

---

# Approche hybride de recommandation des nouvelles collaborations

Nesrine BEN YAHIA<sup>1</sup>, Wala REBHI<sup>2</sup>, Narjès BELLAMINE<sup>3</sup>,  
Henda BEN GHEZELA<sup>4</sup>

<sup>1</sup> RIADI-ENSI, Université de la Manouba, nesrine.benyahia@ensi.rnu.tn

<sup>2</sup> RIADI-ENSI, Université de la Manouba, wala.rebhi@ensi.rnu.tn

<sup>3</sup> RIADI-ENSI, Université de la Manouba, narjes.belamine@ensi.rnu.tn

<sup>4</sup> RIADI-ENSI, Université de la Manouba, henda.benghazala@ensi.rnu.tn

---

*RÉSUMÉ.* Suite à l'émergence des organisations et en particulier celles distribuées, les groupes sont très dynamiques et les gens sont constamment à la recherche de nouveaux collaborateurs. Dans ce contexte, nous proposons une nouvelle approche de recommandation des collaborations qui cherche les collaborateurs potentiels et pertinents. Cette approche se base sur un nouvel algorithme de détection de communautés pour composer dynamiquement les groupes en considérant simultanément les interactions sociales entre les membres de l'organisation et leurs homophilies. L'approche proposée a été expérimentée sur un graphe de terrain représentant un laboratoire de recherche en le considérant comme une organisation où les membres représentent les chercheurs et les interactions représentent leurs relations de co-publication et de co-encadrement.

*ABSTRACT.* With the spread of organizations and especially distributed ones, teams are highly dynamic and people are continually seeking for new collaborators. In this context, we propose an approach of collaboration recommendation that relies on community detection technique to find potential collaborators that can help in problem solving. Our contribution consists on proposing a combined metric that considers both social connections and homophily of participants. The proposed approach has been tested on a real graph representing a research laboratory where members represent researchers and interactions represent their relationships of co-publishing and co-supervision.

*MOTS-CLÉS :* recommandation de collaborations, détection de communautés, interaction sociale, homophilie.

*KEYWORDS:* collaboration recommendation, community detection, social connections, homophily.

---

## 1. Introduction

Le travail des organisations est en grande partie un travail de prise de décisions et de résolution de problèmes. En outre, avec l'émergence des organisations et en particulier celles distribuées, les équipes sont de plus en plus très dynamiques et les gens sont constamment à la recherche de nouveaux collaborateurs pour les aider dans la résolution de leurs problèmes (Matthews *et al.*, 2012). Aujourd'hui, le concept de communauté de pratiques est devenu un objectif d'analyse utile pour identifier et caractériser « les collectivités » (groupes de personnes) pertinentes qui participent à la création des bonnes pratiques et la génération des solutions innovantes via le partage des expériences et la socialisation (Ackerman *et al.*, 2013).

Ainsi, la collaboration est importante pour les organisations qui se veulent compétitives et comment les collaborateurs sont sélectionnés est une chose cruciale pour prendre les bonnes décisions (Matthews *et al.*, 2012). Dans cette direction, des efforts techniques ont été élaborés pour soulever le problème d'identification des collaborateurs les plus pertinents mettant en place des systèmes de recommandation des collaborations (Ackerman *et al.*, 2013).

Dans cet article, nous nous intéressons à la détection et la décomposition dynamique des communautés au sein des organisations pour la recommandation des nouvelles collaborations pour un membre cherchant les personnes pertinentes qui peuvent l'aider dans la résolution de son problème. Pour ce faire, nous proposons une approche hybride de détection de communautés qui prend en considération simultanément les interactions sociales et l'homophilie des individus. Cette recommandation ne cherche pas à identifier les collaborateurs les plus pertinents de façon indépendante mais plutôt elle cherche à identifier tout un groupe de collaborateurs (une communauté).

Cet article est organisé comme suit : dans la section 2, nous présentons un tour d'horizon de différentes techniques utilisées pour la recommandation de collaborations. La section 3 présente l'approche hybride proposée qui combine les interactions sociales et l'homogénéité des individus. Dans la section 4, une expérimentation et une évaluation de cette proposition seront présentées.

## 2. Une vue globale des techniques de recommandation des collaborations

Au sein d'une organisation, (Lu et Zhou, 2011) affirment que les personnes «similaires» sont susceptibles d'être connectées et font la distinction entre deux définitions du mot «similaire» : une similarité basée attributs et une similarité structurelle. La première considère le partage de nombreux traits en commun, quant à la deuxième considère exclusivement la structure du réseau. Dans cette section, nous présentons une revue des principales techniques qui ont été utilisées pour recommander des nouvelles collaborations selon ces deux points de vue.

