

Nouvelle Approche à base de Graphes pour les Systèmes de Recommandation Collaboratifs

Nassira CHEKKAI, Salim CHIKHI
Laboratoire MISC, Université Mentouri
Constantine, 25000, Algérie
n.chekkai@gmail.com, slchikhi@yahoo.com

Hamamache KHEDDOUCI
Laboratoire GAMA, Université Lyon1
hamamache.kheddouci@univ-lyon1.fr

Résumé—Les systèmes de recommandation collaboratifs sont des systèmes qui consistent à recommander à un utilisateur des ressources susceptibles de l'intéresser en se basant sur les avis d'autres utilisateurs qui partagent avec lui les mêmes centres d'intérêt. Cependant, ils souffrent du problème de démarrage à froid d'un nouvel usager qui commence avec un profil vide et rencontre une difficulté de communication avec les membres de sa communauté. Dans ce papier, nous proposons une nouvelle approche visant à pallier ce dernier problème en le modélisant graphiquement et en optant pour la détection des nœuds critiques afin d'identifier les meilleurs représentants de chaque communauté sociale qui aident à faire adapter un nouvel usager.

Mots clés- communautés ; graphes ; problème de démarrage à froid ; problème des nœuds critiques ; scores de confiance ; systèmes de recommandation collaboratifs.

I. INTRODUCTION

Nos aptitudes et besoins en communication, en échange d'idées et en partage d'informations ont nécessité l'apparition de nouvelles fonctions et applications sur les réseaux sociaux, comme en témoignent les *Systèmes de Recommandation Collaboratifs* appelés aussi «*Systèmes de Filtrage Collaboratifs*», ces derniers ont pour principe de recommander à un utilisateur des ressources susceptibles de l'intéresser en se basant sur des votes laissés par des utilisateurs aux profils proches du sien sur ces mêmes ressources.

De ce fait, chaque usager est identifié par un profil comportant ses coordonnées personnelles, ses centres d'intérêt et ses évaluations données aux ressources visitées. Ainsi, les utilisateurs qui partagent un même centre d'intérêt forment une communauté virtuelle où ils sont liés les uns aux autres par des liens pondérés dits « scores de confiance ».

À l'heure actuelle, les systèmes de recommandation collaboratifs prennent une place très importante dans le monde Réseau Sociaux et sont largement investis dans divers domaines.

Toutefois, ils souffrent de certains problèmes, parmi lesquels on peut citer le *démarrage à froid* dans lequel un nouvel utilisateur commence avec un profil vide et risque donc de ne plus recevoir des ressources de la part des usagers qui lui partage le même goût. Dans ce cas, la question qui se pose est la suivante : «Comment le système peut-il positionner un nouvel utilisateur dans une communauté formée sur la base de

l'historique des évaluations alors que cette personne n'a pas encore fourni d'évaluations ?» [4].

Pour pallier ce problème, nous présentons dans ce papier une idée originale permettant d'analyser le système de recommandation collaboratif par un graphe dans lequel toute principale notion de la recommandation collaborative trouve son équivalent. Ainsi, nous introduisons le *Problème des Nœuds Critiques* dit aussi CNP «Critical Node Problem» pour remédier au problème de démarrage à froid d'un nouvel utilisateur, en détectant les connecteurs importants dans chaque communauté du système.

La suite de cet article est organisée de la façon suivante. Dans un premier temps (section deux) nous décrivons la problématique à résoudre. La section trois sera consacrée à la description de l'approche proposée, qui sera suivie par la section quatre où nous montrerons l'enchaînement de notre méthode proposée et enfin nous terminerons par une conclusion et des perspectives à ce travail.

II. PROBLÉMATIQUE

Les systèmes de recommandation collaboratifs, sont nés de la volonté de pallier le problème de surcharge d'information du web combinant des techniques de filtrage d'information, personnalisation, intelligence artificielle, réseaux sociaux et interaction personne-machine [10]. Ces systèmes constituent une technique de personnalisation qui permet de faire des recommandations à un utilisateur en comparant ses préférences sous forme d'évaluations aux jugements portés par d'autres utilisateurs au profil similaire sur des ressources qui peuvent être de toute langue et de tout media (texte, audio, vidéo... etc.)

