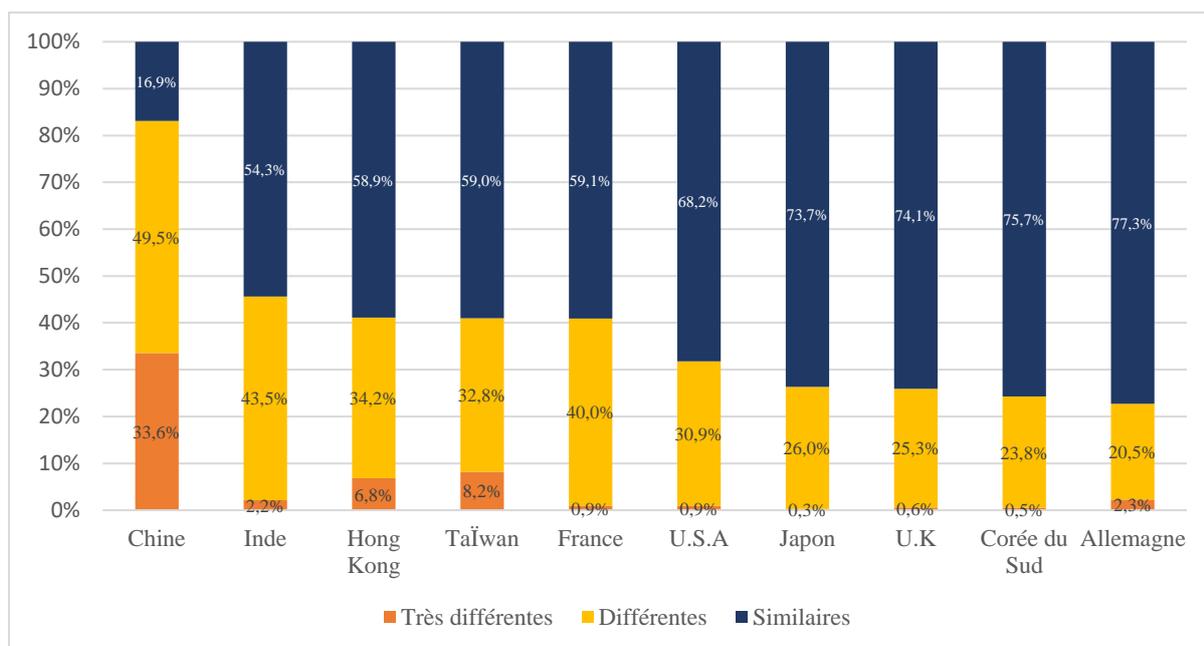


Figure 68 Différence de notes des films des 10 pays ayant le plus grand nombre de films sur les deux plateformes

Comparativement aux utilisateurs de Douban, les internautes sur Maoyan manifestent un plus grand enthousiasme pour les films chinois. Effectivement, 95,1% des films chinois ont reçu une note supérieure sur Maoyan par rapport sur Douban. Ce n'est pas seulement le cas des films chinois, en fait, les résultats montrent que 93,5% des films ayant reçu des notations différentes ou très différentes sur les deux plateformes ont obtenu de meilleures notes sur Maoyan que sur Douban. Cela implique que les utilisateurs de Douban disposent des critères plus sévères par rapport aux utilisateurs de Maoyan. En effet, parmi tous les 1854 films, 88,2% des films ont obtenu une meilleure note sur Maoyan que sur Douban. Dans ce cas, les films qui ont obtenu une meilleure note sur Douban que sur Maoyan semblent particuliers et avoir certaines caractéristiques préférées exclusivement par les utilisateurs de Douban. Ainsi, nous avons examiné les caractéristiques de ces films dont la note sur Douban est supérieure à leur note sur Maoyan.

IX.4.3 Les films les plus appréciés sur Douban

Au total, il y a 883 films sur 3369 (soit 26,2%) qui ont eu une meilleure note sur Douban

que sur Maoyan. Les films chinois représentent seulement 5,2% des films qui ont obtenu de meilleures notes sur Douban, bien que le nombre total de films chinois soit bien supérieur à celui des autres pays (*Figure 64*).

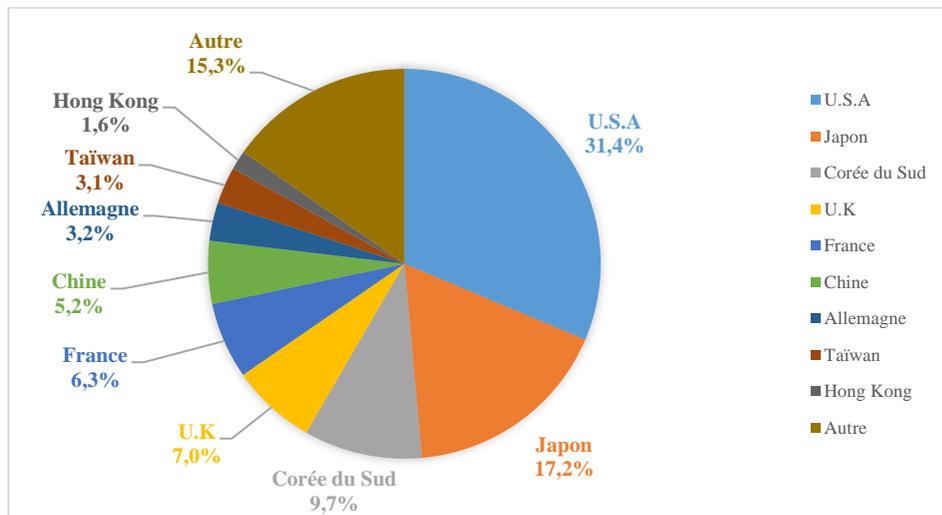


Figure 69 La répartition de nationalité des films qui sont mieux notés sur Douban

En effet, seulement 4% des films chinois ont obtenu de meilleures notes sur Douban, par contraste, 63,6% des films allemands sont plus appréciés sur Douban et 50,9% des films français sont mieux notés sur Douban. Il semble que certaines normes implicites existent sur Douban qui influencent la majorité des utilisateurs lorsqu'ils notent des films (*Figure 70*).

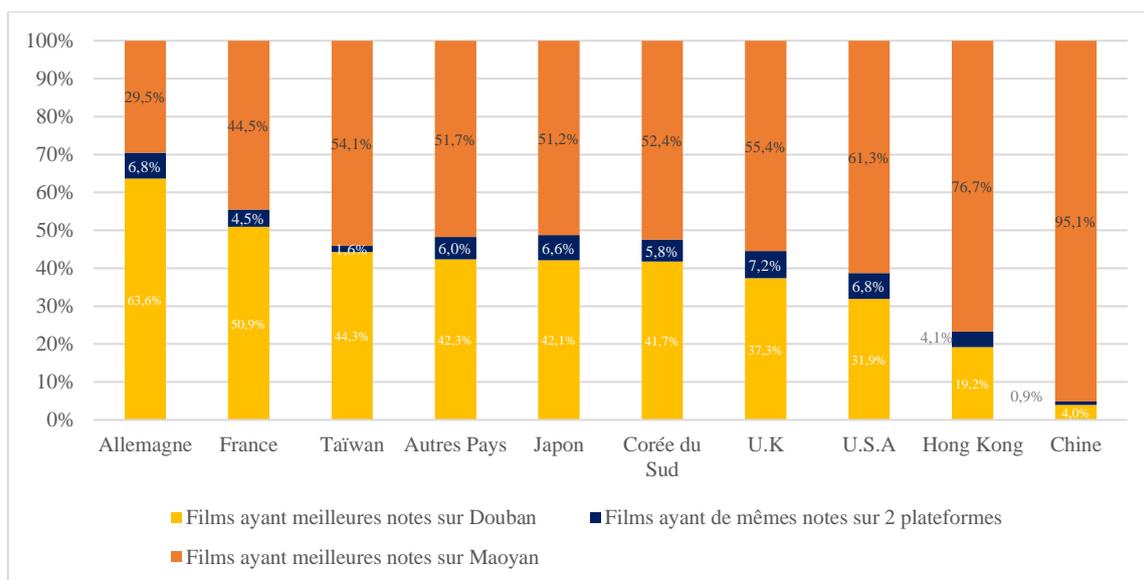


Figure 70 Comparaison des notes des films des dix pays ayant le plus grand nombre de films sur les deux plateformes

Nous constatons que les films de certains genres ou thèmes sont plus appréciés sur Douban que sur Maoyan : 72% des courts métrages, 73,2% des films documentaires et 51,9% des film biopic ont obtenu de meilleures notes sur Douban. Au contraire, les films de péplum, de catastrophe et d'aventure sont le moins appréciés sur Douban. Nous confirmons notre hypothèse H7b : Il existe des films avec certaines caractéristiques qui sont plus appréciés sur Douban que sur Maoyan.

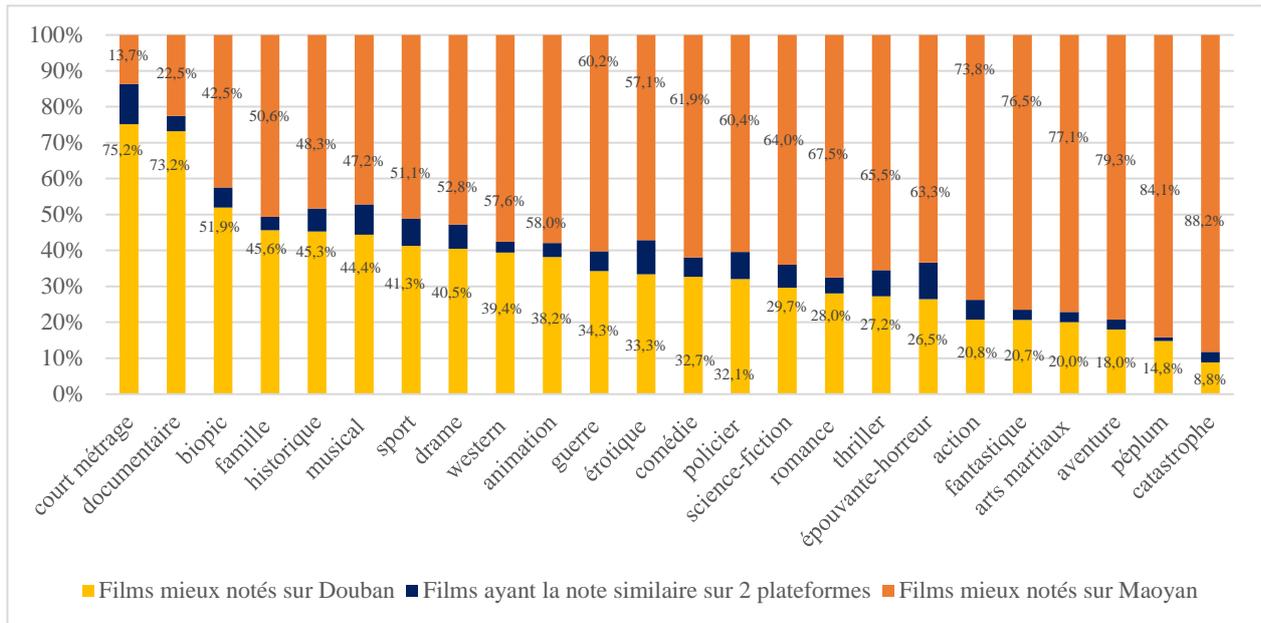


Figure 71 Comparaison des notes des films par genre sur les deux plateformes

En résumé, le résultat suggère que les utilisateurs de ces deux plateformes sont influencés par certaines normes en commun lorsqu'ils notent les films et qu'il existe également certaines normes et préférences spécifiques à chaque plateforme pour évaluer les films. Par rapport aux utilisateurs de Maoyan, la majorité des utilisateurs de Douban disposent des critères plus sévères pour évaluer des films. En tenant compte du résultat de la première comparaison, la norme explicite « de noter le film en fonction de la qualité du film » a une influence plus importante sur les utilisateurs de Douban que ceux de Maoyan. Nous arguons que la norme du groupe a une influence considérable sur les utilisateurs de RSN.

IX.5 Étude comparative entre deux RSN : Douban et Allociné

Les deux études montrent que les notations sur Allociné et sur SensCritique sont plus positivement corrélées que les notations sur Douban et sur Maoyan. Cela suggère que les quatre plateformes partagent certaines normes communes et que l'influence normative est plus puissante sur les RSN. L'étude montre également que les utilisateurs de chaque plateforme ont également leurs propres normes et préférences spécifiques. Nous constatons que les utilisateurs de deux RSN français se comportent de manière beaucoup plus similaire. Nous nous intéresserons à la question de savoir si la différence culturelle est un autre facteur qui peut influencer l'appréciation des films, c'est pourquoi Nous avons mené donc une troisième étude comparative dans laquelle nous avons comparé les comportements des utilisateurs de deux RSN issus de deux cultures différentes : Allociné et Douban.

IX.5.1 Résultats

Les 2728 films identifiés sont tous notés par au moins de 50 utilisateurs à la fois sur Douban et sur Allociné. Les 2728 films proviennent de 75 pays et régions. Près de la moitié des films sont produits par les États-Unis (1320 films, soit 48.4%). Les films de la France (506 films), de la Grande-Bretagne (181 films), de l'Espagne (73 films), du Japon (66 films), de l'Allemagne (59 films), du Canada (51 films), l'Italie (41 films), de la Corée du Sud (36 films) et de la Chine (33 films), représentent 38,3% des films au total (Figure 72).

Chaque film est classifié dans au moins un genre. Au total, nous comptons 29 genres sur Allociné et Douban : action, animation, arts martiaux, aventure, biopic, catastrophe, comédie, comédie dramatique, comédie musicale, court-métrage, documentaire, drame, épouvante-horreur, érotique, espionnage, expérimental, famille, fantastique, guerre, historique, judiciaire, musical, péplum, policier, romance, science-fiction, sport, thriller et western.