### **2.1. Techniques basées sur la similarité structurelle**

S'appuyant sur le principe de la similarité structurelle, il y a une tendance à intégrer les réseaux sociaux dans les systèmes dans le but de faciliter la recherche des nouvelles collaborations. Dans ce contexte, certaines solutions utilisent la visualisation des réseaux sociaux comme une indication et un moyen de trouver une personne avec qui collaborer (Smith *et al.*, 2000). Par exemple, (Kautz *et al.*, 1997) utilisent les relations de copublication et de co-citation pour créer un réseau social et sa visualisation est utilisée pour trouver un expert possible. (Nardi *et al.*, 2002) proposent une application de gestion de communication qui utilise une visualisation de réseau social pour présenter des contacts et favoriser la collaboration entre les individus. D'autres solutions tentent d'utiliser les informations des réseaux sociaux pour recommander des personnes compétentes avec qui collaborer. Par exemple, (Ogata *et al.*, 2001) proposent de trouver des collaborateurs par le biais des réseaux sociaux et des échanges électroniques dans une organisation en considérant un collaborateur potentiel en tant que candidat qui échange fréquemment, récemment et réciproquement des e-mails avec le demandeur de collaboration.

### **2.2. Techniques basées sur la similarité des attributs**

Dans ce contexte, un phénomène appelé homophilie indique que les gens ont tendance à collaborer avec d'autres qui ont les mêmes intérêts et les caractéristiques démographiques similaires (McPherson *et al.*, 2001). Différents travaux se sont basés sur ce principe d'homophilie pour sélectionner des collaborateurs potentiels, mais ils ont utilisé différentes techniques pour trouver ces personnes. (Liben-Nowell et Kleinberg, 2003) ont examiné les différentes méthodes pour la prédiction des liens et ont affirmé que la méthode des voisins communs (la similarité entre deux noeuds  $x$  et  $y$  est le nombre de voisins que  $x$  et  $y$  ont en commun) et le coefficient de Jaccard (métrique très connue dans le domaine de recherche d'information qui permet de mesurer la probabilité que les deux noeuds  $x$  et  $y$  ont une même caractéristique pour une caractéristique choisie au hasard) donnent les meilleurs résultats dans la plupart des cas. (Dong *et al.*, 2012) ont utilisé également la technique de prédiction des liens (link prediction) en se basant sur des mesures de similarité pour analyser la proximité des noeuds (représentant les personnes) dans un réseau. Et, ils ont conclu que les noeuds très similaires sont plus susceptibles d'être reliés dans le futur réseau. Avec le même principe d'homophilie, d'autres travaux ont utilisé la technique de filtrage collaboratif classique utilisée dans les systèmes de recommandation des items (tels que les films) en se basant sur les opinions (aimer ou non) des utilisateurs afin d'aider les personnes dans le choix des éléments. (Cai *et al.*, 2010) ont réutilisé ce principe en se basant sur deux aspects de similarité : deux personnes sont similaires (1) si elles sont sélectionnées par un ensemble d'utilisateurs en commun ou (2) si elles sélectionnent à la fois un ensemble d'utilisateurs en commun.

### 2.3. Synthèse des techniques de recommandation des collaborations

La combinaison des deux points de vue cités précédemment assume une approche de recommandation basée sur une similarité des attributs et une similarité structurale. En outre, et comme les communautés sont considérées selon (Wenger et Snyder, 2000) comme l'un des moyens les plus rapides pour la résolution de problèmes dans les organisations, nous proposerons dans la section suivante une approche hybride de recommandation des nouvelles collaborations en nous basant sur la détection de communautés et en considérant simultanément les interactions sociales (la structure de l'organisation) et l'homophilie des individus (la similarité des attributs).

## 3. Proposition d'une approche hybride de recommandation des collaborations

L'approche proposée se base sur la combinaison d'une technique de fouille de réseaux sociaux i.e. la détection de communautés et d'une méta-heuristique i.e. l'optimisation par essaim particulière pour l'optimisation de la qualité de la communauté détectée. Nous allons commencer par la présentation de ces deux techniques. Ensuite, nous allons présenter notre contribution qui réside dans la proposition d'un nouvel algorithme d'optimisation de détection de communautés pour la recommandation des nouvelles collaborations (estimées les plus pertinentes).

### 3.1. Techniques utilisées

#### 3.1.1. La détection de communautés (DC)

La détection de communauté est considérée comme une technique de classification non supervisée qui est multidisciplinaire où l'utilisation du partitionnement de graphe, l'informatique, et la sociologie sont combinés (Newman et Girvan, 2004). Elle vise à diviser les noeuds du réseau (représentant l'organisation) en groupes (communautés) avec un faible couplage (relations extérieures entre les communautés) et une forte cohésion (relations internes au sein des communautés). Afin d'évaluer la qualité de la détection de communautés, (Newman et Girvan, 2004) ont proposé une métrique, appelée la modularité qui doit être maximisée pour résulter un bon partitionnement. Elle représente, pour un partitionnement donné, la différence entre la fraction d'arêtes dans le réseau considéré et la probabilité (la valeur attendue) de la même quantité dans un réseau avec le même partitionnement mais avec une répartition au hasard des arcs en respectant les degrés des noeuds.