En effet, les systèmes de filtrage collaboratifs se basent sur l'hypothèse que les gens à la recherche d'information devraient pouvoir se servir de ce que d'autres ont déjà trouvé et évalué [5]. Selon cette optique, pour chaque utilisateur, un ensemble de contacts est identifié, et la décision de proposer ou non un document à un utilisateur dépendra des appréciations des membres de son groupe [8]. C'est ainsi qu'émerge la notion de communauté qui consiste à regrouper des utilisateurs partageant un même centre d'intérêt.

Cependant, les systèmes de recommandation collaboratifs souffrent du problème de démarrage à froid pour un *nouvel utilisateur* qui débute avec un profil vide est se retrouve isolé des membres de la communauté, *une*

nouvelle ressource qui risque de ne pas être visitée ou *un nouveau système* où les données sont très rares et les performances sont mauvaises [6].

Pour pallier ces problèmes, plusieurs approches ont été proposées telle que l'approche de Z. Zaier [10] qui a étudié les défis que posent les systèmes de recommandation à savoir le démarrage à froid et a proposé une nouvelle architecture pair-à-pair de recommandation collaborative, basée sur la discrimination du voisinage entre les usagers.

De leur part, A. Nguyen, N. Denos, et C. Berrut [4] ont adopté un processus de formation de communautés d'utilisateurs qui exploite un algorithme de positionnement en deux dimensions pour positionner les utilisateurs et un algorithme classique de classification non supervisée afin d'obtenir de véritables cartes des communautés dans le but de remédier au problème du démarrage à froid.

Ainsi, il existe peu de travaux basés graphes pour la résolution de ce dernier problème. Parmi lesquels, on cite l'approche de Z. Huang, H. Chen, et D. Zeng [9] qui se sont concentrés sur le démarrage à froid d'un nouveau système. Ils ont proposé en effet l'utilisation de la recherche associative en modélisant le problème du démarrage sous forme d'un graphe biparti composé de deux ensembles de nœuds à savoir les utilisateurs et les documents. À cet effet, le parcours des associations transitives entre utilisateurs se fait selon le processus activation/propagation. Intuitivement, plus le chemin entre deux utilisateurs est petit et le nombre de chemin est grand, plus la corrélation est grande.

Selon la même optique et afin de pallier le problème du démarrage à froid d'un nouveau document, K.W. Leung, S.C. Chan, et F. Chung [7] ont proposé l'exploitation des règles d'association sur un graphe multi-niveau en ajoutant aux deux niveaux (utilisateurs et documents) un troisième niveau, celui des (attributs document). Cette solution hybride permet d'atteindre des informations plus précises sur la nature des liens. Ainsi en supportant que toute transaction portant sur un document porte aussi sur ses attributs, les auteurs proposent l'utilisation des règles d'associations Fuzzy dont l'objectif est la découverte d'associations hiérarchiques sur les différents niveaux du graphe.

Cependant, nous constatons en passant en revue des travaux existants portant sur la résolution du problème de démarrage à froid d'un nouvel utilisateur dans un système de recommandation collaboratif, qu'il n'existe pas de solutions basées sur des techniques avancées des graphes.

Dans notre proposition, nous adoptons la modélisation graphique qui s'avère très appropriée pour l'analyse de ce dernier problème.

Nous proposons en effet de pallier ce dernier problème à travers le paramètre CNP qui consiste à un ensemble de nœuds à retirer d'un graphe pour avoir un maximum de composantes connexes. Ce dernier paramètre est lié à la notion de représentant de la communauté sociale afin d'y intégrer et adapter de nouveaux usagers.

Dans la section suivante nous présentons et détaillons cette méthode proposée.