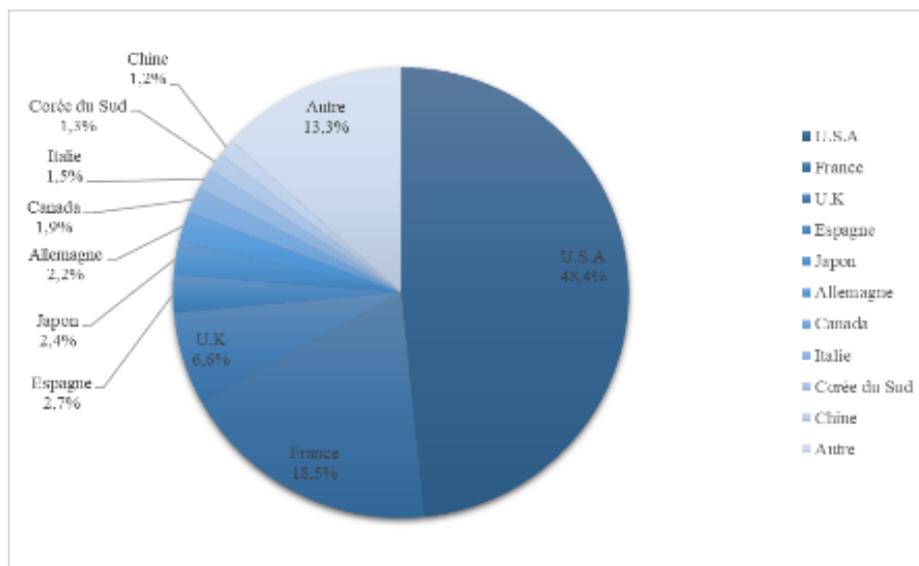


Figure 72 Répartition des films par leur nationalité

En général, la note moyenne de tous les films sur Douban est 6,60 et celle sur Allociné est 6,34¹⁴⁷ (une différence de 0,25 point) (Figure 73 et Figure 74).

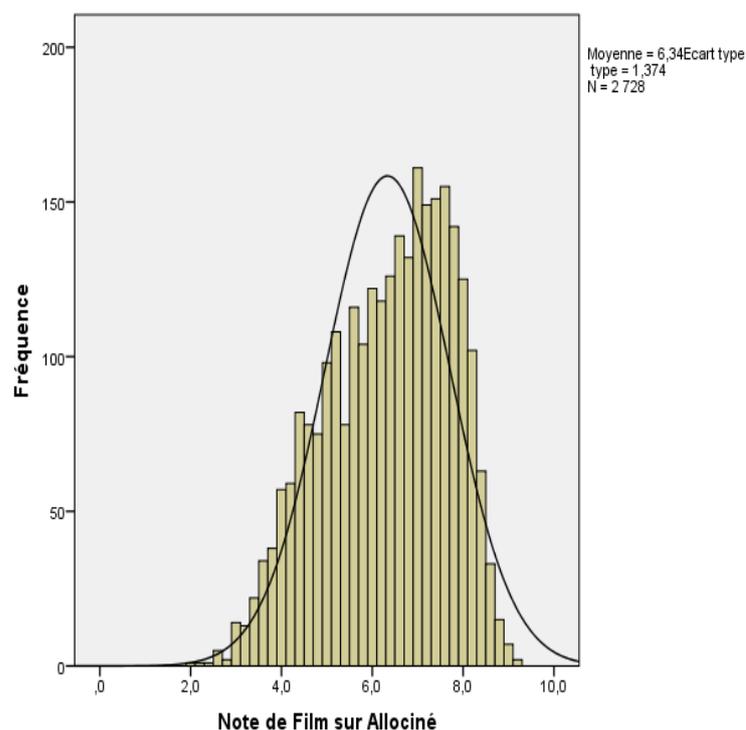


Figure 73 La répartition des notes sur Allociné

¹⁴⁷ Du fait que les notes d'Allociné sont sur une échelle de 1 à 5 points et que celles de Douban sont sur une échelle de 1 à 10 points, nous convertissons les notes d'Allociné en une échelle de 1 à 10 points lorsque nous faisons des analyses.

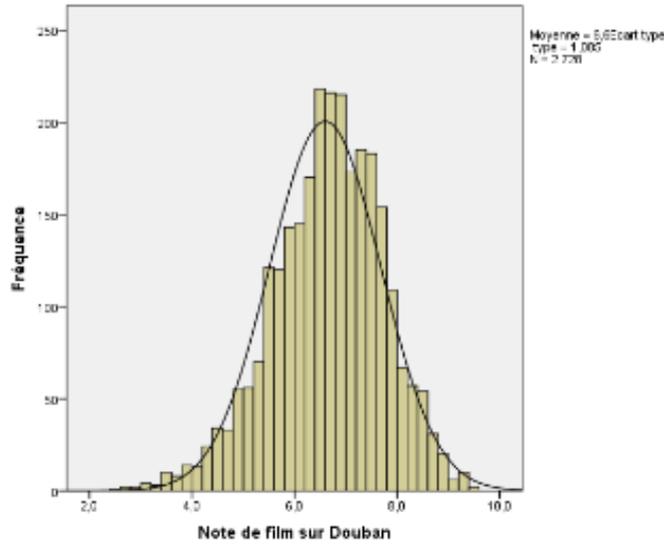


Figure 74 La réparation des notes sur Douban

IX.5.2 Les films ayant les notations différentes sur les deux plateformes

Les notes des mêmes films sur Douban (RSN chinois) et sur Allociné (RSN français) montrent une corrélation positive forte ($r=0,783$) (Tableau 8). Ainsi, l’hypothèse 6c est confirmée. Cela signifie que les utilisateurs de ces deux plateformes partagent également les mêmes normes pour évaluer le film. En tenant compte de deux études précédentes, nous arguons que les utilisateurs de ces quatre plateformes sont tous influencés par la norme explicite du groupe : « de noter le film en fonction de la qualité du film ».

		Douban	Allociné
Notes de Douban	Corrélation de Pearson	1	,783**
	Sig. (bilatérale)		,000
	N	2728	2728
Notes d’Allociné	Corrélation de Pearson	,783**	1
	Sig. (bilatérale)	,000	
	N	2728	2728

** . La corrélation est significative au niveau 0,01 (bilatéral).

Tableau 8 La corrélation entre les notes des mêmes films sur Douban et Allociné

Malgré la différence culturelle, cette corrélation positive est plus forte que celle entre Douban et Maoyan ($r=0,522$). Effectivement, les utilisateurs de Douban et d'Allociné sont plutôt d'accord les uns avec les autres. **Cela implique que l'influence normative joue un rôle plus important que la différence culturelle.**

Si nous continuons à comparer ce résultat avec celui de la première étude, nous constatons que la corrélation positive entre eux est relativement moins forte que celle ($r=0,821$) entre les notations d'Allociné et SensCritique. Nous pouvons confirmer notre dixième hypothèse : les notations sur Allociné et sur SensCritique seraient plus positivement corrélées que les notations sur Allociné et sur Douban. Par conséquent, nous soutenons que la différence culturelle est également un facteur qui influence le comportement des utilisateurs en matière de notation des films, **mais son impact est moins puissant que celui de l'influence normative.**

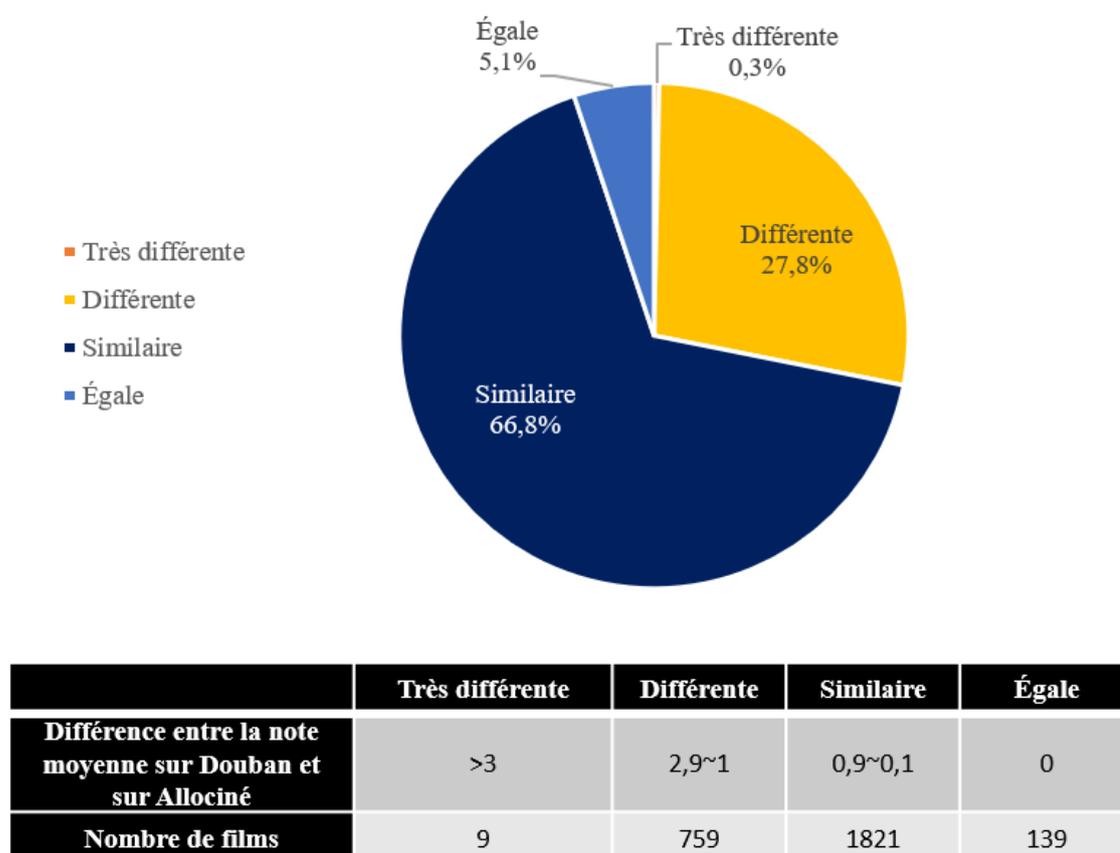


Figure 75 Répartition des films par la différence de notes des mêmes films sur Douban et sur Allociné

En comparant directement leurs notes sur les deux sites, nous avons identifié les films qui ont reçu des notes différentes sur Douban et Allociné. Nous constatons que la plupart des films

partagent certains points en commun. Nous constatons que la plupart de films (soit 71,9%) ont obtenu des notations similaires ou égales. Seulement 28,1% des films ont reçu des notations différentes ou très différentes (Figure 75).

Bien que globalement, les notes des mêmes films sur les deux RSN sont assez similaires, nous constatons qu'il y a des films de certains genres qui sont préférés davantage sur l'un que sur l'autre. En effet, les films d'arts martiaux, de comédie musicale et du sport sont plus favorisés sur Allociné que sur Douban (Figure 76). Par contre, les films d'épouvante-horreur, de comédie et documentaire sont moins appréciés par les utilisateurs d'Allociné que ceux de Douban. Nous constatons également que les utilisateurs des deux RSN notent les films français de manière relativement différente (Figure 77).

En résumé, le résultat montre que les notations des mêmes films sur Douban et Allociné sont similaires avec une forte corrélation positive. Il signifie que les utilisateurs de Douban et d'Allociné partagent des normes assez similaires pour noter le film. Toutefois, les différences existent parmi les utilisateurs de ces deux plateformes. Le résultat a révélé que les films de certains genres sont plus préférés sur une plateforme que sur l'autre. Cela implique que certaines normes implicites ont l'impact sur les utilisateurs. **En comparant les résultats avec ceux des deux études précédentes, nous considérons que les normes de groupe ont un impact significatif sur le comportement des utilisateurs de RSN et que la différence culturelle est également un facteur influençant le comportement des utilisateurs en matière de notation des films, mais que son impact est moins puissant que celui de l'influence normative.**

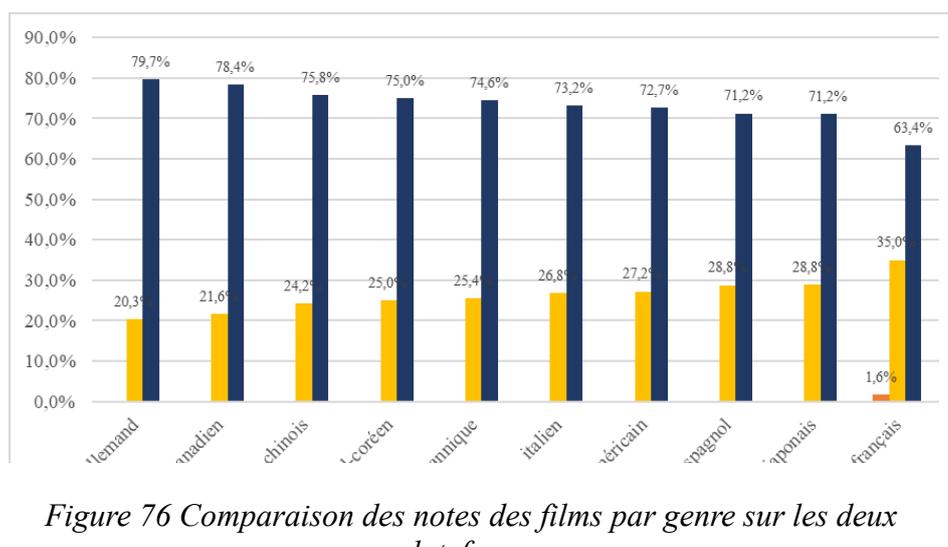


Figure 76 Comparaison des notes des films par genre sur les deux plateformes

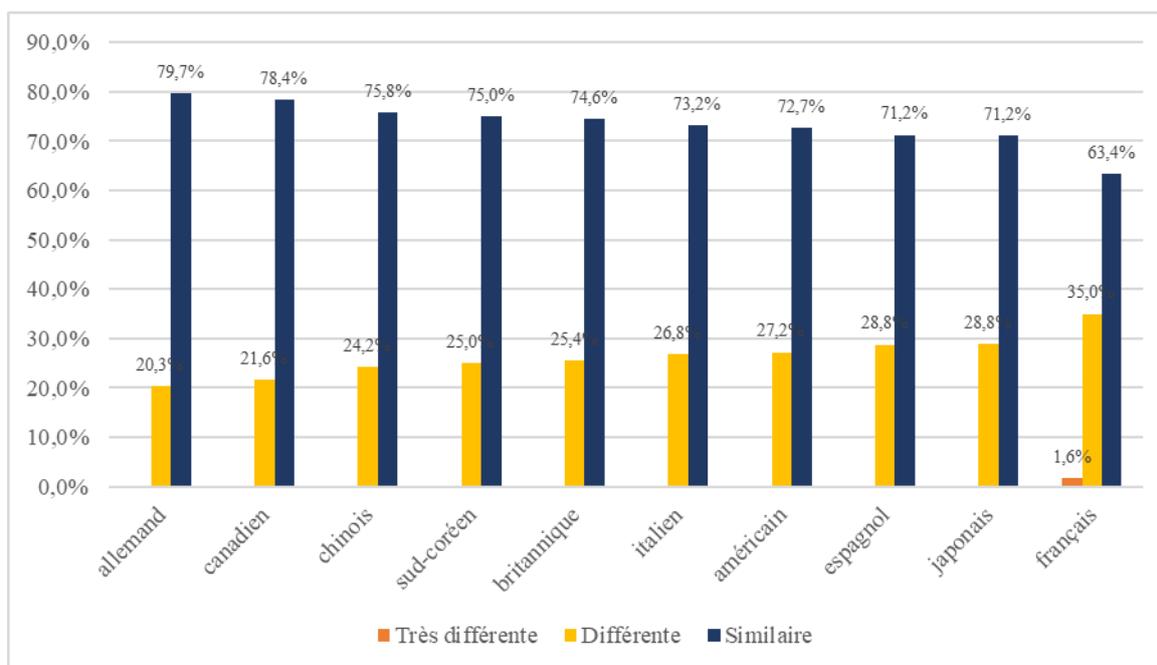


Figure 77 Comparaison des notes des films des dix pays ayant le plus grand nombre de films sur les deux plateformes

IX.6 Conclusion

Dans les trois études comparatives précédentes, nous avons comparé les notations de films attribuées par les utilisateurs de deux RSN de la même culture, par les utilisateurs d'un RSN et d'un non-RSN de la même culture, et par les utilisateurs de deux RSN de deux cultures différentes. Les résultats montrent qu'en général, les notations des mêmes films sur les 4 plateformes montrent des corrélations positives. Cela veut dire que la norme « de noter le film en fonction de la qualité du film » influence plus ou moins les utilisateurs de ces 4 plateformes.