La formule de cette métrique est donnée par (1) où  $n$  représente le nombre des sommets du réseau,  $A_{ij}$  est le nombre d'arêtes entre les sommets  $i$  et  $j$  (qui est normalement 0 ou 1), donc  $A_{ij}$  sont les éléments de la matrice d'adjacence.  $A_{ij}$  représentent la proportion d'arêtes avec un sommet dans la classe (communauté)  $C_i$  et l'autre dans la classe  $C_j$ . Ensuite, le nombre prévu d'arêtes entre les sommets  $i$  et  $j$  si les arêtes sont placées aléatoirement est  $\frac{k_i k_j}{2m}$ , où  $k_i$  et  $k_j$  représentent respectivement les degrés de sommets  $i$  et  $j$ ,  $m$  est le nombre total d'arêtes dans le réseau et  $\delta(C_i, C_j)$  est le

symbole de Kronecker qui vaut 1 si les arguments  $C_i$  et  $C_j$  sont égaux et 0 sinon ( $C_i$  représente la communauté qui contient les sommets  $i$ ). Pour les graphes pondérés, qui possèdent des poids sur les arêtes, la même formule de modularité est utilisée et  $A_{ij}$  devient le poids de l'arête entre  $i$  et  $j$ ,  $k_i$  est la somme des poids des arêtes attachées aux sommets  $i$  et  $m$  est égal à la somme des poids des arêtes (Blondel *et al.*, 2008).

$$M(P) = \frac{1}{2m} \sum_{i,j} [A_{ij} - \frac{k_i k_j}{2m}] \delta(C_i, C_j) \quad (1)$$

Dans le cadre de cet article, nous avons sélectionné cette technique parce que la modularité permet de combiner la technique de la similarité structurelle (en considérant les interactions sociales) et la technique de la similarité basée attributs (en considérant l'homophilie entre les sommets). Toutefois, dans notre contexte, notre objectif est de déterminer pour un membre donné (représenté par un sommet dans le réseau) le groupe (la communauté) le plus pertinent qui peut l'aider dans la résolution de son problème. Donc, nous avons besoin d'une nouvelle métrique qui prend en considération deux types de poids : les poids des arcs et les poids des noeuds. Le poids d'un arc représente le degré de similarité entre les sommets reliés par cet arc. Quant au poids d'un noeud, il représente le degré de similarité entre ce noeud et le noeud représentant le membre demandant l'aide même si ces deux noeuds ne sont pas reliés par un arc. Nous avons présenté dans cette section la première technique que nous allons utiliser pour la recommandation des nouvelles collaborations. Dans la section suivante, nous allons présenter la seconde technique.

### 3.1.2. L'optimisation par essaim particulière (OEP)

Comme l'optimisation de la modularité est typiquement une optimisation NP-difficile qui nécessite un temps d'exécution exponentiel pour trouver la solution optimale (Leskovec *et al.*, 2010), plusieurs travaux se sont basés sur l'utilisation des (méta) heuristiques pour trouver rapidement une solution réalisable et approchée, pas nécessairement la solution optimale exacte, pour un problème d'optimisation difficile. Nous citons par exemple l'utilisation du recuit simulé dans (Guimera *et al.*, 2004), l'algorithme glouton (Newman et Girvan, 2004) et (Blondel *et al.*, 2008), les colonies de fourmis dans (Sadi *et al.*, 2009) et (He *et al.*, 2011) et l'algorithme génétique dans (Pizzuti, 2008) et (Mazur *et al.*, 2010). L'enjeu de toutes ces techniques d'optimisation est de minimiser le temps de convergence de façon qu'elles fournissent rapidement une solution. Cependant, des études empiriques, réalisées dans (Parsopoulos et Vrahatis, 2010) et (El dor, 2012), ont confirmé que, l'optimisation par essaim particulière (OEP), qui consiste en une méta-heuristique pour l'optimisation NP-difficile (Clerc et Siarry, 2004), est plus rapide que les techniques citées précédemment et elle donne de résultats meilleurs. Dans ce qui suit, nous allons expliquer le principe et l'algorithme d'OEP.

L'optimisation par essaim particulière (OEP) est introduite dans (Eberhart et Kennedy, 1995) comme une technique évolutive de computation inspirée de la simulation du comportement social des oiseaux ou de poisson. La population dans l'OEP est appelée un essaim où les individus, les particules, sont candidats à des solutions du pro-

blème d'optimisation dans l'espace de recherche multidimensionnel (D dimensions). Chaque particule propose une solution et évalue sa proposition en calculant la fonction objective (fitness) qui doit être optimisée.