III. NOTRE NOUVELLE APPROCHE PROPOSÉE

L'objectif du CNP est de trouver un ensemble de k -nœuds dans un graphe, dont leur suppression donne une fragmentation maximale du graphe [2]. De ce fait, ces derniers nœuds y jouent le rôle du médiateur parce que leur enlèvement le déconnecte complètement. De plus, l'identification de ces derniers nœuds dans un graphe social est importante pour comprendre ses caractéristiques structurelles et ses propriétés de connectivité.

C'est pourquoi, l'ensemble de k -nœuds critiques peut représenter dans un système de recommandation collaboratif l'ensemble des usagers permettant de déléguer efficacement leur communauté en s'occupant de ses activités sociales à savoir l'adaptation d'un nouvel utilisateur.

En effet, nous avons mentionné auparavant qu'un nouvel utilisateur étant inscrit dans une communauté d'un système de filtrage collaboratif débute avec un profil vide et les recommandations pour lui sont par conséquent très mauvaises.

En représentant la communauté par un ensemble de délégués, ses derniers parce qu'ils ont une grande capacité d'intermédiarité et de connectivité, ils vont lui permettre de s'y intégrer parce qu'ils vont le mettre en contact avec d'autres utilisateurs lui partageant le même goût et les mêmes attentes. De plus, ils lui permettent aussi d'évoluer ses intérêts et découvrir de nouveaux champs en le mettant en communication avec leurs différents contacts ce qui remédie au problème de démarrage à froid.

D'une façon plus parlante, nous proposons de représenter chaque communauté par un ensemble d'usagers volontaires que nous modélisons avec des «Nœud-Représentants».

Par la suite, lorsqu'un nouvel usager vient choisir une communauté en sélectionnant le centre d'intérêt correspondant, il se retrouvera devant une liste comportant les profils de ses représentants et une fois l'utilisateur signale ceux qui semblent intéressants pour lui il leur contactera. Ces derniers vont lui recommander des ressources, lui mettre en contact avec d'autres usagers qui lui partagent le même goût et lui recommander d'autres membres préférant d'autres champs d'intérêt.

De ce fait, chaque communauté du système a besoin des représentants efficaces et bien choisis. En modélisant ce problème à l'aide d'un graphe où les nœuds doivent posséder les propriétés suivantes :

✓ Ils doivent être les nœuds qui jouent les rôles des médiateurs et des connecteurs dans leur communauté. Selon cette optique, ils ont une idée sur tous les groupes qu'ils relient et par conséquent, ils pourront mettre le nouvel usager en contact avec les voisins des tranches adéquates à ses intérêts. ces voisins vont le mettre de leur part en contact avec leurs accointances.

✓ Ils sont alors des nœuds qui ont une grande puissance de liaison des membres de la communauté. De ce fait, leur

enlèvement va engendrer une fragmentation maximale de la communauté.

En passant en revue des concepts enrichissant de la théorie des graphes, nous constatons que le seul propriétaire des caractéristiques citées ci-dessus est le paramètre de graphe appelé « Problème des Nœuds Critiques ». "Un nœud est dit critique si sa suppression déconnecte le graphe" [3].

En effet, les paramètres de graphe induisent des structures intéressantes pour la compréhension et l'analyse des réseaux sociaux et le CNP qui rassemble toutes les caractéristiques signalées précédemment, offrira certainement une représentation efficace de chaque communauté du système de filtrage collaboratif.

Imaginons par exemple que nous avons des équipes de recherche scientifique dont chaque équipe représente une communauté «centre d'intérêt» et les membres sont les utilisateurs du réseau et ont des intérêts dans plusieurs domaines complémentaires.

Ainsi, lorsqu'une équipe reçoit un nouvel élément, soit un nouveau doctorant par exemple. Ce dernier risque de ne pas pouvoir communiquer tôt avec les membres de l'équipe. En effet, il risque de ne pas recevoir des recommandations de leur part parce qu'il est un nouvel élément inconnu. Si c'est le cas, il pourra contacter les représentants ayant les profils les plus intéressants pour lui, ces derniers vont lui recommander de la documentation, lui mettre en contact avec les chercheurs qui travaillent sur des sujets similaires au sien -parmi leur contacts- et lui recommander des chercheurs d'autres domaines afin qu'il puisse enrichir ses connaissances et évoluer ses intérêts.

