

UNIVERSITÉ DE TUNIS EL MANAR (LIPAH)  
UNIVERSITÉ BLAISE PASCAL CLERMONT FERRAND II (LIMOS)

## IC 2014

# Vers des recommandations plus personnalisées dans les folksonomies

Mohamed Nader Jelassi  
Engelbert Mephu Nguifo  
Sadok Ben Yahia

15 mai 2014

# Plan

- 1 Introduction et Motivations
- 2 État de l'art
- 3 Un nouveau système personnalisé de recommandation
- 4 Résultats et Discussion
- 5 Conclusion et Perspectives

# Plan

- 1 Introduction et Motivations
- 2 État de l'art
- 3 Un nouveau système personnalisé de recommandation
- 4 Résultats et Discussion
- 5 Conclusion et Perspectives

# Plan

- 1 Introduction et Motivations
- 2 État de l'art
- 3 Un nouveau système personnalisé de recommandation
- 4 Résultats et Discussion
- 5 Conclusion et Perspectives

# Plan

- 1 Introduction et Motivations
- 2 État de l'art
- 3 Un nouveau système personnalisé de recommandation
- 4 Résultats et Discussion
- 5 Conclusion et Perspectives

# Plan

- 1 Introduction et Motivations
- 2 État de l'art
- 3 Un nouveau système personnalisé de recommandation
- 4 Résultats et Discussion
- 5 Conclusion et Perspectives

# Plan

- 1 Introduction et Motivations
- 2 État de l'art
- 3 Un nouveau système personnalisé de recommandation
- 4 Résultats et Discussion
- 5 Conclusion et Perspectives

# Folksonomie = Porte-Manteau

## Les Utilisateurs

# FOLK



Done by people

## Organisation de Données

# TAXONOMY



Classifications of items

# 3 acteurs principaux

## Contexte triadique : Utilisateurs-Tags-Ressources



**UTILISATEURS**



**TAGS**



**RESSOURCES**

# Zoom...sur les utilisateurs

## Qui sont-ils ?

- Principaux Acteurs : contribuent à la création de l'information
- Accessible à tous
- Responsable du partage
- Choix personnel de tags/ressources : dépend de l'utilisateur (en fonction de son profil : âge, profession, pays, etc.)

# Nouvelle dimension

## Différents Utilisateurs = Différents Profils



## Un contexte à 4 dimensions

- Le profil des utilisateurs comme information supplémentaire
- V-Folksonomy = (Utilisateurs, Tags, Ressources, **Profil**)

# Problématiques

- Surcharge d'information
- Manque d'organisation des données : Quoi et comment choisir ?
- Quelles ressources puis-je partager ?
- Quels tags puis-je utiliser ?

# Solution : Personnaliser les recommandations

## Objectifs

- Comprendre les besoins des utilisateurs
- Refléter le profil de chaque utilisateur
- Aider chaque utilisateur à :
  - Partager les "bonnes" ressources : les plus intéressantes
  - Utiliser les "bons" tags : les plus adéquats
  - Se mettre en contact avec les "bonnes" personnes

# Profil : en quoi peut-il aider ?

## Profil

Durant le processus de personnalisation des recommandations, il est utile d'avoir des informations supplémentaires sur les utilisateurs :

**Information personnelles** âge, genre, profession, pays, etc.

**Partages publics** (Tags, Ressources) partagés

# Plan

- 1 Introduction et Motivations
- 2 État de l'art**
- 3 Un nouveau système personnalisé de recommandation
- 4 Résultats et Discussion
- 5 Conclusion et Perspectives

## Trois principales catégories

- 1 Approches utilisant des mesures de similarité
- 2 Approches se basant sur la popularité
- 3 Approches hybrides utilisant l'historique de tagging et des mesures de similarité

# Approches utilisant des mesures de similarité

## Principales Approches

- (Diederich *et al.*, 2006)
  - Une mesure de similarité entre utilisateurs
  - Deux utilisateurs sont similaires s'ils partagent des tags similaires
- (Landia *et al.*, 2009)
  - Se base sur deux mesures de similarité (entre utilisateurs et entre ressources)
  - Recommandation personnalisée de tags