L'algorithme d'OEP est itératif et le nombre d'itérations n'est pas fixe, il peut être obtenu de façon empirique. Pour chaque itération t, chaque particule i est caractérisée par sa position  $\vec{x}_i(t)$  et sa vitesse  $\vec{v}_i(t)$  qui sont mises à jour de manière synchrone dans chaque itération de l'algorithme. Une particule ajuste sa vitesse en fonction de sa propre expérience de vol (la solution optimale dans son historique) et l'expérience de vol d'autres particules (la solution optimale pour l'essaim entier) de telle manière qu'elle accélère à des positions qui permettent la convergence rapide vers la solution optimale. Par conséquent, il existe deux types de position vers laquelle une particule est accélérée. La première est la meilleure position personnelle d'une particule obtenue jusqu'à l'itération courante, appelée  $Pbest_i$ . L'autre est la meilleure position globale obtenue à l'itération courante par toutes les particules, appelée Gbest. La position  $\vec{x}_i(t+1)$  et la vitesse  $\vec{v}_i(t+1)$  à l'itération t+1 pour une particule i peuvent être calculées en utilisant les formules suivantes :

$$\vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t)(2)$$

$$\vec{v}_i(t+1) = \omega\vec{v}_i(t) + c_1r_1[x\vec{p}_i(t) - \vec{x}_i(t)] + c_2r_2[g(\vec{t}) - \vec{x}_i(t)](3)$$

Les paramètres  $\omega$ ,  $c_1$ , et  $c_2$  sont des coefficients constants fixés par l'utilisateur,  $r_1$  et  $r_2$  sont des nombres aléatoires tirés à chaque itération,  $g(\vec{t})$  est la meilleure solution trouvée jusqu'à l'instant t et  $x\vec{p}_i(t)$  est la meilleure solution trouvée par le particule i. C'est le vecteur vitesse qui dirige le processus de recherche et reflète la "sociabilité" des particules.

L'algorithme d'OEP peut se résumer dans le pseudo-code suivant :

- 1) Les paramètres de l'algorithme sont initialisés,
- 2) Les positions et les vitesses initiales des particules sont générées en utilisant des valeurs aléatoires,
- 3) Chaque particule évalue sa position (sa solution) en calculant la fonction objective,
- 4) La position et la vitesse de chaque particule sont mises à jour en utilisant(2) et(3),
- 5) Si le nombre maximal d'itérations est atteint, les solutions finales sont données et Gbest est considérée comme la solution optimale détectée. Sinon, l'algorithme boucle à 3 et ainsi de suite.

### 3.2. Proposition d'un nouvel algorithme combinant la détection de communautés et l'optimisation par essaim particulière

Nous allons présenter dans cette section notre contribution qui concerne la proposition d'un nouvel algorithme de recommandation des collaborations basé sur la combinaison de la détection de communautés et l'OEP. Cet algorithme cherche à optimiser une nouvelle fonction bi-objective qui tient compte à la fois les interactions sociales (les poids des arcs) et l'homophilie des membres (les poids des noeuds).

#### 3.2.1. Définition d'une nouvelle métrique bi-objective

Rappelons que notre objectif est de déterminer le groupe le plus pertinent qui peut participer dans la résolution de problèmes. Pour ce faire et étant donné un membre qui a un problème et qui cherche des collaborateurs pertinents, nous allons utiliser la détection de communautés pour identifier la communauté à laquelle appartient ce membre et ensuite lui recommander les différents membres de cette communauté.

Dans ce contexte, le graphe qui représente l'organisation, est pondéré par des poids des arcs et des noeuds. Le poids d'un arc représente le degré de similarité entre les sommets déjà reliés par cet arc. Quant au poids d'un noeud, il représente le degré de similarité entre ce noeud et le noeud représentant le membre demandant l'aide même si ces deux noeuds ne sont pas reliés par un arc.

Maintenant, pour prendre en considération ces deux types de poids dans une même métrique, nous proposons une nouvelle fonction bi-objective qui doit être optimisée pour obtenir un partitionnement de bonne qualité. Cette fonction se base sur deux composantes qui doivent être maximisées simultanément :

**La première composante** concerne la fréquence des interactions sociales entre les individus en se basant sur l'hypothèse que les personnes qui socialisent fréquemment (ont des interactions entre elles) sont plus susceptibles d'être des collaborateurs efficaces. Elle s'appuie sur la qualité structurelle (en termes de liens sociaux) pour évaluer le partitionnement. Pour cela, nous proposons de réutiliser la modularité de Newman pour les graphes pondérés.

**La deuxième composante** concerne la fréquence d'homophilie entre l'individu cherchant les collaborateurs et les autres individus en se basant sur l'hypothèse que les personnes qui ont des caractéristiques en commun (par exemple les centres d'intérêts ou les préférences) sont plus susceptibles d'être des collaborateurs efficaces. Elle s'appuie sur la qualité basée attributs (en termes de similarité des attributs) pour évaluer le partitionnement. Pour cela, nous proposons de réutiliser la notion de l'inertie (Lebart *et al.*, 1982).