Particulièrement, nous remarquons que, malgré la différence culturelle, les notations sur **Douban et Allociné montrent une corrélation positive ($r=0,783$) plus forte que la corrélation ($r=0,552$) entre les notations sur Douban et Maoyan. Cela implique que la norme du groupe a une influence plus puissante dans les RSN. Nous en concluons que l'influence normative a un impact très important sur le comportement de notation dans les RSN.**

De plus, les notations de films réalisés par les utilisateurs de deux RSN de la même culture (de SensCritique et d'Allociné) ont montré une très forte corrélation ($r=0,821$) qui est supérieure

à la corrélation des notations sur Douban et sur Allociné. Cela indique que la différence culturelle est également un facteur important qui influence le comportement des utilisateurs en matière de notation des films. Cependant, son influence sur le comportement des utilisateurs est inférieure à celle de la norme du groupe.

En outre, nous avons constaté que chaque plateforme favorise des films avec certaines caractéristiques spécifiques. Les films expérimentaux, documentaires, les court-métrages sont plus appréciés sur Douban. Les films d'art martiaux sont mieux notés sur Allociné que sur Douban. Les utilisateurs de Maoyan favorisent plutôt les films chinois et ceux de catastrophes. Cela implique que, sauf la norme explicite telle que « de noter le film en fonction de la qualité du film », chaque plateforme dispose respectivement de certaines normes implicites qui influencent également le comportement de notations.

Chapitre X Interprétation des résultats

X.1 Deux critères pour évaluer un produit culturel

Il est extrêmement important pour le SR d'obtenir des rétroactions de la part des utilisateurs. La rétroaction explicite telle que l'évaluation de l'utilisateur est considérée comme la rétroaction la plus directe et la plus efficace. En effet, pour les SR, les évaluations des utilisateurs sont une base importante pour déduire les préférences des utilisateurs et faire des recommandations précises. Cependant, le fait est que la plupart des problèmes auxquels de nombreux SR sont confrontés sont dus à un manque de données, tels que le problème du démarrage à froid, celui du nouvel utilisateur et celui de nouveau produit. En fait, de nombreuses recherches sont consacrées à la résolution de ces problèmes.

Cependant, notre étude a révélé que l'interprétation des évaluations des utilisateurs est également un problème qui doit être pris en compte lorsqu'il s'agit de recommandations du film. Contrairement à d'autres types de produits (comme le vêtement, l'appareil ménager ou l'automobile etc.) qui peuvent être évalués en fonction de l'apparence, de la couleur et des paramètres techniques etc., l'évaluation d'un film ne semble pas avoir des critères objectifs clairs. En fait, la notation d'un film est souvent considérée comme une évaluation subjective du film par l'utilisateur, c'est pourquoi des sites comme Netflix et MovieLens demandent aux utilisateurs de noter les films en fonction de leur propre appréciation. Cependant, lorsque nous examinons des RSN comme Douban, Maoyan, Allociné et SensCritique, ils fournissent un autre critère plus objectif en demandant à leurs utilisateurs de noter les films en fonction de la qualité du film. Nous observons donc deux critères différents pour évaluer les films.

Pourtant, de nombreux SR appliqués aux sites pour les cinéphiles ne se rendent pas compte de ce problème. Nous constatons qu'ils utilisent simultanément deux critères de notation différentes et les confondent : d'une part, ils ont tendance à encourager les utilisateurs à noter les films en fonction de leur qualité, et d'autre part, ils interprètent les notes obtenues simplement comme les préférences des utilisateurs pour les films. Ainsi, les notes élevées (5 étoiles) qui représentaient à l'origine des « chefs-d'œuvre » lorsque les utilisateurs évaluent les films sont interprétées par le SR comme la preuve que les utilisateurs aimaient beaucoup les

films. Basé sur ces interprétations, le SR a fourni des recommandations telles que « ce qui ont aimé ce film aimeraient également celui-ci ». Il semble que le SR confond les deux critères en supposant que les films d'excellente qualité sont forcément appréciés par les utilisateurs.

Cependant, les résultats de notre étude montrent que cette logique est erronée. En fait, notre enquête auprès des utilisateurs de Douban confirme déjà que les deux critères coexistent et qu'ils ne coïncident pas toujours. La majorité des répondants ont déclaré qu'ils pouvaient percevoir les deux critères (objectif et subjectif) lors de l'évaluation des films et que les deux critères ne s'accordent pas toujours. Pour eux, il y a des films d'excellente qualité qu'ils n'aimaient pas et des films de mauvaise qualité qu'ils appréciaient.

La coexistence de ces deux critères qui ne s'accordent pas toujours, peut entraîner une confusion lors de l'évaluation des films, ce qui peut également rendre difficile l'interprétation des notations des utilisateurs. Les résultats de notre enquête confirment également que tous les utilisateurs de Douban n'utilisent pas un seul critère pour évaluer les films, et que chaque utilisateur emploie différents critères lorsqu'il note des films différents. Par conséquent, les évaluations des utilisateurs deviennent difficiles à interpréter. Lorsqu'un utilisateur évalue favorablement un film, nous ne pouvons pas être sûrs que c'est parce que l'utilisateur aime le film ou parce qu'il pense que le film est de bonne qualité. Cela est confirmé par les résultats de notre enquête dans laquelle plus de la moitié des répondants ont déclaré qu'ils donnaient de bonnes notes à la fois aux films de mauvaise qualité qu'ils aimaient et aux films d'excellente qualité qu'ils n'aimaient pas.

Cette confusion dans l'interprétation des notations due à l'incertitude des critères d'évaluation peut brouiller le jugement du SR et affecter la précision ses recommandations. En fait, ce problème existe non seulement dans le domaine de recommandation pour le film, mais **aussi dans toutes les recommandations pour les produits culturels et même pour toutes les recommandations pour les produits expérientiels**. Pour ces produits, l'excellente qualité ne garantit pas qu'ils seront définitivement aimés par les utilisateurs. Cela signifie que les SR concernant les produits culturels doivent distinguer le critère objectif et le critère subjectif. Une solution possible est de fournir aux utilisateurs deux systèmes d'évaluation différents : d'une part, on demande aux utilisateurs de noter un film en fonction de sa qualité, et d'autre part, on leur demande s'ils souhaitent recevoir des recommandations pour des films similaires à celui-ci. La première note peut aider les autres utilisateurs à déduire la qualité du film, tandis que la seconde peut aider le SR à obtenir une image plus précise des préférences personnelles de

l'utilisateur.

Outre les deux critères différents qui coexistent, l'influence de la norme de groupe sur le comportement des utilisateurs peut également entraver le bon fonctionnement du SR. Dans la section suivante, nous l'abordons.

X.2 L'influence normative et culturelle

Les résultats suggèrent également que les influences normatives et sociales ont un impact sur le comportement de notation sur les RSN. En fait, le critère objectif (de noter des films en fonction de leur qualité) dont nous avons parlé est l'une des normes de groupe les plus fondamentales pour ces RSN. Les quatre plateformes de notre étude ont toutes cette norme de groupe. Nous constatons que les notations de ces quatre plateformes sont corrélées positivement. Parmi eux, les notations de trois RSN (Douban, Allociné et SensCritique) où la norme de groupe a une plus grande influence sur les comportements des utilisateurs montrent une corrélation positive ($r=0,783$ et $r=0,821$) plus forte que celle ($r=0,552$) d'un RSN (Douban) et d'un non-RSN (Maoyan). Cela démontre l'influence de la norme du groupe sur le comportement des utilisateurs en matière de notation des films. Le modèle SIDE s'applique également aux RSN pour les cinéphiles.

En plus de la norme du groupe, la différence culturelle est également un facteur important qui influence le comportement des utilisateurs. Nous constatons que les notations de deux RSN (Allociné et SensCritique) issus de la même culture montrent une corrélation positive très forte ($r=0,821$). En revanche, la corrélation positive entre les notations des deux RSN (Allociné et Douban) de cultures différentes est relativement faible ($r=0,786$). Cependant, l'effet de la différence culturelle sur le comportement des utilisateurs est clairement inférieure à l'effet des normes du groupe sur les utilisateurs. En effet, on observe que les notations sur un RSN (Douban) et un non RSN (Maoyan) d'une même culture, bien qu'elles soient également corrélées positivement, sont beaucoup moins positivement corrélées ($r=0,552$) que celles de deux RSN (Allociné et Douban) de deux cultures différentes.

Sans aucun doute, les normes de groupe qui existent sur RSN pour les cinéphiles entravent également le fonctionnement du SR, car les notations des utilisateurs obtenues par le SR ne reflètent pas entièrement la véritable opinion de l'utilisateur, mais plutôt un point de vue plus

conforme aux normes de groupe. En fait, notre enquête auprès des utilisateurs de Douban le confirme. La majorité des répondants déclarent qu'ils sont hésitants lors de notation si leur opinion est en désaccord avec les autres membres. La plupart des répondants donnent une note plus conforme aux opinions des autres malgré leur propre avis. Évidemment, les recommandations basées sur ces notations ne correspondent pas vraiment aux préférences personnelles de l'utilisateur.

La solution consiste à minimiser l'impact de l'opinion des autres et de la norme de groupe sur les utilisateurs. Les RSN pour les cinéphiles actuels permettent toujours aux utilisateurs de noter un film directement sur la page de ce film qui affiche à la fois la note moyenne de ce film et les commentaires des autres membres en même temps. Cela renforce sans aucun doute l'impact de l'opinion des autres utilisateurs sur lui. Par conséquent, une solution possible consiste à isoler l'utilisateur des autres membres en n'affichant pas les notes et les commentaires d'autrui lorsqu'il note des films.

Dans la section précédente, nous avons mentionné qu'une façon d'aborder la coexistence de deux critères différents est de fournir aux utilisateurs deux systèmes d'évaluation. L'un d'eux demande à l'utilisateur s'il souhaite continuer à recevoir des recommandations comme ce film à l'avenir. En fait, nous pouvons la demander quelques jours après que l'utilisateur a noté un film en fonction de la qualité du film. À ce moment-là, l'influence des opinions des autres utilisateurs sur les utilisateurs sera considérablement réduite.

X.3 Le RSN et le SR : désindividuation et personnalisation

Une série d'études a montré que l'anonymat dans la communication virtuelle favorise et renforce l'influence des facteurs sociaux au sein du groupe et que l'interaction de la CMC (*Computer Mediated Communication*) est fortement influencée par les normes de groupe dans un média où l'utilisateur est relativement anonyme ou désindividualisé (Spears, Lea et Lee, 1990; Lea et Spears, 1991; Postmes, Spears et Lea, 1999; Postmes *et al.*, 2001a). Selon la théorie de SIDE, la désindividuation n'est pas une perte de soi, mais la transformation de soi d'une identité personnelle en une identité sociale. Il en résulte le respect des normes du groupe. Les résultats de nos études démontrent que le modèle SIDE est également applicable sur les RSN pour les cinéphiles. Cela signifie que les normes du groupe peuvent avoir un impact assez important sur le comportement des utilisateurs de RSN. Toutefois, cela affectera inévitablement

le fonctionnement d'un SR conçu pour fournir des recommandations personnalisées.

Dans notre enquête auprès des utilisateurs de Douban, la majorité des répondants ont admis que leur évaluation était influencée par l'opinion des autres utilisateurs. Les trois études comparatives que nous avons menées confirment également que les normes de groupe peuvent avoir une influence significative sur le comportement de notation des utilisateurs. Cela signifie que les utilisateurs ont tendance à donner des notes conformes aux normes du groupe en dépit de leur propre opinion. Basé sur ces notes, le SR proposera naturellement des recommandations correspondant à ce qui conforme aux normes du groupe. Au fur et à mesure que les utilisateurs continueront à recevoir telles recommandations, ils renforceront leur respect aux normes du groupe et noteront davantage de films selon les normes du groupe. **Dans un tel cycle, le SR ne peut plus proposer des recommandations personnalisées.** Cela détruit évidemment l'intention initiale de la conception du SR.

Le SR est conçu pour aider aux utilisateurs à découvrir ceux qu'ils ne connaissent pas mais qui pourraient les intéresser. Mais avec l'intervention de la norme du groupe sur les RSN, les utilisateurs ne fourniront plus des rétroactions personnalisées qui se conforment à leurs préférences personnelles, mais plutôt celles qui se conforment aux normes du groupe. Ce qui signifie que le SR qui est conçu pour aider les utilisateurs à découvrir **est devenu un outil pour renforcer leur identité de groupe** ainsi que leur respect aux normes du groupe. À long terme, le SR peut devenir plutôt une entrave qui empêche les utilisateurs à découvrir.

Cependant, avec le développement rapide de l'Internet, les utilisateurs sont confrontés à une grande quantité d'informations chaque jour et le SR est devenu un outil essentiel pour résoudre le problème de la surcharge d'informations. L'utilisation de SR sur les RSN est donc inévitable. Mais il est important d'être conscient des normes de groupe qui existent sur les RSN, et d'éviter leurs influences possibles sur le SR.