Nous proposons, à cet effet, de modéliser ceci par un graphe pondéré dont chaque communauté soit représentée par l'ensemble des nœuds critiques les plus efficaces, cette efficacité dépend de plusieurs critères que nous allons détailler ci-après.

IV. MODELISATION GRAPHIQUE DE L'APPROCHE PROPOSÉE

Nous proposons de représenter notre réseau social «système de recommandation collaboratif» par un graphe pondéré dans lequel nous représentons:

- Chaque utilisateur du système par un nœud ;
- Chaque relation entre deux utilisateurs par une arête ;
- Chaque score de confiance entre deux utilisateurs par un poids (pondération de l'arête qui les relie) ;
- Chaque communauté (représentant un centre d'intérêt) par une couleur ;
- Un nœud peut appartenir à plus d'une communauté. Les arêtes peuvent donc être multiples ;
- L'arête prend la couleur de la communauté correspondante ;
- Et les nœuds des usagers volontaires pour la représentation de leur communauté prennent sa couleur.

La figure suivante schématise notre proposition.

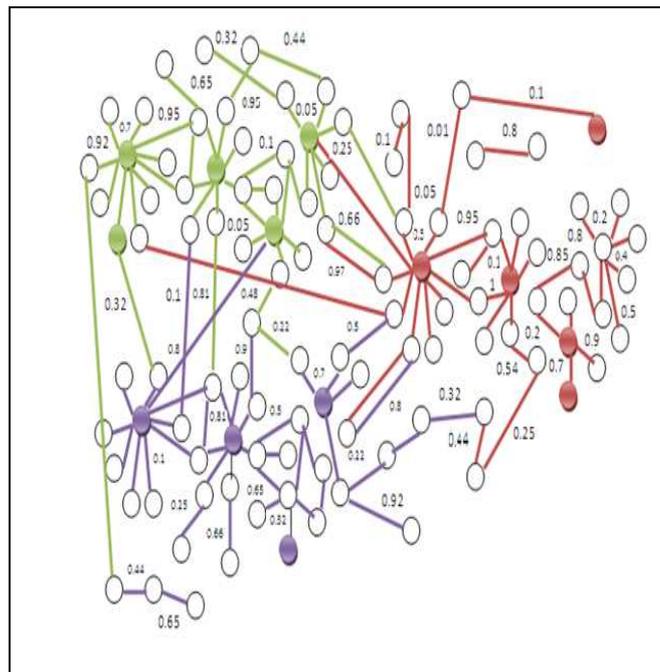


Figure 1. Modélisation du système de recommandation collaboratif par un graphe.

Il est à noter que les pondérations des arrêtes modélisent en quelques sortes des relations sociales de similarité entre les goûts des individus qui ne sont pas constants. De ce fait, ces pondérations expriment la confiance mutuelle entre deux acteurs, le lien qui relie ces derniers est alors bidirectionnel c'est à dire l'indice de confiance de l'acteur a vers l'acteur b est-il le même que celui de l'acteur b vers a.

Nous avons choisi de limiter les valeurs de ces indices entre 0 et 1, où le 1 correspond à un lien très fort entre deux utilisateurs (relation fructueuse) et le 0 correspond à une relation non fructueuse.

De plus, ces scores de confiances peuvent varier dans le temps en se renforçant ou contrairement se diminuant au fur et à mesure des interactions au sein des différentes communautés lors d'échange de ressources ou des idées entre les utilisateurs ce qui génère une évolution dynamique du graphe.

V. ENCHAINEMENT DE LA METHODE PROPOSÉE

Afin d'identifier les nœuds critiques les plus efficaces dans chaque communauté à part, nous proposons de suivre la démarche suivante :

A. Coloration des nœuds modélisant les usagers volontaires pour la représentation de la communauté

Nous proposons d'étiqueter les nœuds indiquant les personnes volontaires voulant représenter leur communauté par une couleur grise. Comme le montre la figure suivante.