# Approches se basant sur la popularité

## Principales Approches

- (Lipczak, 2008)
  - Recommandation de tags par co-occurrence sur d'autres ressources
- (Hotho *et al.*, 2007)
  - Recommandation de tags les plus populaires (les plus utilisés)
  - Recours aux règles d'association entre tags

# Approches hybrides utilisant l'historique de tagging et des mesures de similarité

## Principales Approches

- (Hu *et al.*, 2011)
  - Recommandation se basant sur les contacts sociaux des utilisateurs et sur les tags/ressources déjà partagés
- (Basile *et al.*, 2007)
  - Recommandation de tags à partir de similarité entre ressources et de tags/ressources déjà partagés par les utilisateurs

# Contributions

- Combiner le profil des utilisateurs avec les (tags,ressources) partagés
- Zoomer sur les (tags, ressources) les plus utilisés
- Un nouveau système personnalisé de recommandation

# Plan

- 1 Introduction et Motivations
- 2 État de l'art
- 3 Un nouveau système personnalisé de recommandation**
- 4 Résultats et Discussion
- 5 Conclusion et Perspectives

# Un système de recommandation en 2 temps

- 1 **Extraction des Quadri-Concepts** : Regrouper sous forme de structures quadratiques les utilisateurs, tags et ressources les plus utilisés
  - Data Peeler (Cerf *et al.*, 2009)
  - QuadriCons (Jelassi *et al.*, 2014)
- 2 **Recommandation Personnalisée** : L'algorithme PersoRec

## Phase de pré-traitement : Extraction de Quadri-concepts

### C'est quoi un quadri-concept ?

Un Quadri-concept est une structure quadratique  $(U, T, R, P)$

- 1  $U$  : l'ensemble d'utilisateurs
  - 2  $T$  : l'ensemble de tags
  - 3  $R$  : l'ensemble de ressources
  - 4  $P$  : l'ensemble de profils
- **Définition** : Chaque utilisateur de  $U$  (avec les profils de  $P$ ) a tagué chaque ressource de  $R$  avec tous les tags de  $T$ .
  - **Propriété** : Le Quadri-Concept est **maximal** : aucun de ces ensembles ne peut être augmenté sans faire diminuer l'une des trois autres dimensions.
  - **Pourquoi les quadri-concepts ?** Les quadri-concepts sont une représentation réduite d'une folksonomie qui peut contenir des milliers de quadruplets. ( $|QC| \ll |\mathcal{F}|$ )

## Phase de pré-traitement : Extraction de Quadri-concepts

### Exemple d'un quadri-concept

$QC = (\{john, peter, dana\},$   
 $\{papers, books, reviews\},$   
 $\{sciencedirect.com, springer.com\},$   
 $\{18-25 \text{ ans}, \text{étudiant}\})$

- Nous lisons : : **John**, **Peter** et **Dana**, trois **étudiants** âgés entre **18 et 25 ans** ont partagé les sites web **sciencedirect.com**, **springer.com** avec les tags **papers**, **books** et **reviews**.

# Algorithme PersoRec

## Entrées

- $QC$  : Un ensemble de quadri-concepts
- $(u,p)$  : l'utilisateur cible et son profil

## Sorties

- 1 Un ensemble d'amis proposés
- 2 Un ensemble de tags suggérés
- 3 Un ensemble de ressources recommandées

# Algorithme PersoRec

## Pseudo-Code

### Algorithme 1 : PERSOREC

**Données :** l'ensemble des quadri-concepts fréquents  $QC$ , un utilisateur  $u$  avec sa variable  $v$  et une ressource  $r$ .

**Résultats :** l'ensemble d'utilisateurs proposés  $PU$ , l'ensemble des tags suggérés  $ST$  et l'ensemble des ressources recommandées  $RR$ .

```
1  début
2  pour chaque quadri-concept  $qc \in QC$  faire
3      si  $v \in qc.Variables$  alors
4          si  $u \notin qc.Utilisateurs$  alors
5              /*Proposition d'utilisateurs*/
6               $PU = PU \cup qc.extent$ ;
7              /*Suggestion de Tags*/
8              si  $r \in qc.Ressources$  alors
9                   $ST = ST \cup qc.Tags$ ;
10             /*Recommandation de Ressources*/
11              $RR = RR \cup qc.Ressources$ ;
12  retourner  $(PU, ST, RR)$ ;
13  fin
```

# Algorithme PersoRec

## Remarques

- La phase d'extraction des quadri-concepts est une phase **hors-ligne** exécutée une **seule** fois.
- L'algorithme PersoRec est totalement indépendant de la phase de pré-traitement.