CONCLUSION GENERALE

Dans cette étude, nous explorons l'application du SR personnalisé dans les RSN, plus précisément dans les RSN pour les cinéphiles. Nous constatons que le SR appliqué dans les RSN présente deux caractéristiques contradictoires : la personnalisation et la désindividuation. D'une part, le SR cherche à fournir aux utilisateurs des recommandations personnalisées qui correspondent à leurs préférences personnelles. Cela implique qu'il est primordial que le SR puisse obtenir des rétroactions qui reflètent les préférences réelles et personnelles des utilisateurs. D'autre part, les RSN pour les cinéphiles ont tendance à conduire les utilisateurs dans un état de la désindividuation où leurs pensées et leurs comportements sont fortement influencés par les normes du groupe et où ils se comportent conformément aux normes du groupe dans lequel ils se trouvent, de sorte que la plupart de leurs rétroactions sont souvent conformes aux normes du groupe plutôt que le reflet de leur personnalité. Basé sur de telles rétroactions, le SR fournira plutôt des recommandations conformes aux normes du groupe que des recommandations qui correspondent aux préférences personnelles des utilisateurs. De telles recommandations non seulement vont à l'encontre de l'objectif du SR mais aussi renforcent encore le respect de l'utilisateur aux normes du groupe, en rendant les recommandations personnalisées impossibles.

Cela veut dire que la désindividuation existant dans les RSN pour les cinéphiles affecte le comportement de notation de leurs utilisateurs et qui interfère donc le fonctionnement du SR personnalisé. Selon le modèle SIDE, l'anonymat et les interactions entre les utilisateurs dans les plateformes numériques renforcent l'adhésion des utilisateurs aux normes du groupe et influencent leurs pensées et leurs comportements. Par conséquent, nous nous sommes concentré sur la question de vérifier s'il existe des normes du groupe (explicites et implicites) dans les RSN pour les cinéphiles qui peuvent influencer le comportement des utilisateurs en matière de notation du film.

L'observation des quatre plateformes concernant notre étude a révélé que les RSN pour les cinéphiles partagent tous une norme explicite de groupe commune, c'est-à-dire que les utilisateurs sont encouragés à donner une note objective au film en fonction de sa qualité. Nous avons constaté que ce critère objectif était incompatible avec le critère subjectif que demande le SR aux utilisateurs (de noter les films en fonction de leurs préférences personnelles). En fait, nos résultats démontrent la coexistence de ces deux critères : objectif et subjectif. Les

utilisateurs peuvent évaluer les films en fonction de la qualité ou de leurs préférences personnelles. Cependant, ces deux critères ne sont pas toujours convergents pour la plupart des utilisateurs. Les utilisateurs pourraient ne pas aimer des films qu'ils jugent de bonne qualité et ils pourraient aimer des films qu'ils jugent de moindre qualité. Lorsque ces deux critères ne sont pas cohérents, le choix du critère affectera la notation du film.

Les résultats montrent que la norme explicite du groupe « de noter le film en fonction de sa qualité » peut avoir un impact sur le comportement des utilisateurs. Cela explique pourquoi la plupart des utilisateurs ont déclaré qu'ils avaient donné de bonnes notes aux films dont la qualité qu'ils jugeaient excellente malgré le fait qu'ils ne les aimaient pas personnellement. Néanmoins, ils ont avoué également qu'ils insistaient pour donner de bonnes notes aux films qu'ils aiment personnellement même s'ils les jugeaient de moindre qualité. Cela conduit à une ambiguïté de l'interprétation dans la notation du film. Si un utilisateur donne une bonne note à un film, on ne sait pas si c'est parce qu'il aime ce film ou parce que la qualité de ce film est excellente pour lui. Cette ambiguïté de la signification de la notation du film empêchera le SR de fournir des recommandations pertinentes et personnalisées. Les utilisateurs des RSN pour les cinéphiles risqueront de recommander de nombreux films « d'excellente qualité » qu'ils n'aiment pas, s'ils respectent la norme du groupe « de noter le film en fonction de la qualité du film ». Alors, les recommandations ne seront donc pas différentes des recommandations non personnalisées telles que la liste des « 250 meilleurs films ».

Cette ambiguïté causée par la coexistence du critère objectif et subjectif n'existe pas seulement dans le domaine de la recommandation de film. En fait, les deux critères coexistent dans les domaines de la recommandation de tous les biens d'expérience (Nelson, 1970, 1974), tels que le livre, la musique, le restaurant etc. Cette coexistence conduit à l'ambiguïté de l'interprétation de la notation et ensuite nuit aux performances du SR personnalisé. Malheureusement, cette ambiguïté causée par la coexistence des deux critères n'attire pas suffisamment l'attention. Au contraire, les deux critères sont souvent confondus dans la pratique.

Nous montrons qu'il est nécessaire de faire la distinction entre ces deux critères dans le domaine de recommandation des biens d'expérience. Une solution possible est d'inviter les utilisateurs à noter deux fois le film en fonction de deux critères différents. Les utilisateurs pourraient, d'abord, noter le film en fonction du critère objectif. Ensuite, ils pourraient également signaler au système s'ils veulent voir davantage de films similaires de celui-ci ou pas.

Bien que les répondants du Douban dans notre première enquête aient déclaré qu'ils ne notaient pas des films exactement selon la norme explicite « de noter le film en fonction de la qualité », notre deuxième étude comparative démontre que cette norme a une influence assez puissante sur le comportement de notation des utilisateurs car les notations de quatre plateformes disposant de cette même norme montrent les corrélations positives. Cela signifie que l'influence normative est un facteur qui affecte le comportement de notation dans les RSN pour les cinéphiles. En comparant les notations des mêmes films sur 4 plateformes pour les cinéphiles, nous remarquons que les notations sur les trois RSN (Douban, Allociné et SensCritique) montrent des corrélations positives plus fortes. Cela implique que l'influence de cette norme est plus puissante sur les utilisateurs de RSN. Cette constatation correspond à la théorie de SIDE (Spears et Lea, 1992, 1994; Postmes et Spears, 1998) qui indique que l'influence normative peut parfois exercer un fort impact sur le comportement de l'utilisateur dans les groupes de CMC. Dans notre cas, les utilisateurs de RSN pour les cinéphiles ont tendance à noter le film selon le critère objectif au lieu de leurs propres préférences.

Effectivement, dans les RSN où les utilisateurs sont anonymes, les utilisateurs sont plus ou moins objets de désindividuation. Suite à l'interaction, les utilisateurs développent un sens d'identité de groupe significatif. Ce sens de l'identité de groupe renforce ensuite le respect des normes de groupe qui définissent la façon dont les membres du groupe doivent penser, ressentir ou se comporter. Cela veut dire qu'une notation d'un film reflète l'opinion qu'un utilisateur pense être conforme aux normes du groupe au lieu de sa propre opinion. Selon les résultats, lorsque les utilisateurs notent le film, ils sont hésitants si leur opinion n'est pas en accord avec celle des autres. La plupart des utilisateurs déclarent qu'ils donnent une note plus conforme aux opinions des autres. En réalité, nous avons constaté que chaque plateforme favorise des films avec certaines caractéristiques spécifiques. Ces caractéristiques peuvent être le genre, la nationalité, le thème ou bien d'autres. Cela prouve l'existence de certaines normes implicites qui définit le comportement de notation dans les RSN.

En outre, la différence culturelle peut également être un facteur qui influence le comportement des utilisateurs en matière de notation du film. Par conséquent, nous avons également inclus la différence culturelle dans notre recherche. En comparant les notations des films de 3 plateformes issues de la Chine (Douban) et de la France (Allociné et SensCritique), nous avons constaté une forte corrélation positive entre les notations de films sur deux RSN ayant le même contexte culturel (Allociné et SensCritique). En fait, nous considérons que la culture peut être considérée également comme un ensemble de normes du groupe. Cela explique

aussi pourquoi la corrélation positive entre les notations des mêmes films sur un RSN (Douban) et un non-RSN (Maoyan) issue d'une même culture est beaucoup plus faible que la corrélation positive entre les notations des mêmes films sur deux RSN de deux cultures différentes (Allociné et Douban).

En fait, les résultats de notre recherche suggèrent que les normes du groupe influencent fortement le comportement de notation dans les RSN pour les cinéphiles. Pourtant, l'influence normative nuit gravement la performance du SR qui cherche à fournir des recommandations personnalisées. Car, d'une part, le SR ne peut pas comprendre les véritables pensées des utilisateurs en se basant sur leurs notations. Les recommandations que le SR proposent sont plutôt des films favorisés par les normes du groupe. D'autre part, au fur et à mesure que les utilisateurs continuent à recevoir les recommandations conformes aux normes du groupe, leurs préférences sont changées et deviennent de plus en plus comme les autres. La recommandation personnalisée n'aura aucun sens, si les goûts de tout le monde deviennent les mêmes. Il faudrait individualiser les utilisateurs lorsqu'ils sont invités à noter le film en fonction de leur préférence, par exemple, et ne pas leur montrer les évaluations des autres ou de ne pas publier leurs notations qui sont en fonction du critère subjectif.

Il existe plusieurs limites à cette étude. Premièrement, nous n'avons pas pu terminer l'enquête par questionnaire auprès des utilisateurs d'Allociné faute de suffisamment de réponses. Cela nous empêche de comprendre les véritables pensées des utilisateurs français concernant leur comportement de notation. Nous aurions pu mieux comprendre l'impact de la différence culturelle sur le comportement de notation des utilisateurs en comparant avec le résultat de l'enquête auprès des utilisateurs de Douban. Deuxièmement, en raison du nombre insuffisant de films notés en commun par les utilisateurs des quatre plateformes, nous ne pouvons comparer les notations des quatre plateformes que deux par deux, plutôt que de comparer les notations des 4 plateformes simultanément. Troisièmement, nous nous sommes concentrés sur une norme explicite. Mais pour les normes implicites, nous n'avons examiné que superficiellement l'influence des opinions d'autrui et les préférences des utilisateurs pour les genres et les nationalités des films. Il est certain que ces plateformes doivent avoir davantage de normes implicites qui influencent les préférences et les notations des utilisateurs, telles que des acteurs et des réalisateurs spécifiques ou l'expérience qu'un film présente au spectateur. Si l'on pouvait identifier le plus grand nombre possible de normes explicites et implicites, il serait plus utile d'aider à libérer les utilisateurs de l'influence normative lors de la notation afin qu'ils puissent maximiser l'affichage de leurs préférences et opinions dans leurs évaluations.

En résumé, le résultat de notre recherche montre qu'il existe deux critères (objectif et subjectif) pour évaluer les biens d'expérience et que l'influence normative joue un rôle très important dans les RSN qui affecte le comportement de notation. Cette coexistence de deux critères différents et l'influence normative empêchent le SR de fournir les recommandations pertinentes et personnalisées. En outre, les utilisateurs des RSN sont affectés par la désindividuation dans leur comportement de notation, ce qui interfère finalement avec le fonctionnement des SR personnalisés. Cette recherche s'est basée sur l'enquête par questionnaire et trois études comparatives après une analyse rigoureuse de quatre plateformes. Des recherches complémentaires, surtout des études expérimentales sont encore nécessaires à une meilleure compréhension de ces enjeux que sont la désindividuation et la recommandation personnalisée en ligne.

BIBLIOGRAPHIE

Abrams, D. et Hogg, M. A. (2004) « Metatheory: Lessons from social identity research », *Personality and social psychology review*, 8(2), p. 98-106.

Abrams, D. et Hogg, M. A. (2006) *Social identifications: A social psychology of intergroup relations and group processes*. Routledge.

Adomavicius, G. et Tuzhilin, A. (2005) « Toward the Next Generation of Recommender Systems: a Survey of the State of the Art and Possible Extensions », *IEEE Transactions on Knowledge and Data Engineering*, 17(6), p. 734-749. doi: 10.1109/TKDE.2005.99.

Adorno, T. W. *et al.* (1950) *The authoritarian personality*. New York: Harper.

Aggarwal, C. C. *et al.* (1999) « 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*, p. 201-212. doi: 10.1145/312129.312230.

Aggarwal, C. C. (2016) *Recommender Systems*, Springer. Yorktown Heights, NY, USA.

Aggarwal, C. C. et Parthasarathy, S. (2001) « Mining Massively Incomplete Data Sets by Conceptual Reconstruction », in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery (KDD '01), p. 227-232. doi: 10.1145/502512.502543.

Agrawal, R. et Srikant, R. (1994) « Fast algorithms for mining association rules », in *Proc. 20th int. conf. very large data bases, VLDB*, p. 487-499.

Ahn, J. W. *et al.* (2007) « Open user profiles for adaptive news systems: help or harm? », in *Proceedings of the 16th international conference on World Wide Web*. New York, NY, USA: ACM, p. 11-20. doi: 10.1145/1242572.1242575.

Aichner, T. et Jacob, F. (2015) « Measuring the degree of corporate social media use », *International Journal of Market Research*, 57(2), p. 257-275. doi: 10.2501/IJMR-2015-018.

Alspector, J., Koicz, A. et Karunanithi, N. (1997) « Feature-based and clique-based user models for movie selection: A comparative study », *User Modeling and User-Adapted Interaction*, 7(4), p. 279-304.

Altemeyer, B. (1998) « The other “authoritarian personality” », in *Advances in*

experimental social psychology. Elsevier, p. 47-92.

Amatriain, X. *et al.* (2011) « Data mining methods for recommender systems », in *Recommender Systems Handbook*. New York: Springer, p. 39-71. doi: 10.1007/978-0-387-85820-3.

Amatriain, X., Pujol, J. M. et Oliver, N. (2009) « I like it... i like it not: Evaluating user ratings noise in recommender systems », in *Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization: Formerly UM and AH*, p. 247-258. doi: 10.1007/978-3-642-02247-0_24.

Anand, S. S. et Mobasher, B. (2003) « Intelligent techniques for web personalization », in *Proceedings of the 2003 international conference on Intelligent Techniques for Web Personalization*. Springer-Verlag, p. 1-36.

Asnicar, F. et Tasso, C. (1997) « IfWeb: a prototype of user model-based intelligent agent for document filtering and navigation in the World Wide Web », in Jameson, Anthony, Paris, Cecile, Tasso, C. (éd.) *User Modeling: Proceedings of the Sixth International Conference UM97 Chia Laguna, Sardinia, Italy June 2--5 1997*. Vienna: Springer Vienna, p. 2-5.