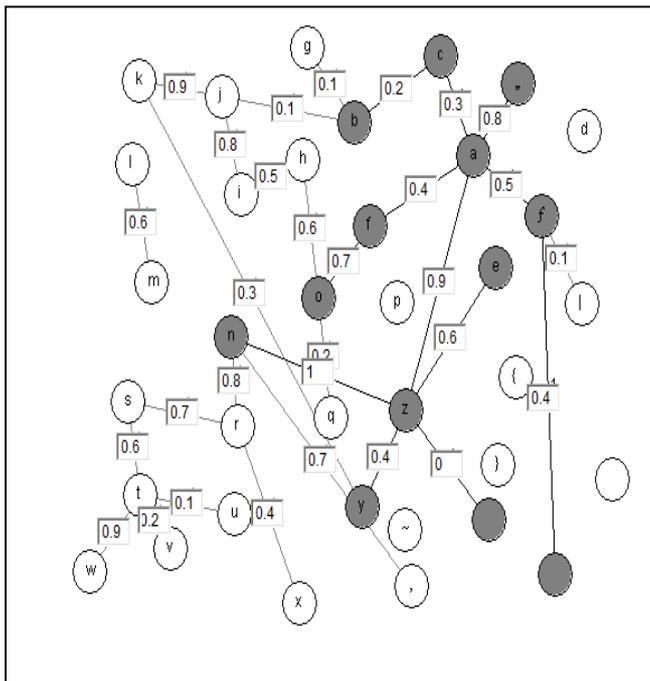


Figure 2. Graphe représentant les nœuds modélisant les usagers volontaires.

B. Détection des nœuds critique

La définition formelle du problème des nœuds critique est donnée par :

Entrée : un graphe non orienté $G(V, E)$ et un entier k .

Sortie : $A \leftarrow \arg \min \sum_{i,j \in (V \setminus A)} u_{ij} (G(V \setminus A)) : |A| \leq k$, où

$$U_{ij} : \begin{cases} 1, & \text{si } i \text{ et } j \text{ se trouvent dans la même composante de } \\ & G(V \setminus A), \\ 0, & \text{sinon [1]} \end{cases}$$

L'objectif est de trouver un sous ensemble A de nœuds appartenant à l'ensemble V qui est en effet l'ensemble de tous les nœuds du graphe G , tel que A est inférieur ou égal à un entier k précisé à l'entrée et dont la suppression de ses nœuds minimise la connectivité entre les nœuds restants du graphe $G(V \setminus A)$.

Dans cette étape, nous exploitons l'heuristique de A. Arulsevan, C. W. Commander, P. M. Pardalos, et O. Shylo [2] pour la détection des nœuds critiques parmi les nœuds colorés en gris afin d'identifier les représentants efficaces de la communauté parmi les utilisateurs volontaires inscrits.

L'heuristique utilisée est décrite comme suit :

```

procedure NoeudsCritiques(G,k)
1. MIS ← EnsembleIndépendantMaximal(G)
2. tant que (MIS ≠ |V|-k) faire
3.  $i \leftarrow \arg \min \{ \sum_{s \in S} \frac{s_i(s_i-1)}{2} : S \in G(MIS \cup \{i\}), i \in V \setminus MIS \}$ 
4. MIS ← MIS ∪ {i}
5. fin tant que
6. retourner V \ MIS /*ensemble de k nœuds critiques*/
fin procedure NoeudsCritiques

```

Où :

G : est le graphe général;

V : est l'ensemble des nœuds du graphe;

MIS : est l'ensemble indépendant maximal du graphe ;

k : est le nombre entier des nœuds critique;

S : est l'ensemble de toutes les composantes connexes du graphe.