## Complexité

- Complexité (Extraction des quadri-concepts) =  $\mathcal{O}(2^n)$  avec  $n = |\mathcal{T}| + |\mathcal{R}| + |\mathcal{P}|$ .
- Complexité (PersoRec) =  $\mathcal{O}(m)$  avec  $m = |\mathcal{QC}| \ll n$

# Plan

- 1 Introduction et Motivations
- 2 État de l'art
- 3 Un nouveau système personnalisé de recommandation
- 4 Résultats et Discussion**
- 5 Conclusion et Perspectives

# Données et Plan d'évaluation

## Jeux de données

- 1 Le jeu de données filmographique MovieLens : 95580 tags appliqués à 10681 films par 71567 utilisateurs
- 2 Le jeu de données de livres BookCrossing : 278858 utilisateurs donnant 1149780 notes à 271379 livres
  - Profils : Genre, Age, Profession, Localisation

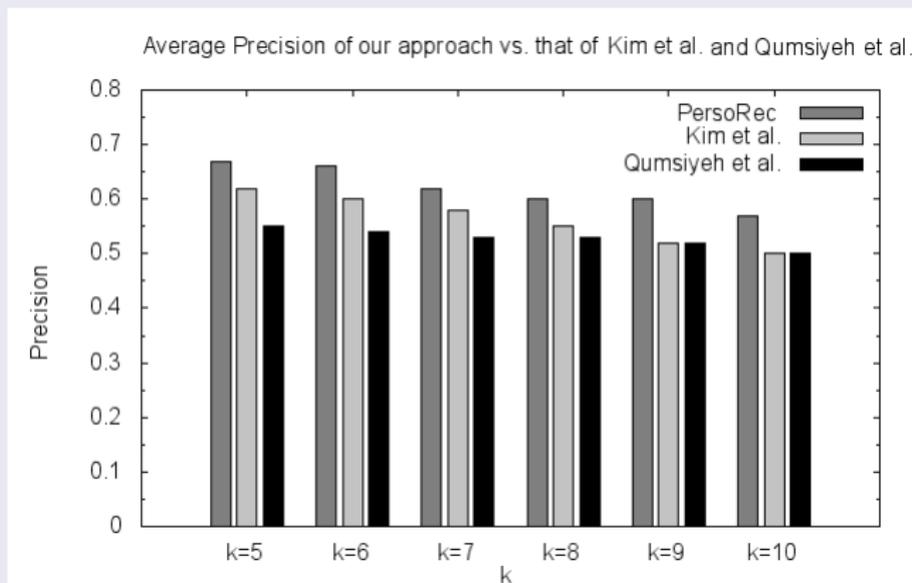
## Plan de l'évaluation

- 1 Métriques de Qualité : Précision, Rappel et F1-Score
- 2 Évaluation Sociale
- 3 Propriétés des systèmes de recommandations

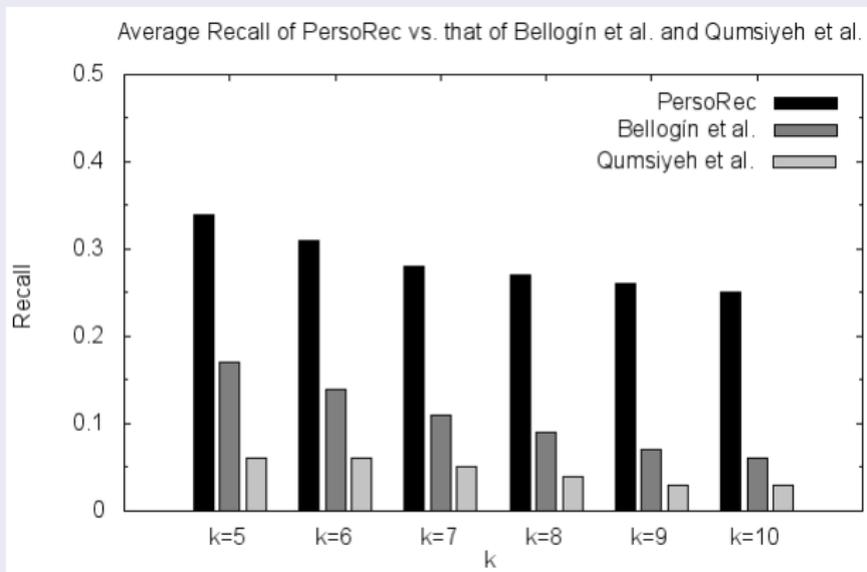
## Exemples de quadri-concepts (MovieLens)