L'inertie est une métrique qui permet de mesurer la dispersion d'un nuage pondéré (un ensemble de noeuds où chaque noeud possède un poids). Si nous considérons  $I_{total}$  l'inertie totale d'un nuage alors selon la formule de Huygens (4),  $I_{total}$  est composé de deux quantités : l'inertie intra-classes  $I_{intra}$  et l'inertie inter-classes  $I_{inter}$ .

$$I_{total} = I_{inter} + I_{intra} \quad (4)$$

L'inertie intra-classes  $I_{intra}$ , donnée par la formule (5), permet de mesurer le degré d'homogénéité entre les noeuds appartenant à la même classe. Elle calcule leurs distances par rapport au centre de gravité de la classe (le point représentant la moyenne de la classe). Dans la formule (3),  $p_i$  représente le poids de la classe  $C_i$  (qui est la somme des poids de tous les noeuds appartenant à la même classe  $C_i$ ). Ensuite,  $I_i$  représente l'inertie de chaque classe  $C_i$  (qui est la somme des carrés des distances entre chaque sommet dans la classe  $C_i$  et le centre de gravité  $g_i$  de  $C_i$  divisées par la cardinalité de  $C_i$ ). Enfin,  $d(x,y)$  est la distance entre  $x$  et  $y$  (qui est généralement de la distance euclidienne).

$$I_{intra} = \sum_i p_i I_i = \sum_i p_i \sum_{x_j \in C_i} \frac{d^2(x_j, g_i)}{|C_i|} \quad (5)$$

L'inertie inter-classes  $I_{inter}$ , donnée par la formule (6), permet de mesurer le degré d'hétérogénéité entre les classes. Elle calcule les distances entre les centres de gravité  $g_i$  de chaque classe  $C_i$  de la partition par rapport au centre de gravité global  $g$  de tout le nuage ( $g$  est le poids du nuage divisé par sa cardinalité).

$$I_{inter} = \sum_i p_i d^2(g_i, g) \quad (6)$$

Ensuite, une inertie inter-classes forte reflète l'hétérogénéité entre les classes et une inertie inter-classes faible reflète l'homogénéité au sein des classes. Pour obtenir une bonne partition du nuage, nous considérons le rapport  $I$  entre l'inertie inter-classes  $I_{inter}$  et l'inertie totale  $I_{total}$  qui doit être maximisé.

$$I = \frac{I_{inter}}{I_{total}} \quad (7)$$

Par conséquent, la nouvelle métrique  $Q$  que nous allons utiliser pour évaluer la qualité de la détection de communautés, donnée par (8), représente une fonction bi-objective. En effet, son objectif est de maximiser simultanément la modularité et l'inertie inter-classes de façon à obtenir des communautés denses et homogènes en termes d'interactions sociales fortes et en termes d'homophilie importante. Cette nouvelle métrique est basée sur une combinaison pondérée de  $M$  et  $I$  et doit être maximisée pour obtenir une bonne division du réseau ( $\alpha$  est un facteur de pondération dynamique où  $0 < \alpha < 1$ , si nous voulons obtenir l'équité entre la modularité et la proportion de l'inertie inter-classes, nous pouvons mettre  $\alpha$  à 0,5).

$$Q = \alpha M + (1 - \alpha) I = \alpha \frac{1}{2m} \sum_{i,j \in V} [A_{ij} - \frac{k_i k_j}{2m}] \delta(C_i, C_j) + (1 - \alpha) \frac{I_{inter}}{I_{total}} \quad (8)$$

Dans ce qui suit, nous allons montrer comment cette métrique va être exploitée et maximisée pour assister la recommandation des collaborations.

### 3.2.2. L'algorithme d'optimisation proposé

Comme nous l'avons déjà mentionné et justifié, nous proposons d'utiliser l'optimisation par essaim particulaire pour l'optimisation de la nouvelle fonction bi-objective proposée. Ainsi, l'algorithme de détection de communautés que nous proposons est un algorithme d'optimisation qui se base sur l'OEP.

L'objectif de cet algorithme consiste à partitionner l'organisation en deux clusters disjoints, l'un représente la communauté qui englobe le membre cherchant les collaborateurs et les membres qui lui sont similaires et l'autre représente le groupe qui englobe le reste de l'organisation. Pour déterminer les membres similaires (à recommander), nous avons besoin de : (1) identifier la fonction de similarité entre les membres pour calculer leurs degrés de similarité et (2) identifier un seuil au-dessus duquel un membre sera considéré comme similaire au membre demandeur.