Autin, F. (2010) « La théorie de l'identité sociale de Tajfel et Turner », *Préjugés & Stéréotypes*. Disponible sur: www.prejuges-stereotypes.net/autinIdentiteSociale.pdf.

Barnes, J. A. (1954) « Class and Committees in a Norwegian Island Parish », *Human Relations*, 7(1), p. 39-58. doi: 10.1177/07399863870092005.

Barreto, M. et Ellemers, N. (2000) « You can't always do what you want: Social identity and self-presentational determinants of the choice to work for a low-status group », *Personality and Social Psychology Bulletin*, 26(8), p. 891-906.

Barreto, M. et Ellemers, N. (2002) « The impact of anonymity and group identification on progroup behavior in computer-mediated groups », *Small Group Research*, 33(5), p. 590-610. doi: 10.1177/104649602237680.

Bates, M. E. (1994) « Electronic clipping services: A new life for SDIs. », *Online*, 18(4), p. 43-50.

Béchet, N. (2012) « Etat de l'art sur les systèmes de recommandation », *Projet AxIS de l'INRIA, dans le cadre du projet Addictrip*. Disponible sur: <http://citeseerx.ist.psu.edu> (Consulté le: 18 novembre 2019).

Belkin, N. et Croft, W. B. (1992) « Information filtering and information retrieval: two sides of the same coin? », *Communications of the ACM*, 35(12), p. 29-38. doi: 10.1145/138859.138861.

Bell, R., Koren, Y. et Volinsky, C. (2007) « Modeling Relationships at Multiple Scales to Improve Accuracy of Large Recommender Systems », in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM (KDD '07), p. 95-104. doi: 10.1145/1281192.1281206.

Bell, R. M. et Yehuda, K. (2007) « Scalable collaborative filtering with jointly derived neighborhood interpolation weights », *Proceedings - IEEE International Conference on Data Mining, ICDM*, p. 43-52. doi: 10.1109/ICDM.2007.90.

Bennett, J. et Lanning, S. (2007) « The Netflix prize », in *Proceedings of KDD cup and workshop*, p. 35. doi: 10.1145/1562764.1562769.

Berkowitz, L. (1962) « Aggression: A social psychological analysis. »

Berkowitz, L. (1969) *Perspectives: simple views of aggression: an essay review*. JSTOR.

Berkowitz, L. (1974) « Some determinants of impulsive aggression: Role of mediated associations with reinforcements for aggression. », *Psychological review*, 81(2), p. 165.

Bernardes, D. *et al.* (2015) « A Social Formalism and Survey for Recommender Systems », *ACM SIGKDD Explorations Newsletter*, 16(2), p. 20-37. doi: 10.1145/2783702.2783705.

Berrut, C. et Denos, N. (2003) « Filtrage collaboratif », in *Assistance intelligente à la recherche d'informations*. Hermes Science Publications, p. 241-269. doi: 10.3166/dn.11.1-2.13-35.

Billsus, D. et Pazzani, M. J. (1998) « Learning collaborative information filters », *Proceedings of the Fifteenth International Conference on Machine Learning*, 54, p. 47. Disponible sur: <http://www.aaai.org/Papers/Workshops/1998/WS-98-08/WS98-08-005.pdf>.

Billsus, D. et Pazzani, M. J. (1999) « A hybrid user model for news story classification », in *UM99 User Modeling*. Springer, p. 99-108.

Billsus, D. et Pazzani, M. J. (2000) « User modeling for adaptive news access », *User modeling and user-adapted interaction*, 10(2-3), p. 147-180.

Blei, D. M., Ng, A. Y. et Jordan, M. I. (2003) « Latent Dirichlet allocation », *Journal of Machine Learning Research*, 3(4-5), p. 993-1022. doi: 10.1016/b978-0-12-411519-4.00006-9.

Blood, R. (2000) *Weblogs: A History And Perspective, Rebecca 'S Pocket*. Disponible sur: http://www.rebeccablood.net/essays/weblog_history.html.

Blood, R. (2002) *The Weblog Handbook: Practical Advice on Creating and Maintaining Your Blog, Library Journal*. Disponible sur: <http://www.amazon.co.uk/dp/073820756X>.

Bobadilla, J. *et al.* (2013) « Recommender systems survey », *Knowledge-Based Systems*, 46, p. 109-132. doi: 10.1016/j.knosys.2013.03.012.

Bollacker, K. D., Lawrence, S. et Giles, C. L. (1998) « CiteSeer: An autonomous web agent for automatic retrieval and identification of interesting publications », in *Proceedings of the second international conference on Autonomous agents*. ACM, p. 116-123.

Le Bon, G. (1900) *Psychologie des foules*. F. Alcan.

Boone, G. (1998) « Concept features in re: agent, an intelligent email agent », in *International Conference on Autonomous Agents: Proceedings of the second international conference on Autonomous agents*, p. 141-148.

Bouas, K. S. et Arrow, H. (1995) « The development of group identity in computer and face-to-face groups with membership change », *Computer supported cooperative work (CSCW)*, 4(2-3), p. 153-178.

Boyd, danah m. et Ellison, N. B. (2007) « Social Network Sites: Definition, History, and Scholarship », *Journal of Computer-Mediated Communication*, 13(1), p. 210-230. doi: 10.1111/j.1083-6101.2007.00393.x.

Breese, J. S., Heckerman, D. et Kadie, C. (1998) « Empirical analysis of predictive algorithms for collaborative filtering », *Proceedings of the 14th conference on Uncertainty in Artificial Intelligence*, 461(8), p. 43-52. doi: 10.1111/j.1553-2712.2011.01172.x.

Brewer, M. B. et Gardner, W. (1996) « Who is this "We"? Levels of collective identity and self representations. », *Journal of personality and social psychology*, 71(1), p. 83.

Burke, R. (2007) « Hybrid Web Recommender Systems », in Brusilovsky, P., Kobsa, A., et Nejdl, W. (éd.) *The Adaptive Web: Methods and Strategies of Web Personalization*. Berlin, Heidelberg: Springer Berlin Heidelberg, p. 377-408. doi: 10.1007/978-3-540-72079-9_12.

Butler, B. (2001) « Membership size, communication activity and sustainability: The internal dynamics of networked social structures », *Information Systems Research*, 12(4), p. 346-362.

Campbell, D. T. (1965) « Ethnocentric and other altruistic motives », in *Nebraska symposium on motivation*, p. 283-311.

Canavese, P. et Besser, H. (1994) *The future of information filtering*. Disponible sur: http://besser.tsoa.nyu.edu/impact/s94/students/paul/paul_final.html.

Canny, J. (2002) « Collaborative filtering with privacy », in *Proceedings - IEEE Symposium on Security and Privacy*, p. 45-57. doi: 10.1109/SECPRI.2002.1004361.

Cavazza, F. (2019) *Panorama des médias sociaux 2019 – FredCavazza.net*. Disponible sur: <https://fredcavazza.net/2019/05/12/panorama-des-medias-sociaux-2019/> (Consulté le: 7 janvier 2020).

Chaimbault, T. (2007) « Web 2.0 : l'avenir du web ? », p. 83.

Chakrabarti, S. (2002) *Mining the Web: Discovering knowledge from hypertext data*. Elsevier.

Chen, L. et Sycara, K. (1998) « WebMate : A Personal Agent for Browsing and Searching », *Knowledge Acquisition*, p. 132-139. doi: 10.1145/280765.280789.

Chiu, C., Ip, C. et Silverman, A. (2012) « Understanding social media in China », *McKinsey Quarterly*, (2), p. 78-81.

Christopherson, K. M. (2007) « The positive and negative implications of anonymity in Internet social interactions: "On the Internet, Nobody Knows You're a Dog" », *Computers in Human Behavior*, 23(6), p. 3038-3056.

Chuang, E. et Sher, D. (1993) « χ^2 test for feature detection », *Pattern Recognition*, 26(11), p. 1673-1681.

Cormode, G. et Krishnamurthy, B. (2008) « Key differences Web 1.0 and Web 2.0 », *First Monday*, 13(6). doi: 10.5210/fm.v13i6.2125.

Courrier, S. (2007) *Utiliser les fils RSS et Atom*. ADBS.

D'Souza, Q. (2007) « Web 2.0 ideas for educators », *A Guide to RSS and More*. Disponible sur: http://cent.uji.es/wiki/_media/seminari:100ideasweb2educators_es.pdf.

Daelemans, W., Van Den Bosch, A. et Ubler, S. (2005) « Memory-Based Language Processing », *Computational Linguistics*, 32(4).

Degemmis, M., Lops, P. et Semeraro, G. (2007) « A content-collaborative recommender

that exploits WordNet-based user profiles for neighborhood formation », *User Modeling and User-Adapted Interaction*, 17(3), p. 217-255.

Dempster, A. P., Laird, N. M. et Rubin, D. B. (1977) « Maximum likelihood from incomplete data via the EM algorithm », *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), p. 1-22.

Deshpande, M. et Karypis, G. (2004) « Item-Based Top-N Recommendation Algorithms », *ACM Trans. Inf. Syst.*, 22(1), p. 143–177. doi: 10.1145/963770.963776.

Desrosiers, C. et Karypis, G. (2011) « A Comprehensive Survey of Neighborhood-based Recommendation Methods », in Ricci, F. et al. (éd.) *Recommender Systems Handbook*. Boston, MA: Springer US, p. 107-144. doi: 10.1007/978-0-387-85820-3_4.

Dice, L. R. (1945) « Measures of the Amount of Ecologic Association Between Species », *Ecology*, 26(3), p. 297-302. doi: 10.2307/1932409.

Diener, E. (1979) « Deindividuation, self-awareness, and disinhibition », *Journal of Personality and Social Psychology*, 37(7). doi: 10.1037/0022-3514.37.7.1160.

Diener, E. (1980) « Deindividuation: The absence of self-awareness and self-regulation in group members », *The psychology of group influence*, 209242.

Diener, E. (2006) « Deindividuation: causes and consequence », *Social Behavior and Personality: an international journal*, 5(1). doi: 10.2224/sbp.1977.5.1.143.

Diener, E. et al (1975) « Effects of altered responsibility, cognitive set, and modeling on physical aggression and deindividuation », *Journal of Personality and Social Psychology*, 31(2). doi: 10.1037/h0076279.

Diener, E. et Wallbom, M. (1976) « Effects of self-awareness on antinormative behavior », *Journal of Research in Personality*, 10(1). doi: 10.1016/0092-6566(76)90088-X.

DiNucci, D. (1999) « Fragmented future », *Print*, 53(4), p. 32-33.

Dollard, J. et al. (1939) « Frustration and aggression. »

Duval, S. et Wicklund, R. A. (1972) « A theory of objective self awareness », *A theory of objective self awareness*.

Ekstrand, M. D., Riedl, J. T. et Konstan, J. A. (2011) « Collaborative Filtering Recommender Systems », *Foundations and Trends® in Human-Computer Interaction*, 4(2), p. 81-173. doi: 10.1561/11000000009.

Fayon, D. (2010) *Web 2.0 et au-delà: Nouveaux internautes: du surfeur à l'acteur*. Economica.

Feinberg, M. (2006) « An Examination of Authority in Social Classification Systems », in *Proceedings of the 17th Annual ASIS&T SIG/CR Classification Research Workshop*, p. 1-11. doi: 10.7152/acro.v17i1.12490.

Festinger, L., Pepitone, A. et Newcomb, T. (1952) « Some consequences of deindividuation in a group. », *The Journal of Abnormal and Social Psychology*, 47(2S), p. 382.

Festinger, L., Schachter, S. et Back, K. (1964) « Action et fonctionnement des “normes de groupe” », in *Éléments de sociologie générale*. Paris: Presses de Sciences Po (Hors collection), p. 207-228.

Fiedler, F. (1967) *A theory of leadership effectiveness*. New York McGraw-Hill. Disponible sur: [https://www.scirp.org/\(S\(351jmbntvnsjt1aadkposzje\)\)/reference/ReferencesPapers.aspx?ReferenceID=895085](https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=895085).

Finholt, T. et Sproull, L. S. (1990) « Electronic groups at work », *Organization Science*, 1(1), p. 41-64.

Forsé, M. (2008) « Définir et analyser les réseaux sociaux », *Informations sociales*, (3), pp. 10–19.

Fu, X., Budzik, J. et Hammond, K. J. (2000) « Mining navigation history for recommendation », in *Proceedings of the 5th international conference on Intelligent user interfaces*. ACM, p. 106-112.

Funk, T. (2009) *Web 2.0 and beyond: understanding the new online business models, trends, and technologies*. Praeger.

Gallardo-Lopez, L. (2005) *Accès à l'information par un système de filtrage collaboratif contrôlé*. Disponible sur: <https://tel.archives-ouvertes.fr/tel-01683549> (Consulté le: 18 novembre 2019).

George, T. et Merugu, S. (2005) « A scalable collaborative filtering framework based on co-clustering », in *Fifth IEEE International Conference on Data Mining (ICDM'05)*, p. 4 pp. doi: 10.1109/ICDM.2005.14.

Gergen, K., Gergen, M. et Barton, W. H. (1973) « Deviance in the dark », *Psychology*

Today, 7, p. 129-130.

Getoor, L. et Sahami, M. (1999) « Using Probabilistic Relational Models for Collaborative Filtering », in *Working Notes of the KDD Workshop on Web Usage Analysis and User Profiling*.

Gibbons, J. D. et Kendall, M. (1990) « Rank correlation methods », *Edward Arnold*.

Gillmor, S. (2009) *Rest in Peace, RSS*. Disponible sur: <https://techcrunch.com/2009/05/05/rest-in-peace-rss/>.

Goldberg, D. *et al.* (1992) « Using collaborative filtering to weave an information Tapestry », 35(12), p. 61+.

Goldberg, K. *et al.* (2001) « Eigentaste: A Constant Time Collaborative Filtering Algorithm », *Information Retrieval*, 4(2), p. 133-151. doi: 10.1023/A:1011419012209.

Golder, S. A. et Huberman, B. A. (2006) « Usage patterns of collaborative tagging systems », *Journal of Information Science*, 32(2), p. 198-208. doi: 10.1177/0165551506062337.

Golub, G. et Kahan, W. (1965) « Calculating the singular values and pseudo-inverse of a matrix », *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, 2(2), p. 205-224.

Good, N. *et al.* (1999) « Combining collaborative filtering with personal agents for better recommendations », *Proceedings of the National Conference on Artificial Intelligence*, p. 439-446.

Gordon-Murnane, L. (2006) « Social bookmarking, folksonomies, and Web 2.0 Tools », *Searcher*, 14(6), p. 26-28. Disponible sur: https://www.researchgate.net/publication/279895554_Social_bookmarking_folksonomies_and_Web_2_0_Tools (Consulté le: 5 janvier 2020).