Utilisateurs	Tags	Ressources	Profil
{saloua, yasmine, wafa}	{classic, dialog, oscar}	{Star Wars, Magician of OZ, Rear Window}	{Femme, 46-73 ans, retraité}
{mulder, scully, krycek}	{bestmovie, cult}	{Usual Suspects, Silence of the Lambs, X-Files}	{Homme , 25-35 ans, médecin}
{ross, anlucia, franela}	{classic, oldmovie, quotes}	{Rear Window, Magician of OZ, Gone with the Wind}	{Homme, 36-45 ans, Ecrivain}

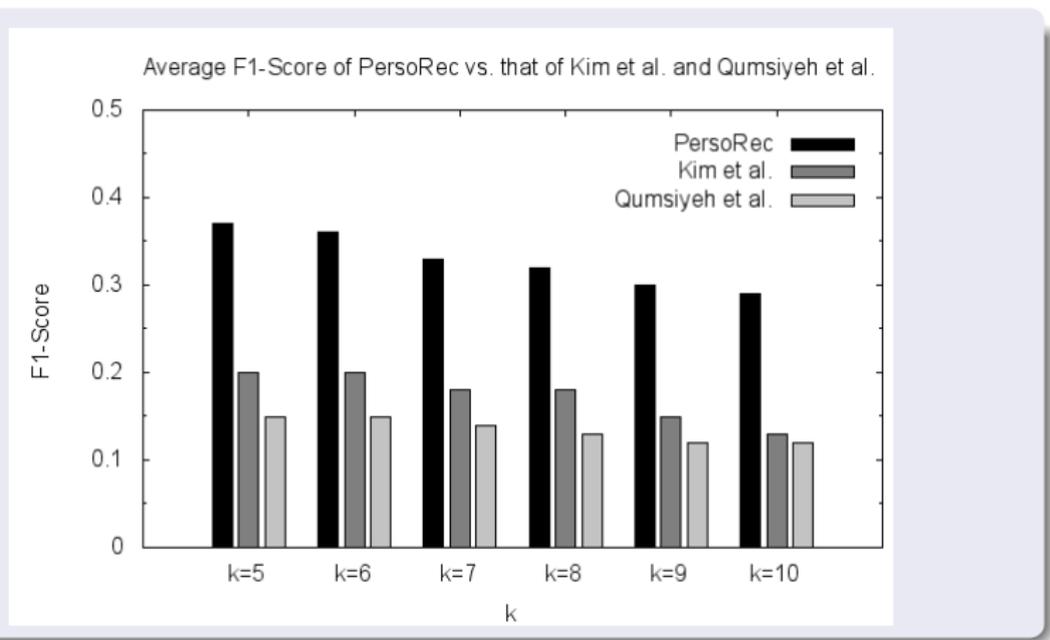
# Précision



# Rappel



# F1-Score



# Métriques d'évaluation

## Explications

- les travaux de la littérature s'appuient surtout sur les items les plus utilisés (livres, films, tags)
- le profil des utilisateurs : Information utile qui améliore les recommandations
- avantage des quadri-concepts : Recommander les tags et ressources les plus partagés

# Évaluation Sociale

## L'idée

- qu'est-ce qui se passe après la phase de recommandation ?
- l'utilisateur cible a-t-il réellement apprécié les recommandations ?
- les utilisateurs (amis) qu'on lui a proposé adoptent-ils le même comportement ?