En ce qui concerne la fonction de similarité, nous allons modéliser l'organisation et ses membres par un patron conceptuel ontologique (PCO) que nous avons proposé dans (Ben Yahia *et al.*, 2013). Ce PCO permet de représenter les caractéristiques de l'organisation (ses objectifs, sa structure et ses groupes) ainsi que les profils de ses membres. Donc, la fonction de similarité représente la similarité sémantique entre les concepts des profils des membres.

En ce qui concerne le seuil, nous allons utiliser l'OEP pour déterminer la meilleure valeur qui permet de maximiser la fonction bi-objective. Donc, pour le présent problème à optimiser, dans chaque itération, chaque particule propose un seuil puis évalue la fonction en se basant sur ce seuil.

Pour l'initialisation des paramètres de l'algorithme proposé, nous utilisons les valeurs suivantes :

- La population initiale est composée de 10 particules, le nombre d'itérations est de 1000,  $c_1 = 2,8$  et  $c_2 = 1,3$  en justifiant ces choix sur l'étude empirique réalisée dans (Carlisle et Dozier, 2001) qui confirme que ces valeurs sont les meilleures.

- Les positions et les vitesses initiales des particules sont générées en utilisant des valeurs aléatoires dans ( $W_{nmin}$ ,  $W_{nmax}$ ) où  $W_{nmin}$  et  $W_{nmax}$  sont respectivement les valeurs minimales et maximales des poids des noeuds.

- La valeur de  $Pbest_i$  est initialisée à la position initiale de la particule  $i$  et  $Gbest$  est initialisée à la meilleure valeur de tous les  $Pbest_i$ .

L'algorithme proposé peut se résumer comme suit. Après l'initialisation des paramètres :

- 1) Chaque particule prend une position.
- 2) Ensuite, elle évalue cette position en calculant la modularité  $M$  et la proportion d'inertie inter-classes  $I$  et elle déduit la valeur de  $Q$ .
- 3) Enfin, elle met à jour la valeur de  $Pbest_i$  (si elle a amélioré sa position sinon elle garde la valeur ancienne) et même chose pour  $Gbest$  (si une particule a donné une valeur  $Pbest_i$  mieux que  $Gbest$  donc celle-ci va être écrasée par  $Pbest_i$  sinon  $Gbest$  garde la valeur ancienne).

4) Si le nombre maximal d'itérations est atteint, Gbest représente la meilleure solution obtenue par tout l'essaim. Sinon, chaque particule va refaire le même travail.

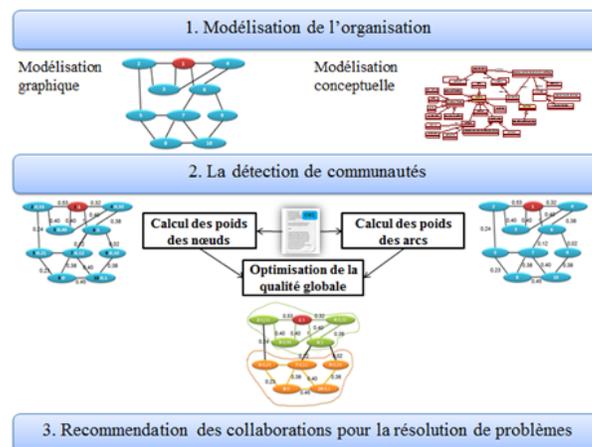
### 3.3. Présentation de l'approche de recommandation proposée

L'approche de recommandation des collaborations, illustrée par la figure 1, que nous proposons se base sur trois phases :

1) La première phase concerne la modélisation de l'organisation. En fait, cette dernière est modélisée graphiquement par un graphe ou les noeuds représentent les membres et les arcs représentent les interactions sociales entre ces membres. Elle est modélisée conceptuellement par l'instanciation du PCO organisationnel proposé dans (Ben Yahia *et al.*, 2013).

2) La deuxième phase concerne l'application de l'algorithme proposé pour détecter la communauté qui peut aider un membre dans la recherche des collaborateurs pertinents. Cet algorithme calcule les poids des arcs et des noeuds et ensuite optimise la qualité bi-objective pour détecter la bonne communauté.

3) La troisième phase recommande les membres de la communauté détectée (considérés comme des collaborateurs pertinents et efficaces) au membre demandant l'aide.