Gordon, G. (2007) « Latent Factor Models Simple case : snapshot , no side information Simple case : snapshot , no side information ».

Grčar, M. *et al.* (2006) « kNN Versus SVM in the Collaborative Filtering Framework », in Batagelj, V. *et al.* (éd.) *Data Science and Classification*. Berlin, Heidelberg: Springer Berlin Heidelberg, p. 251-260.

Hammond, T. *et al.* (2005) « Social bookmarking tools (I): A general review », *D-Lib Magazine*, 11(4), p. 1-21. doi: 10.1045/april2005-hammond.

Han, J., Kamber, M. et Tung, A. K. H. (2001) « Spatial clustering methods in data mining »,

Geographic data mining and knowledge discovery, p. 188-217.

Hanani, U., Shapira, B. et Shoval, P. (2001) « Information filtering: Overview of issues, research and systems », *User Modeling and User-Adapted Interaction*, 11(3), p. 203-259. doi: 10.1023/A:1011196000674.

Harper, F. M. *et al.* (2005) « An Economic Model of User Rating in an Online Recommender System », in, p. 307-316. doi: 10.1007/11527886_40.

Hartigan, J. A. (1975) « Clustering algorithms ».

Hayman, S. (2007) « Folksonomies and tagging: New developments in social bookmarking », in *Ark group conference: Developing and improving classification schemes*. Disponible sur: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.138.8884> (Consulté le: 5 janvier 2020).

Herlocker, J. et Konstan, J. (1999) « An algorithmic framework for performing collaborative filtering », in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, p. 230-237. doi: 10.1145/312624.312682.

Herlocker, J., Konstan, J. A. et Riedl, J. (2002) « An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms », *Information Retrieval*, 5(4), p. 287-310. doi: 10.1023/A:1020443909834.

Herlocker, J. L. *et al.* (2004) « Evaluating collaborative filtering recommender systems », *ACM Transactions on Information Systems (TOIS)*, 22(1), p. 5-53. doi: 10.1145/963770.963772.

Herlocker, J. L. *et al.* (2017) « An algorithmic framework for performing collaborative filtering », in *ACM SIGIR Forum*. ACM New York, NY, USA, p. 227-234.

Hill, W. *et al.* (1995) « Recommending and evaluating choices in a virtual community of use », in Katz, Irvin R., Mack, Robert L., Marks, Linn, Rosson, Mary Beth, Nielsen, J. (eds. . (éd.) *Conference on Human Factors in Computing Systems - Proceedings*. Denver, Colorado: ACM Press/Addison-Wesley Publishing Co., p. 194-201. doi: 10.1145/223904.223929.

Hofmann, T. (2003) « Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis », in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*. New York, NY, USA: Association for Computing Machinery (SIGIR '03), p. 259–266. doi: 10.1145/860435.860483.

Hofmann, T. (2004) « Latent Semantic Models for Collaborative Filtering », 22(1), p. 89-115.

Hogg, M. A. (2003) « Social identity. », in *Handbook of self and identity*. New York, NY, US: The Guilford Press, p. 462-479.

Hogg, M. A. (2005) « The social identity perspective », in *The handbook of group research and practice*. Sage, p. 133-157.

Hogg, M. A. (2006) « Social identity theory », in *Contemporary social psychological theories*. Stanford University Press, p. 111-136.

Hogg, M. A. (2013) « Intergroup relations », in *Handbook of social psychology*. Springer, p. 533-561.

Hogg, M. A. et Reid, S. A. (2006) « Social identity, self-categorization, and the communication of group norms », *Communication Theory*, 16(1), p. 7-30. doi: 10.1111/j.1468-2885.2006.00003.x.

Howe, A. E. et Forbes, R. D. (2008) « Re-considering neighborhood-based collaborative filtering parameters in the context of new data », in *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, p. 1481-1482.

Ickes, W., Layden, A. et Barnes, R. (1978) « Objective self-awareness and individuation: An empirical link », *Journal of personality*, 46, p. 146-161.

Jannach, D. et al. (2010) *Recommender Systems An Introduction*. 1 edition. New York: Cambridge University Press. Disponible sur: <http://ebooks.cambridge.org/ref/id/CBO9781107415324A009>.

Jannach, D. et Friedrich, G. (2013) *Tutorial: Recommender Systems, International Joint Conference on Artificial Intelligence Beijing*.

Johnson, R. D. et Downing, L. L. (1979) « Deindividuation and valence of cues: Effects on prosocial and antisocial behavior. », *Journal of Personality and Social Psychology*, 37(9), p. 1532.

Jorgenson, D. O. et Dukes, F. O. (1976) « Deindividuation as a function of density and group membership », *Journal of Personality and Social Psychology*, 34(1). doi: 10.1037/0022-3514.34.1.24.

Kalin, R. et Marlowe, D. (1968) « The effects of intergroup competition, personal drinking

habits and frustration in intra-group cooperation », in *Proceedings of the 76th Annual Convention of the American Psychological Association (APA)*, p. 405-406.

Kantar (2019) *China Social Media Whitepaper*.

Kaplan, A. M. et Haenlein, M. (2010) « Users of the world, unite! The challenges and opportunities of Social Media », *Business Horizons*, 53(1), p. 59-68. doi: 10.1016/j.bushor.2009.09.003.

Karypis, G. (2001) « Evaluation of Item-Based Top-N Recommendation Algorithms », in *Proceedings of the Tenth International Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery (CIKM '01), p. 247–254. doi: 10.1145/502585.502627.

Kaufman, L. et Rousseeuw, P. J. (2009) *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Kaufman, L., Rousseeuw, P. J. et Dodge, Y. (1987) « Clustering by Means of Medoids in Statistical Data Analysis Based on the L_1 Norm », orth-Holland, Amsterdam.

Khan, I. et Card, H. C. (1997) « Personal adaptive Web agent: A tool for information filtering », in *Canadian Conference on Electrical and Computer Engineering*. IEEE, p. 305-308. doi: 10.1109/ccece.1997.614850.

Kim, D. J. *et al.* (2009) « Global diffusion of the internet XV: Web 2.0 technologies, principles, and applications: A conceptual framework from technology push and demand pull perspective », *Communications of the Association for Information Systems*, 24(1), p. 657-672. doi: 10.17705/1cais.02438.

Kipp, M. E. I. et Campbell, D. G. (2007) « Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices », *Proceedings of the American Society for Information Science and Technology*, 43(1), p. 1-18. doi: 10.1002/meet.14504301178.

Koh, N. S., Hu, N. et Clemons, E. K. (2010) « Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures », *Electronic Commerce Research and Applications*, 9(5), p. 374-385. doi: 10.1016/j.elerap.2010.04.001.

Konstan, J. A. *et al.* (1997) « GroupLens: Applying Collaborative Filtering to Usenet News », *Communications of the ACM*, 40(3), p. 77-87. doi: 10.1145/245108.245126.

Koren, Y. (2008) « Factorization Meets the Neighborhood: A Multifaceted Collaborative

Filtering Model », in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery (KDD '08), p. 426–434. doi: 10.1145/1401890.1401944.

Koren, Y. et Bell, R. (2015) « Advances in collaborative filtering », in *Recommender systems handbook*. Springer, p. 77-118.

Koren, Y., Bell, R. et Volinsky, C. (2009) « Matrix Factorization Techniques for Recommender Systems », *Computer*, 42(8), p. 42-49. doi: 10.1109/MC.2009.263.

Lam, X. N. *et al.* (2008) « Addressing cold-start problem in recommendation systems », in *Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication, ICUIMC-2008*, p. 208-211. doi: 10.1145/1352793.1352837.

Lambiotte, R. et Ausloos, M. (2005) « Collaborative tagging as a tripartite network », *Lecture Notes in Computer Science*, 3993, p. 1114-1117. doi: 10.1007/11758532_152.

Lang, K. (1995) « Newsweeder: Learning to filter netnews », in *Machine Learning Proceedings 1995*. Elsevier, p. 331-339.

Lea, M. et Spears, R. (1991) « Computer-mediated communication, de-individuation and group decision-making », *International Journal of Man-Machine Studies*, 34(2), p. 283-301. doi: 10.1016/0020-7373(91)90045-9.

Lea, M., Spears, R. et de Groot, D. (2001) « Knowing me, knowing you: Anonymity effects on social identity processes within groups », *Personality and Social Psychology Bulletin*, 27(5), p. 526-537.

Lee Rodgers, J. et Nicewander, W. A. (1988) « Thirteen ways to look at the correlation coefficient », *The American Statistician*, 42(1), p. 59-66.

Lewis, C. et Rieman, J. (1993) « Task-centered user interface design: A practical introduction ».

Licata, L. (2007) « La théorie de l'identité sociale et la théorie de l'autocatégoriesation : le Soi, le groupe et le changement social », *Revue électronique de Psychologie Sociale*, 1, p. 19-33. doi: 10.1002/mrdd.20019.

Lieberman, H. (1995) « Letizia: An agent that assists web browsing », *IJCAI (1)*, 1995, p. 924-929.

Lin, W. (2002) « Efficient Adaptive-Support Association Rule Mining for Recommender

Systems », p. 83-105.

Linden, G., Smith, B. et York, J. (2003) « Amazon.com Recommendations: Item-to-Item Collaborative Filtering », *{IEEE} Internet Computing*, 7(1), p. 76-80. doi: 10.1109/MIC.2003.1167344.

Van Loan, C. F. et Golub, G. H. (1983) *Matrix computations*. Johns Hopkins University Press.

Lops, P., de Gemmis, M. et Semeraro, G. (2011) « Content-based Recommender Systems: State of the Art and Trends », in *Recommender Systems Handbook*. Springer US, p. 73-105. doi: 10.1007/978-0-387-85820-3_3.

Luhn, H. P. (1958) « The Automatic Creation of Literature Abstracts », *IBM Journal of Research and Development*, 2(2), p. 159-165. doi: 10.1147/rd.22.0159.

Macmanus, R. (2005) *Web 2.0 Definition and Tagging*. Disponible sur: https://readwrite.com/2005/02/01/web_20_definiti/.

MacQueen, J. (1967) « Some methods for classification and analysis of multivariate observations », in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA, p. 281-297.

Mak, H., Koprinska, I. et Poon, J. (2003) « Intimate: A web-based movie recommender using text categorization », in *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*. IEEE, p. 602-605.

Malone, T. *et al.* (1987) « Intelligent information sharing systems », *Communications of the ACM*, 30(5)(8), p. 390-402. doi: 10.1108/eb028253.

Manning, C. D., Raghavan, P. et Schütze, H. (2008) *An Introduction to Information Retrieval, An Introduction to Information Retrieval*. Cambridge: Cambridge University Press. doi: 10.1109/LPT.2009.2020494.

Marieke Guy, E. T. (2006) « Folksonomies: Tidying up Tags? — the University of Bath's research portal », *D-Lib Magazine*, 12(1). Disponible sur: <https://researchportal.bath.ac.uk/en/publications/folksonomies-tidying-up-tags>.

Markus, H. R. et Kitayama, S. (1991) « Culture and the self: Implications for cognition, emotion, and motivation. », *Psychological review*, 98(2), p. 224.

Marlin, B. (2004) « Modeling user rating profiles for collaborative filtering », *Advances in*

Neural Information Processing Systems, p. 627-634.

McCallum, A. et Nigam, K. (1998) « A comparison of event models for naive bayes text classification », in *AAAI-98 workshop on learning for text categorization*. Citeseer, p. 41-48.

McKenna, K. Y. A. et Bargh, J. A. (2000) « Plan 9 from cyberspace: The implications of the internet for personality and social psychology », *Personality and Social Psychology Review*, 4(1). doi: 10.1207/S15327957PSPR0401_6.

Mercklé, P. (2004) « Sociologie des réseaux sociaux », p. 121.

Mladenic, D. (1999a) « Machine learning used by personal webwatcher ».

Mladenic, D. (1999b) « Text-learning and related intelligent agents: a survey », *IEEE intelligent systems and their applications*, 14(4), p. 44-54.

Montaner, M., López, B. et De La Rosa, J. L. (2003) « A taxonomy of recommender agents on the internet », *Artificial Intelligence Review*, 19(4), p. 285-330. doi: 10.1023/A:1022850703159.

Mooney, R. J. et Roy, L. (2000) « Content-based book recommending using learning for text categorization », in *Proceedings of the fifth ACM conference on Digital libraries*. ACM, p. 195-204.

Moukas, A. (1997) « Amalthea information discovery and filtering using a multiagent evolving ecosystem », *Applied Artificial Intelligence*, 11(5), p. 437-457.

Mukherjee, R. *et al.* (2001) « Movies2go: an online voting based movie recommender system », in *Proceedings of the fifth international conference on Autonomous agents*, p. 114-115.

Negre, E. (2015) *Systèmes de recommandations*. Collection. ISTE Editions.

Nelson, P. (1970) « Information and consumer behavior », *Journal of political economy*, 78(2), p. 311-329.

Nelson, P. (1974) « Advertising as information », *Journal of political economy*, 82(4), p. 729-754.

Ng, A. Y. et Jordan, M. (2000) « PEGASUS: A Policy Search Method for Large MDPs and POMDPs », in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. (UAI'00), p. 406-415.

Noll, M. G. et Meinel, C. (2007) « Web search personalization via social bookmarking and tagging », in *Lecture Notes in Computer Science*. Springer Verlag, p. 367-380. doi: 10.1007/978-3-540-76298-0_27.

O'Reilly, T. (2007) « What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software », *Design*, 65, p. 17-37. doi: 10.2139/ssrn.1008839.

O'Reilly, T. (2009) *What is Web 2.0*. Disponible sur: [https://books.google.fr/books?hl=en&lr=&id=NpEk_WFCMdIC&oi=fnd&pg=PT3&dq=what+is+Web+2.0+&ots=OZSDN9hwET&sig=RGxXllaKu3sF5SOgQZOzFAEJ794&redir_esc=y#v=onepage&q=what is Web 2.0&f=false](https://books.google.fr/books?hl=en&lr=&id=NpEk_WFCMdIC&oi=fnd&pg=PT3&dq=what+is+Web+2.0+&ots=OZSDN9hwET&sig=RGxXllaKu3sF5SOgQZOzFAEJ794&redir_esc=y#v=onepage&q=what+is+Web+2.0&f=false) (Consulté le: 2 décembre 2019).