L'heuristique commence alors par trouver le MIS du graphe, puis elle sélectionne les nœuds qui minimisent la fonction 'objectif' du graphe $G(MIS + \{i\})$ et les ajoute au MIS Lorsque MIS soit égale au $|V| - k$, à ce moment là l'heuristique se termine et l'ensemble des nœuds critiques est l'ensemble des nœuds j tel que j s'appartient à $V \setminus MIS$.

La figure suivante montre l'identification de l'ensemble des nœuds critiques dans le graphe précédent en appliquant l'heuristique expliquée ci-dessus.

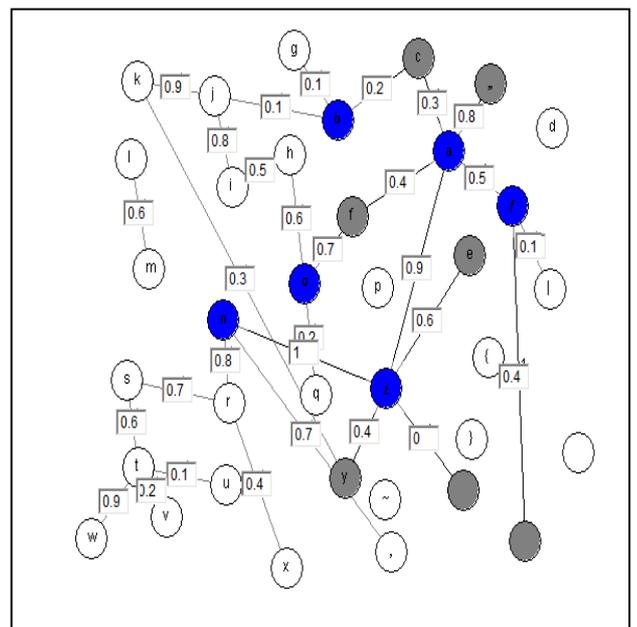


Figure 3. Détection des nœuds critiques.

Il est à noter que nous avons choisi de représenter les communautés par des noeuds critiques au lieu des noeuds denses parce qu'un noeud critique a l'avantage de contacter tous les groupes d'une même communauté tant qu'il est l'intermédiaire entre eux. Bien qu'un noeud dense n'a pas cette puissance. En effet, ce dernier possède un grand nombre de connectivité dans sa communauté, cependant, il n'est pas forcément l'intermédiaire.

Ainsi, nous avons opté pour le choix des nœuds critiques au lieux des nœuds centraux parce que la suppression de ces derniers ne garantie pas une fragmentation du réseau [2].

C. Identification des nœuds critiques les plus efficaces

Du fait que les arrêtes du graphe social proposé sont pondérées. Il faut donc tenir compte de leurs pondérations qui représentent les scores de confiance entre les utilisateurs d'un système de filtrage collaboratif.

À cet effet, nous gardons dans cette étape que les nœuds critiques ayant des bons scores de confiance avec leurs accointances ce qui garantit une bonne représentation de la communauté.

Si nous optons par exemple pour le choix de quatre représentants, nous calculons alors la moyenne de pondérations des arêtes de chaque nœuds critique à part. Par la suite nous choisissons les quatre nœuds ayant les meilleures moyennes.

Le tableau suivant classifie les nœuds critiques du graphe précédent selon la moyenne de pondérations des arêtes.

TABLE I. CLASSEMENT DES NOEUDS CRITIQUES SELON LA MOYENNE DE PONDERATIONS DES ARETES

<i>Nœud critique</i>	<i>Moyenne de pondérations de ses arêtes</i>
n	0,83
z	0,58
a	0,50
o	0,50
f	0,33
b	0,13

En se référant aux résultats présentés dans le tableau ci-dessus, les nœuds critiques modélisant les représentants de la communauté sont (n, z, a et o).

Il est à noter finalement que ces derniers nœuds vont être remplacés par d'autres nœuds critiques choisis de la même manière après une durée précise pour les raisons suivantes :

- Leurs liens avec leurs accointances peuvent devenir non fructueux au fur et à mesure des interactions dans le graphe social.