## Sujets de test

- 1 BookCrossing : *skinner* (38 ans, New York, USA)
- 2 MovieLens : *thiago\_motta* (47 ans, Homme, Édicateur)

# Évaluation Sociale (BookCrossing)

## Recommandations

- trois livres de la franchise *Harry Potter*
- quatre nouveaux utilisateurs : *snowh* (43 ans, *Illinois, USA*), *char\_dav* (54 ans, *California, USA*), *emma* (40 ans, *Oregon, USA*) et *henry90* (36 ans, *Teheran, Iran*).

## Conclusions

- ses nouveaux amis ont également partagé tous les livres de la franchise *Harry Potter*
- *skinner* et ses nouveaux amis ont noté les livres recommandés avec une note égale à 10  
⇒ ils ont apprécié les recommandations

# Propriétés de PersoRec

## Couverture de l'Espace Utilisateur

- Tous les utilisateurs sont couverts
- Profils couverts : 100% de genre (homme et femme), 100% des catégories d'âge, 100% de professions et 88% des pays

## Cold Start : Nouveaux Utilisateurs

- Recommandations selon le profil.
- Tous les utilisateurs reçoivent des recommandations qu'ils aient déjà partagé ou non.

# Propriétés de PersoRec

## Diversité

- Les utilisateurs préfèrent des recommandations diverses : e.g., 5 livres de 5 auteurs différents plutôt que 5 livres d'un même auteur
- $d$  mesure la distance entre les items recommandés
- Un score moyen de 56%

# Propriétés de PersoRec

## Scalabilité (Passage à l'échelle)

	<i>minsupp-u</i>	$ QC $	# UU	Tâche1 (ms)	Tâche2 (ms)
(MovieLens)	20	221	526	0,1	2,6
	16	500	605	0,2	3,9
	12	1295	668	0,7	6,1
	8	5123	805	4,0	13,5
	6	13461	865	12,7	23,3
(BookCrossing)	30	553	6789	0,9	149,8
	20	1486	9092	4,9	296,9
	16	2638	10397	13,0	415,5
	12	5698	12239	45,0	542,3

- Quadri-Concepts = Représentation réduite d'une folksonomie
- # UU = # Utilisateurs uniques = Nombre de recommandations
- Un temps de réponse moyen de 2 ms pour la recommandation de ressources (**Tâche1**) et de 8 ms pour la proposition d'utilisateurs (**Tâche2**)

## Comparaison avec les travaux de la littérature

	<b>M-Mode</b>	<b>Couv.</b>	<b>C-Start</b>	<b>Div.</b>	<b>Sca.</b>
(Diedrich <i>et al.</i> , 2006)	Non	Non	Non	Oui	Non
(Basile <i>et al.</i> , 2007)	Non	Non	Non	?	?
(Hotho <i>et al.</i> , 2007)	Non	Non	Non	Oui	?
- (Lipczak, 2008)	Non	Non	Non	Oui	?
(Landia <i>et al.</i> , 2009)	Non	Non	Non	Non	?
(De Meo <i>et al.</i> , 2010)	Non	Non	Non	?	?
(Hu <i>et al.</i> , 2011)	Non	Non	Non	Oui	?
Notre approche	Oui	Oui	Oui	Oui	Oui

- M-Mode : Algorithme multi-mode (recommandation d'utilisateurs, tags et ressources)
- Couv : Couverture de l'espace utilisateur
- C-Start : Cold Start
- Div : Diversité des recommandations
- Sca : Scalabilité

# Plan

- 1 Introduction et Motivations
- 2 État de l'art
- 3 Un nouveau système personnalisé de recommandation
- 4 Résultats et Discussion
- 5 Conclusion et Perspectives

# Conclusion

- Prise en compte du profil des utilisateurs
- Recommandations plus ciblées
- Analyse des propriétés du système de recommandation

# Perspectives

- Extension de la 4e dimension à d'autres variables : le temps, historique de connexion, etc.
- Système en ligne : un suivi temps-réel des partages des utilisateurs
- Algorithme incrémental afin de suivre les mises à jour constantes des folksonomies

Merci !  
Vos Questions ?