O'Riordan, C. et Sorensen, H. (1997) *Information Filtering and Retrieval: An Overview*, *citeseer.nj.nec.com/483228.html*. doi: 10.1109/IEMBS.1994.412140.

Oufaida, H. et Nouali, O. (2009) « Exploiting semantic web technologies for recommender systems a multi view recommendation engine », in *CEUR Workshop Proceedings*, p. 87-92.

Oyserman, D., Coon, H. M. et Kimmelmeier, M. (2002) « Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. », *Psychological bulletin*, 128(1), p. 3.

Palme, J. (1998) « Information Filtering », in *Proceedings of the 12th Biennial ITS (International Telecommunications Society) Conference, Stockholm*, p. 1-10.

Park, D. H. *et al.* (2012) « A Literature Review and Classification of Recommender Systems on Academic Journals », *Expert systems with applications*, 39(11), p. 10059-10072.

Paterek, A. (2007) « Improving regularized singular value decomposition for collaborative filtering », *KDD Cup and Workshop*, p. 2-5. doi: 10.1145/1557019.1557072.

Pavlov, D. Y. et Pennock, D. M. (2003) « A Maximum Entropy Approach to Collaborative Filtering in Dynamic, Sparse, High-Dimensional Domains », in Becker, S., Thrun, S., et Obermayer, K. (éd.) *Advances in Neural Information Processing Systems 15*. MIT Press, p. 1465-1472. Disponible sur: <http://papers.nips.cc/paper/2278-a-maximum-entropy-approach-to-collaborative-filtering-in-dynamic-sparse-high-dimensional-domains.pdf>.

Pazzani, M. et Billsus, D. (1997) « Learning and revising user profiles: The identification of interesting web sites », *Machine learning*, 27(3), p. 313-331.

Pazzani, M. J. et Billsus, D. (2007) « Content-based recommendation systems », in *The*

adaptive web. Springer, p. 325-341.

Pazzani, M. J., Muramatsu, J. et Billsus, D. (1996) « Syskill & Webert: Identifying interesting web sites », in *AAAI/IAAI, Vol. 1*, p. 54-61.

Peis, E., Morales del Castillo, J. . et Delgado-López, J. A. (2008) « Semantic Recommender Systems. Analysis of the state of the topic », *Hipertext. net*, 6(2008), p. 1-5. Disponible sur: <https://www.researchgate.net/publication/37764044> (Consulté le: 18 novembre 2019).

Peters, I. (2009) *Folksonomies. Indexing and Retrieval in Web 2.0*. 1 édition. K. G. Saur. doi: 10.1515/9783598441851.

Peters, I. *et al.* (2011) « Social tagging & folksonomies: Indexing, retrieving...and beyond? », *Proceedings of the ASIST Annual Meeting*, 48(January). doi: 10.1002/meet.2011.14504801069.

Postmes, T. *et al.* (2001a) « Social influence in computer-mediated communication: The effects of anonymity on group behavior », *Personality and Social Psychology Bulletin*, 27(10), p. 1243-1254. doi: 10.1177/01461672012710001.

Postmes, T. *et al.* (2001b) « Social influence in computer-mediated communication: The effects of anonymity on group behavior », *Personality and Social Psychology Bulletin*, 27(10), p. 1243-1254.

Postmes, T. et Spears, R. (1998) « Deindividuation and Antinormative Behavior: A Meta-Analysis », *Psychological Bulletin*, 123(3), p. 238-259. doi: 10.1037/0033-2909.123.3.238.

Postmes, T., Spears, R. et Lea, M. (1999) « Social identity, group norms, and “deindividuation”: Lessons from computer-mediated communication for social influence in the group », *Social identity: Context, commitment, content*. Disponible sur: <http://dare.uva.nl/record/78456>.

Postmes, T., Spears, R. et Lea, M. (2000) « The formation of group norms in computer-mediated communication », *Human Communication Research*, 26(3), p. 341-371. doi: 10.1111/j.1468-2958.2000.tb00761.x.

Postmes, Tom, Spears, R. et Lea, M. (2000) « The Formation of Group Norms in Waking Suggestion », *Human Communication Research*, 26(3), p. 341-371. doi: 10.1080/00224545.1948.9918909.

Prentice-Dunn, S. et Rogers, R. W. (1980) « Effects of deindividuating situational cues and

aggressive models on subjective deindividuation and aggression », *Journal of Personality and Social Psychology*, 39(1). doi: 10.1037/0022-3514.39.1.104.

Prentice-Dunn, S. et Rogers, R. W. (1982) « Effects of public and private self-awareness on deindividuation and aggression. », *Journal of Personality and Social Psychology*, 43(3), p. 503.

Prentice-Dunn, S. et Rogers, R. W. (1989) *Deindividuation and the self-regulation of behavior*. Lawrence Erlbaum Associates, Inc.

Quinlan, J. R. (1986) « Induction of decision trees », *Machine Learning*, 1(1), p. 81-106. doi: 10.1007/BF00116251.

Quinlan, J. R. (2014) *C4. 5: programs for machine learning*. Elsevier.

Reicher, S. D. (1984) « Social influence in the crowd: Attitudinal and behavioural effects of de-individuation in conditions of high and low group salience », *British Journal of Social Psychology*, 23(4), p. 341-350.

Reicher, S. D. (1987) « Crowd behaviour as social action », *Rediscovering the social group: A self-categorization theory*, 10.

Reicher, S. D., Spears, R. et Postmes, T. (1995) « A Social Identity Model of Deindividuation Phenomena », *European Review of Social Psychology*, 6(1), p. 161-198. doi: 10.1080/14792779443000049.

Resnick, P. et al. (1994) « GroupLens: An Open Architecture for Collaborative Filtering of Netnews », in *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*. New York, NY, USA: Association for Computing Machinery (CSCW '94), p. 175-186. doi: 10.1145/192844.192905.

Resnick, P. et Varian, H. (1997) « Recommender systems », *Communications of the ACM*, 3(7), p. 222-229. doi: 10.1162/153244302760200641.

Ricci, F. et al. (2011) *Recommender Systems Handbook*. 2011 editi, *Recommender Systems Handbook*. 2011 editi. Édité par F. Ricci et al. Springer. doi: 10.1007/978-0-387-85820-3.

Rocchio, J. (1971) « Relevance feedback in information retrieval », *The Smart Retrieval System-Experiments in Automatic Document Processing*, p. 313-323. Disponible sur: <http://ci.nii.ac.jp/naid/10000074359/en/> (Consulté le: 11 janvier 2020).

Rokach, L. et Maimon, O. (2014) *Data Mining With Decision Trees: Theory and*

Applications. 2nd éd. USA: World Scientific Publishing Co., Inc.

Rokeach, M. (1960) *The open and closed mind*. New York: Basic Books.

Rosario, B. (2000) « Latent Semantic Indexing : An overview », p. 1-16.

Rosenberg, S. (2009) *Say Everything: How Blogging Began, What It's Becoming, and Why It Matters*. Crown. Disponible sur: <https://www.goodreads.com/book/show/6294828-say-everything> (Consulté le: 26 décembre 2019).

Salakhutdinov, R., Mnih, A. et Hinton, G. (2007) « Restricted Boltzmann Machines for Collaborative Filtering », in *Proceedings of the 24th International Conference on Machine Learning*. New York, NY, USA: Association for Computing Machinery (ICML '07), p. 791–798. doi: 10.1145/1273496.1273596.

Saleh, I., Kembellec, G. et Chartron, G. (2014) *Les moteurs et systèmes de recommandation*. Collection. Édité par ISTE. ISTE (September 1, 2014). Disponible sur: http://ulb.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwbV1LS8NAEB5KBREEbdVoVchZSNhsdvM4GrEItSeriJeSfQSEVNQk_8n_4R9zJg8R6WmZveywzOw3M-zMBxByn3n_3gRdGJYajCZ4HCulpYitECy0GH1Lzdr6_kMmF9dilSVPIxg4EfWrKu172VR-Uyo6QA1Uad0HS2IIScIgopHbYcTIzh-zX3tDrEsxGqFkLIg.

Salton, G. (1971) *The SMART Retrieval System---Experiments in Automatic Document Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Salton, G. (1989) « Automatic text processing: The transformation, analysis, and retrieval of », *Reading: Addison-Wesley*, 169.

Salton, G. et Buckley, C. (1988) « Term-weighting approaches in automatic text retrieval », *Information processing & management*, 24(5), p. 513-523.

Salton, G., Wong, A. et Yang, C. S. (1975) « A Vector Space Model for Automatic Indexing », 18(11).

Sammut, C. et Webb, G. I. (2017) *Encyclopedia of Machine Learning and Data Mining*. 2nd éd. Springer Publishing Company, Incorporated.

Sankar, K. et Bouchard, S. A. (2009) *Enterprise Web 2.0 fundamentals*. Cisco Press.

Sarwar, B. *et al.* (2000a) « Analysis of recommendation algorithms for e-commerce », in *EC'00 Proceedings of the 2nd ACM Conference on Electronic Commerce*. Minneapolis, MN, Etat-Unis: ACM, p. 158-167.

Sarwar, B. *et al.* (2000b) « Application of dimensionality reduction in recommender system-a case study », *ACM WebKDD 2000 Web Mining for ECommerce Workshop*. doi: 10.3141/1625-22.

Sarwar, B. *et al.* (2001) « Item-based collaborative filtering recommendation algorithms », in *Proceedings of the 10th International Conference on World Wide Web, WWW 2001*, p. 285-295. doi: 10.1145/371920.372071.

Schafer, J. Ben *et al.* (2007) « Collaborative filtering recommender systems », in *The adaptive web*. Springer, p. 291-324.

Schein, A. I. *et al.* (2002) « Methods and metrics for cold-start recommendations », *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, (August), p. 253-260.

Schmidt, D. (2003) « Social Software: Facilitating Information-, Identity- and Relationship Management », in Burg, Thomas N.; Schmidt, J. (éd.) *BlogTalks reloaded. Social Software - Research & Cases*. BlogTalks. Books On Demand, p. 31-49. doi: 10.1016/s1525-5050(03)00040-4.

Semeraro, G. *et al.* (2009) « User profiles for personalizing digital libraries », in *Handbook of Research on Digital Libraries: Design, Development, and Impact*. IGI Global, p. 149-158.

Sen, S. *et al.* (2006) « Tagging, communities, vocabulary, evolution », *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, p. 181-190. doi: 10.1145/1180875.1180904.

Sfetcu, N. (2017) *Web 2.0 / Social Media / Social Networks: Illustrated Edition*. 1^{re} éd. CreateSpace Independent Publishing Platform. Disponible sur: <https://dl.acm.org/doi/book/10.5555/3153671>.

Shardanand, U. et Maes, P. (1995) « Social information filtering: algorithms for automating “word of mouth” », *Proceedings of the ACM CHI 95 Human Factors in Computing Systems Conference*, 1, p. 210-217.

Sherif, M. (1958) « Superordinate goals in the reduction of intergroup conflict », *American journal of Sociology*, 63(4), p. 349-356.

Sherif, M. *et al.* (1961) *Intergroup conflict and cooperation: The Robbers Cave experiment*. University Book Exchange Norman, OK.

Sherif, M. (1966) « In common predicament: Social psychology of intergroup conflict and cooperation: The robber's cave experiment », *Norman: University of Oklahoma Book Exchange*.

Sherif, M. (2017) *Social interaction: Process and products*. Routledge.

Sherif, M. E. (1962) *Intergroup relations and leadership: Approaches and research in industrial, ethnic, cultural and political areas*. John Wiley.

Sheth, B. et Maes, P. (1993) « Evolving agents for personalized information filtering », in *Proceedings of 9th IEEE Conference on Artificial Intelligence for Applications*. IEEE, p. 345-352.

Sidanius, J. et Pratto, F. (2019) « Social Dominance Theory: A New Synthesis », in *Political Psychology*. Psychology Press, p. 315-332. doi: 10.4324/9780203505984-18.

Singer, J. E., Brush, C. A. et Lublin, S. C. (1965) « Some aspects of deindividuation: Identification and conformity », *Journal of Experimental Social Psychology*, 1(4). doi: 10.1016/0022-1031(65)90015-6.

Smith, G. (2004) *Folksonomy: Social Classification*. Disponible sur: http://atomiq.org/%0Aarchives/2004/08/folksonomy_social_classification.html (Consulté le: 4 janvier 2020).

Sorensen, H. et McElligott, M. (1995) « PSUN: a profiling system for Usenet news », in *Proceedings of CIKM*. Citeseer, p. 1-2.

Sparling, E. I. et Sen, S. (2011) « Rating: How difficult is it? », in *RecSys '11 - Proceedings of the 5th ACM Conference on Recommender Systems*, p. 149-156. doi: 10.1145/2043932.2043961.

Spears, R. et Lea, M. (1992) *Social influence and the influence of the 'social' in computer-mediated communication*. Harvester Wheatsheaf.

Spears, R. et Lea, M. (1994) « Panacea or Panopticon? », *Communication Research*, 21(4), p. 427-459. doi: 10.1177/009365094021004001.

Spears, R., Lea, M. et Lee, S. (1990) « De-individuation and group polarization in computer-mediated communication », *British Journal of Social Psychology*, 29(2), p. 121-134. doi: 10.1111/j.2044-8309.1990.tb00893.x.

Spiteri, L. (2006) « The use of collaborative tagging in public library catalogues », in *Advances in Classification Research Online*, p. 29-30. doi: 10.1002/meet.14504301214.

Stadnyk, I. et Kass, R. (1992) « Modeling Users' Interests in Information Filters », *Commun. ACM*, 35(12), p. 49-50. doi: 10.1145/138859.138864.

Su, X. et Khoshgoftaar, T. M. (2009) « A survey of collaborative filtering techniques », *Advances in artificial intelligence*, 2009.

Sumner, W. G. (1906) *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*. Ginn.

Tajfel, H. (1970) « Experiments in Intergroup Discrimination », *Scientific American*, 223(5), p. 96-103. Disponible sur: <http://www.jstor.org/stable/24927662>.

Tajfel, H. *et al.* (1971) « Social categorization and intergroup behaviour. », *European Journal of Social Psychology*, 1(2), p. 149-178. doi: 10.1002/ejsp.2420010202.

Tajfel, H. (1972) « Social categorization, English manuscript of La catégorization sociale », *Introduction a la psychologie sociale*, 1.

Tajfel, H. (1978) « Interindividual behaviour and intergroup behaviour », *Differentiation between social groups: Studies in the social psychology of intergroup relations*, p. 27-60.

Tajfel, H. (1981) *Human groups and social categories: Studies in social psychology*. Cup Archive.