**Figure 1.** L'approche proposée de recommandation des collaborations.

#### **4. Expérimentation de l'approche proposée sur un graphe de terrain : un laboratoire de recherche**

Afin d'évaluer la performance de l'algorithme de recommandation proposé, nous proposons de comparer les résultats de cet algorithme par les résultats donnés par l'application de l'algorithme de Louvain (Blondel *et al.*, 2008) sur un graphe de terrain qui représente le laboratoire RIADI<sup>1</sup>. En effet, nous avons testé les deux approches en considérant le laboratoire RIADI comme une organisation où les membres (les noeuds) représentent les chercheurs et les interactions (les arcs) représentent les relations de co-publication et de co-encadrement. Le graphe ainsi que le PCO organisationnel ont été construit progressivement (en ajoutant à chaque fois des nouveaux chercheurs et leurs copublications) en utilisant les fiches-chercheurs de différents membres. Le tableau 1 expose les différents résultats obtenus lors de l'application de notre approche et ceux obtenus lors de l'application de l'approche de Louvain (un bloc de chercheurs est à chaque fois ajouté jusqu'à l'obtention d'un graphe englobant 100 noeuds). Dans ce tableau, nous comparons les résultats obtenus par les deux approches en nous basant sur la qualité de détection de communautés et le temps d'exécution. Pour notre approche, nous donnons séparément les résultats liés à la modularité et ceux liés à la proportion d'inertie inter-classes. Ensuite, nous supposons le facteur de pondération  $\alpha$  est égal à 0,5 (nous obtenons ainsi l'équité entre les interactions sociales et l'homophilie des participants) et nous donnons les résultats liés à la qualité globale.

Pour mieux visualiser ces mesures, nous avons exploité quelques valeurs pour tracer les deux courbes suivantes.

La première courbe donnée par la figure 2 représente une comparaison du temps d'exécution des deux algorithmes. D'après cette courbe, nous remarquons que l'algorithme de Louvain est plus rapide uniquement pour des graphes de petites tailles. Cependant l'algorithme utilisé dans notre application a un temps d'exécution qui est presque stable. Cela signifie que ce dernier permet de converger indépendamment de la taille du réseau exploité, ce qui n'est pas le cas pour l'algorithme de Louvain.

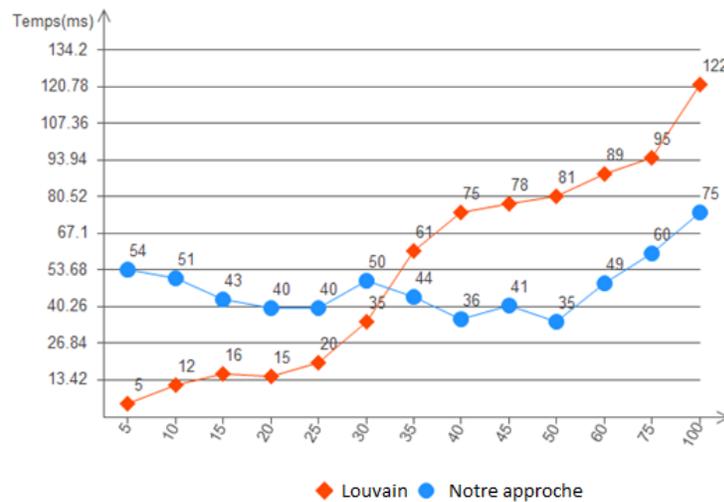
La deuxième courbe donnée par la figure 3 représente une comparaison de la modularité de la communauté détectée. D'après cette courbe, nous pouvons affirmer que la recommandation est plus performante et plus précise dans le cas de l'application du nouvel algorithme que dans le cas où l'algorithme de Louvain est appliqué pour détecter la communauté la plus proche d'un noeud donné. De plus l'algorithme que nous avons choisi tient compte de la similarité entre les attributs des noeuds, ce qui n'est pas le cas pour l'algorithme de Louvain qui considère uniquement les liens entre les noeuds.

---

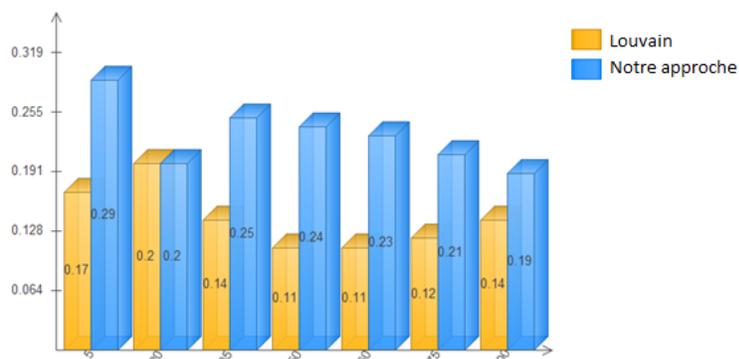
1. [www.riadi.rnu.tn](http://www.riadi.rnu.tn)