- Les usagers correspondants à ces nœuds peuvent choisir de quitter la communauté pour la raison précédente ou même pour d'autres raisons.

- Les usagers correspondants pourront aussi choisir de ne plus représenter la communauté.

De ce fait, le graphe social va être évolué périodiquement en plus de son évolution dynamique.

VI. CONCLUSION

Tout au long de cet article, nous avons présenté une nouvelle approche visant à pallier le problème du démarrage froid d'un nouvel usager dans un système de recommandation collaboratif en optant pour le problème des nœuds critiques.

Nous avons proposé en effet de représenter chaque communauté correspondante à un centre d'intérêt par un

ensemble de nœuds critiques ayant la capacité de lier le graphe social de la communauté et y faire adapter un nouveau nœud représentant un nouvel usager tout en le mettant en contact avec des nœuds semblant importants pour lui.

Pour ce faire, nous avons présenté une méthode débutant avec l'identification des nœuds correspondants aux volontaires voulant représenter leur communauté, et en extraire par la suite les nœuds critiques. La dernière étape sert à comparer les moyennes de pondérations des arêtes des nœuds critiques afin d'identifier les nœuds critiques les plus efficaces correspondants bien-sur aux bonnes moyennes ce qui assure une meilleure représentation de la communauté et donc une meilleure prise en charge des nouveaux inscrits dans le système ce qui remédie au problème du démarrage à froid.

Cette étude est préliminaire. Les premiers résultats sur des graphes de taille moyenne sont intéressants. Notre objectif est de la tester sur des graphes sociaux connus et de la comparer à d'autres méthodes existantes.

Nous comptons aussi tester notre approche sur un graphe orienté.

REFERENCES

- [1] A. Arulselvan, "Network Models for Disaster Management," Report written for Candidacy Admission Examination, Industrial and Systems Engineering, University of Florida, 2008.
- [2] A. Arulselvan, C. W. Commander, P. M. Pardalos, and O. Shylo, "Managing network risk via critical node identification," Risk Management in Telecommunication Networks, N. Gulpinar and B. Rustem (editors), Springer, 2009.
- [3] A. Casteigtsyz et L. Bouchardz, "Intégration de la Dimension Temporelle dans l'Analyse des Réseaux de Santé en Français," Rapport préliminaire pour le compte du Réseau de Recherche ésur la Santé des Francophones en Ontario (RRASFO), Ecole d'Ingénierie et de Technologie de l'Information et Institut de Recherche en Santé des Populations, Université d'Ottawa, 2010.
- [4] A. Nguyen, N. Denos, et C. Berrut, "Exploitation des données disponibles à froid pour améliorer le démarrage à froid dans les systèmes de filtrage d'information hybride," Actes de la 3ème Conférence en Recherche Information et Applications (CORIA'06), 303-314, 2006.
- [5] C. Berrut et N. Denos, "Filtrage collaboratif. Assistance intelligente à la recherche d'information," Chapitre 8, Hermes-Lavoisier, 2003, pp. 255-283.
- [6] H. Oufaida, "Amélioration du processus de démarrage à froid dans les systèmes de filtrage d'information collaboratif," Mémoire En vue de l'obtention du diplôme de Magister En Informatique, Ecole nationale Supérieure d'Informatique (E.S.I), 2008-2009.
- [7] K.W. Leung, S.C. Chan, and F. Chung, "Applying Cross-Level Association Rule Mining to Cold-Start Recommendations," IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology Workshops, 2007.
- [8] W. Gao, S. Wang, and N. Cerrone, "A dynamic recommendation system based on log mining," In International journal of foundations of computer science, Vol. 13, N° 4, 2002, pp. 521-530.
- [9] Z. Huang, H. Chen, and D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsty Problem in Collaborative filtering," ACM Trans. Information Systems, vol. 22, N° 1, 2002, pp. 116 142.
- [10] Z. Zaier, "Modèle multi-agent pour le filtrage collaboratif de l'information," Thèse de doctorat de l'université du Québec à Montréal, 2010.