Tajfel, H. et Turner, J. (2001) « An integrative theory of intergroup conflict. », in *Intergroup relations: Essential readings*. New York, NY, US: Psychology Press (Key readings in social psychology.), p. 94-109.

Tajfel, H. et Turner, J. C. (1979) « An integrative theory of intergroup conflict », in Worchel, S. et Austin, W. (éd.) *The social psychology of intergroup relations*. Pacific Grove, p. 33-48.

Tajfel, H. et Turner, J. C. (1986) « The social identity theory of intergroup behavior », in S. Worchel et W. Austin (éd.) *Psychology of intergroup relations*. 2nd éd. Chicago : Nelson-Hall, p. 7-24.

Takács, G. *et al.* (2008) « Investigation of Various Matrix Factorization Methods for Large Recommender Systems », in *2008 IEEE International Conference on Data Mining Workshops*, p. 553-562. doi: 10.1109/ICDMW.2008.86.

Tan, P. N., Steinbach, M. et Kumar, V. (2006) « Introduction to data mining, Addison Wesley Publishers ».

Tarde, G. (1890) *Les lois de l'imitation, Les lois de l'imitation*. Paris: Librairie Felix Alcan.

Taylor, D. M. et Moghaddam, F. M. (1994) *Theories of intergroup relations: International social psychological perspectives*. Greenwood Publishing Group.

Tenenbaum, J. M. et Hughes, K. (2006) « AI meets web 2.0: Building the web of tomorrow, today », *AI Magazine*, 27(4), p. 47-68.

Terveen, L. et Hill, W. (2001) « Beyond recommender systems: Helping people help each other », *HCI in the New Millennium*, (1), p. 487-509. doi: 10.1.1.26.2437.

Thelwall, M. (2009) « Social Network Sites : Users and Uses », *Advances*, 76(09), p. 19-73. doi: 10.1016/S0065-2458(09)01002-X.

Turner, J. C. (1982) « Towards a cognitive redefinition of the social group », *Social identity and intergroup relations*, p. 15-40.

Turner, J. C. (1985) « Social categorization and the self-concept: A social cognitive theory of group behavior. », *Advances in Group Processes: Theory and Research*. Vol. 2. Édité par E. J. Lawler, 2, p. 77-122.

Turner, J. C. *et al.* (1987) *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

Turner, J. C. (1991) *Social influence.*, *Social influence*. Belmont, CA, US: Thomson Brooks/Cole Publishing Co (Mapping social psychology series.).

Turner, J. C. et Giles, H. (1981) *Intergroup behavior*. Blackwell Oxford, UK.

Turner, J. C. et Oakes, P. J. (1986) « The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence », *British Journal of Social Psychology*, 25(3), p. 237-252.

Ungar, L. H. et Foster, D. P. (1998) « Clustering methods for collaborative filtering », *AAAI Workshop on Recommendation Systems*, p. 114-129. doi: 10.1.1.33.4026.

Viappiani, P., Faltings, B. et Pu, P. (2006) « Preference-based search using example-critiquing with suggestions », *Journal of artificial intelligence Research*, 27, p. 465-503.

Vilanova, F. *et al.* (2017) « Deindividuation: From Le Bon to the social identity model of deindividuation effects », *Cogent Psychology*, 4(1), p. 1-21. doi: 10.1080/23311908.2017.1308104.

Vinacke, W. E. (1964) « Intra-group power relations, strategy, and decisions in inter-triad competition », *Sociometry*, p. 25-39.

Vander Wal, T. (2005) *Explaining and Showing Broad and Narrow Folksonomies*. Disponible sur: <http://www.vanderwal.net/random/entrysel.php?blog=1635>.

Vander Wal, T. (2007) *Folksonomy: Folksonomy Coinage and Definition*. Disponible sur: <http://www.vanderwal.net/folksonomy.html> (Consulté le: 4 janvier 2020).

We Are Social & Hootsuite (2019a) « Digital 2019 China », *We Are Social & Hootsuite*, p. 76. Disponible sur: https://es.slideshare.net/DataReportal/digital-2019-argentina-january-2019-v01?from_action=save.

We Are Social & Hootsuite (2019b) « Digital 2019 France », *We Are Social & Hootsuite*, p. 76. Disponible sur: https://es.slideshare.net/DataReportal/digital-2019-argentina-january-2019-v01?from_action=save.

Weber, S. et Rech, J. (2009) « An overview and differentiation of the evolutionary steps of the Web X.Y movement: The Web before and beyond 2.0 », *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications*, 1, p. 12-38. doi: 10.4018/978-1-60566-384-5.ch002.

Westford, K. L., Diener, E. et Diener, C. (1973) « Deindividuating effects of group presence and arousal on stealing by Halloween trick-or-treaters. », *Proceedings of the Annual Convention of the American Psychological Association*, p. 219-220.

Wicklund, R. A. (1975) « Objective Self-Awareness11Much of the research reported in this paper as well as the writing of this paper were supported by NSF Grant GS-31890. Sharon S. Brehm, William J. Ickes, Michael F. Scheier, and Melvin L. Snyder are acknowledged for their suggest », in Berkowitz, L. B. T.-A. in E. S. P. (éd.). Academic Press, p. 233-275. doi: [https://doi.org/10.1016/S0065-2601\(08\)60252-X](https://doi.org/10.1016/S0065-2601(08)60252-X).

Wilson, D. W. *et al.* (2011) « Web 2.0: A Definition, Literature Review, and Directions for Future Research », *AMCIS 2011 Proceedings*, (November 2014), p. Paper 368. Disponible sur: http://aisel.aisnet.org/amcis2011_submissions%5Cnhttp://aisel.aisnet.org/amcis2011_submissions.

Yan, T. W. et Garcia-Molina, H. (1999) « The SIFT information dissemination system », *ACM Transactions on Database Systems (TODS)*, 24(4), p. 529-565.

Yang, X. *et al.* (2014) « A survey of collaborative filtering based social recommender systems », *Computer Communications*, 41, p. 1-10. doi: 10.1016/j.comcom.2013.06.009.

Zeuzula, P. *et al.* (2005) *Similarity Search: The Metric Space Approach (Advances in*

Database Systems). Berlin, Heidelberg: Springer-Verlag.

Zhang, T., Ramakrishnan, R. et Livny, M. (1996) « BIRCH: an efficient data clustering method for very large databases », in *ACM Sigmod Record*. ACM, p. 103-114.

Zhao, X. M., Ye, X. Q. et Zhu, D. Y. (2008) *A novel compound isolated from the peels of Citrus changshan-huyou* Y. B. Chang, *Yaoxue Xuebao*.

Ziller, R. C. (1964) « Individuation and Socialization: A Theory of Assimilation in Large Organizations », *Human Relations*, 17(4). doi: 10.1177/001872676401700403.

Zimbardo, P. G. (1969) « The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. », *Nebraska Symposium on Motivation*, 17, p. 237-307.

Zitnick, C. L. et Kanade, T. (2004) « Maximum Entropy for Collaborative Filtering », in *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*. Arlington, Virginia, USA: AUAI Press (UAI '04), p. 636–643.

ANNEXES

1. Questionnaire auprès des utilisateurs de Douban Film

Introduction

Bonjour,

Actuellement étudiant de l'université Charles-de-Gaulle de Lille 3, nous avons élaboré une enquête pour notre étude sur Douban Film. Plus précisément, nous avons enquêté sur les manières par lesquelles les utilisateurs de notent les films et s'ils sont influencés par les notations des autres membres.

Votre réponse est très importante pour notre étude : c'est la raison pour laquelle nous souhaiterions que vous répondiez au questionnaire ci-dessous qui comporte seulement 19 questions et qui ne dura pas plus de 10 minutes. Cette enquête est anonyme et nous ne vous demanderons aucune information privée.

Première partie

Votre sexe :

Votre âge

- a. Moins de 18 ans
- b. De 19 – 25 ans
- c. De 26 -29 ans
- d. De 30- 39 ans
- e. De 40- 49 ans
- f. Plus de 50 ans

Deuxième partie : quand vous vous renseignez sur les informations des films sur Douban avant d'aller voir des films

- 1. A quelle fréquence découvrez-vous des films qui vous intéressent ?
 - a. Toujours
 - b. Souvent
 - c. Rarement
 - d. Jamais
- 2. Que pensez-vous d'un film très bien noté (5 étoiles) sur Douban ?
 - a. C'est un film très bien filmé
 - b. C'est un film aimé par beaucoup d'utilisateurs
- 3. Que pensez-vous d'un film très mal noté (1 étoile) sur Douban ?
 - a. C'est un film très mal filmé
 - b. C'est un film détesté par beaucoup d'utilisateurs

Troisième partie : quand vous allez noter un film sur Douban après l'avoir vu

4. En général, votre évaluation d'un film dépend
 - a. De la qualité d'un film
 - b. Plutôt de la qualité d'un film que votre préférence
 - c. A la fois de la qualité d'un film et de votre préférence
 - d. Plutôt de votre préférence que la qualité d'un film
 - e. De votre préférence
5. Pour les films appartenant aux genres que vous préférez, vos évaluations dépendent
 - a. De la qualité d'un film
 - b. Plutôt de la qualité d'un film que votre préférence
 - c. A la fois de la qualité d'un film et de votre préférence
 - d. Plutôt de votre préférence que la qualité d'un film
 - e. De votre préférence
6. Pour les films ne pas appartenant aux genres que vous préférez, vos évaluations dépendent
 - a. De la qualité d'un film
 - b. Plutôt de la qualité d'un film que votre préférence
 - c. A la fois de la qualité d'un film et de votre préférence
 - d. Plutôt de votre préférence que la qualité d'un film
 - e. De votre préférence
7. A votre avis, la qualité d'un film dépend
 - a. Du réalisateur
 - b. Du scénario
 - c. Des acteurs et des actrices
 - d. De la photographie
 - e. De la musique
 - f. De l'effet visuel
 - g. Autres _____
8. Trouvez-vous forcément bons les films ceux qui sont bien filmés ?
 - a. Oui
 - b. Non
9. Trouvez-vous forcément mauvais les films ceux qui ne sont pas bien filmés ?
 - a. Oui
 - b. Non
10. Donnez-vous une bonne note à un film bien filmé bien que vous ne le trouviez pas bon vraiment ?
 - a. Oui
 - b. J'hésite souvent
 - c. Non
11. Donnez-vous une bonne note à un film que vous trouvez bon bien qu'il ne soit pas bien filmé ?
 - a. Oui
 - b. J'hésite souvent
 - c. Non
12. Hésitez-vous lors de noter un film (entre 5 étoiles et 4 étoiles pour un bon film ; entre 2 étoiles et 1 étoile pour un mauvais film) ?
 - a. Toujours
 - b. Souvent
 - c. Rarement
 - d. Jamais
13. Si vous hésitez lors de noter un film, tenez-vous compte d'opinions d'autrui ?

- a. Toujours
 - b. Souvent
 - c. Rarement
 - d. Jamais
14. Lors de noter un film, hésitez-vous si votre avis n'est pas en accord avec l'opinion de la majorité ?
- a. Toujours
 - b. Souvent
 - c. Rarement
 - d. Jamais
15. Pour un film très bien noté (4-5 étoiles) sur Douban mais que vous trouvez mauvais (1-2 étoiles)
- a. Vous donne une bonne note (4-5 étoiles)
 - b. Vous donne une bonne note mais moins bonne que ce que les autres pensent (3-4 étoiles)
 - c. Vous donne une mauvaise note mais meilleure que ce que vous pensez (2-3 étoiles)
 - d. Vous donne une mauvaise note (1-2 étoiles)
16. Pour un film très mal noté (1-2 étoiles) sur Douban mais que vous trouvez honorable (4-5 étoiles)
- a. Vous donne une mauvaise note (1-2 étoiles)
 - b. Vous donne une mauvaise note mais meilleure que ce que les autres pensent (2-3 étoiles)
 - c. Vous donne une bonne note mais moins bonne que ce que vous pensez (3-4 étoiles)
 - d. Vous donne une bonne note (4-5 étoiles)
17. Lors de noter un film, hésitez-vous si votre avis n'est pas en accord avec l'opinion de la plupart de vos amis ?
- a. Toujours
 - b. Souvent
 - c. Rarement
 - d. Jamais
18. Pour un film très bien noté (4-5 étoiles) par la plupart de vos amis mais que vous trouvez mauvais (1-2 étoiles)
- a. Vous donne une bonne note (4-5 étoiles)
 - b. Vous donne une bonne note mais moins bonne que ce que vos amis pensent (3-4 étoiles)
 - c. Vous donne une mauvaise note mais meilleure que ce que vous pensez (2-3 étoiles)
 - d. Vous donne une mauvaise note (1-2 étoiles)
19. Pour un film très mal noté (1-2 étoiles) par la plupart de vos amis mais que vous trouvez honorable (4-5 étoiles)
- a. Vous donne une mauvaise note (1-2 étoiles)
 - b. Vous donne une mauvaise note mais meilleure que ce que la plupart de vos amis pensent (2-3 étoiles)
 - c. Vous donne une bonne note mais moins bonne que ce que vous pensez (3-4 étoiles)
 - d. Vous donne une bonne note (4-5 étoiles)
20. Lors de noter un film, hésitez-vous si votre avis n'est pas en accord avec l'opinion des critiques de cinéma que vous suivez ?

- a. Toujours
 - b. Souvent
 - c. Rarement
 - d. Jamais
21. Pour un film très bien évalué (4-5 étoiles) par les critiques de cinéma que vous suivez mais que vous trouvez mauvais (1-2 étoiles)
- a. Vous donne une bonne note (4-5 étoiles)
 - b. Vous donne une bonne note mais moins bonne que ce que les critiques de cinéma pensent (3-4 étoiles)
 - c. Vous donne une mauvaise note mais meilleure que ce que vous pensez (2-3 étoiles)
 - d. Vous donne une mauvaise note (1-2 étoiles)
22. Pour un film très mal noté (1-2 étoiles) par les critiques de cinéma que vous suivez mais que vous trouvez honorable (4-5 étoiles)
- a. Vous donne une mauvaise note (1-2 étoiles)
 - b. Vous donne une mauvaise note mais meilleure que ce que les critiques de cinéma pensent (2-3 étoiles)
 - c. Vous donne une bonne note mais moins bonne que ce que vous pensez (3-4 étoiles)
 - d. Vous donne une bonne note (4-5 étoiles)
23. Pensez-vous que vos évaluations seraient différentes si elles n'étaient pas publiques ?
- a. Complètement
 - b. Largement
 - c. Plus ou moins
 - d. Pas du tout