**Table 1.** Tableau des mesures.

Réseau social		Algorithme de Louvain		Algorithme de recommandation proposé			
Nombre de noeuds	Nombre de liens	Temps d'exécution (ms)	Modularité de la communauté détectée	Nombre de communautés détectées	Temps d'exécution (ms)	Modularité de la communauté détectée	Nombre de communautés détectées
5	7	5	0.17	1	54	0.29	2
10	16	12	0.24	2	51	0.24	2
15	23	16	0.23	3	43	0.23	2
20	30	15	0.20	4	40	0.20	2
25	38	20	0.16	5	40	0.25	2
30	42	35	0.26	2	50	0.26	2
35	47	61	0.14	7	44	0.25	2
40	53	75	0.06	6	36	0.24	2
45	59	78	0.12	8	41	0.25	2
50	63	81	0.11	9	35	0.24	2
60	73	89	0.11	10	49	0.23	2
75	92	95	0.12	12	60	0.21	2
100	114	122	0.14	16	75	0.19	2



**Figure 2.** Comparaison du temps d'exécution des deux algorithmes.



**Figure 3.** Comparaison de la modularité de la communauté détectée par les deux algorithmes.

## 5. Conclusion et perspectives

Nous avons proposé dans cet article une approche hybride de recommandation des collaborations en combinant la détection de communautés (une technique de partitionnement de graphe en considérant l'organisation comme un réseau social) et l'optimisation par essaim particulière (une méta heuristique). Cette approche se base sur une fonction bi-objective qui considère la modularité (les interactions sociales entre les individus) et l'homophilie (l'homogénéité des individus) pour évaluer la qualité de la communauté détectée. Ensuite, nous avons eu recours à l'optimisation par essaim particulière pour maximiser cette qualité bi-objective. A ce stade de réflexion, nous proposons les deux perspectives suivantes : i) vérifier la scalabilité de l'approche proposée pour des graphes de terrains de taille importante, ii) étudier la résilience des communautés en se basant sur leurs productivités et leurs rendements.

## Bibliographie

- Ackerman M. S., Dachtera J., Pipek V., Wulf V., « Sharing Knowledge and Expertise : The CSCW View of Knowledge Management », *J Computer Supported Cooperative Work*, vol. 22(4-6), p. 531-573, 2013.
- Ben Yahia N., Bellamine N., Ben Ghezala H., « Vers une architecture multicouche d'ontologies dédiée à la résolution mixte de problèmes », In : *Int Proc of Extraction et gestion des connaissances, Toulouse*. 263-268, 2013.
- Blondel V., Guillaume J. L., Lambiotte R., Lefebvre E., « Fast unfolding of communities in large networks », *J Statistical Mechanics : Theory and Experiment*, vol. 10, p. 10008-10020, 2008.

- Cai X., Bain M., Krzywicki A. and Wobcke W., Kim Y., Compton P., Mahidadia A., « Collaborative Filtering for People to People Recommendation in Social Networks », *In : IntProc of Australasian Conference on Artificial Intelligence, South Australia*. 476-485, 2010.
- Carlisle G., Dozier G., « An Off-The-Shelf PSO », *In : IntProc of Particle Swarm Optimization Workshop, USA*. 1-6, 2001.
- Clerc M., Siarry P., « Une nouvelle méta-heuristique pour l'optimisation difficile : la méthode des essaims particuliers », *J l'enseignement des sciences et technologies de l'information et des systèmes*, 2004.
- Dong Y., Tang J., Wu S., Tian J., Chawla N., Rao J., Cao H., « Link Prediction and Recommendation across Heterogeneous Social Networks », *In : IntProc of IEEE 12th International Conference on Data Mining, Belgium*. 181-190, 2012.
- Eberhart R. C., Kennedy J., « New optimizer using particle swarm theory », *In Proceedings of the 6th International Symposium on Micro Machine and Human Science, Nagoya, Japan*. 39-43, 1995.
- El dor A., Perfectionnement des algorithmes d'Optimisation par Essaim Particulaire. Applications en segmentation d'images et en électronique, Thèse de doctorat, Paris-Est University, 2012.
- Guimera R., Sales-Pardo M., Amaral L., « Modularity from fluctuations in random graphs and complex networks », *Physical Review Ep*. 70, 2004.
- He D., Liu J., Liu D., Jin D., Jia Z., « Ant colony optimization for community detection in large-scale complex networks », *In : IntProc of International Conference on Natural Computation (ICNC), Chinap*. 1151-1155, 2011.
- Kautz H., Selman B., Shah M., « Referral Web : Combining Social Networks and Collaborative Filtering », *Magazine Communications of the ACM*, vol. 40(3), p. 63-65, 1997.
- Lebart L., Maurineau A., Piron M., *Traitement des données statistiques*, Dunod Paris, 1982.
- Leskovec J., Lang K., Mahoney M., « Empirical Comparison of Algorithms for Network Community Detection », *In : IntProc of International Conference on World Wide Web, USA*. 631-640, 2010.
- Liben-Nowell D., Kleinberg J., « The link prediction problem for social networks », *In : IntProc of Conference on Information and Knowledge Management, USA*. 556-559, 2003.
- Lu L., Zhou T., « Link prediction in complex networks : a survey », *J Physica A : Statistical Mechanics and its Applications*, vol. 390(6), p. 1150-1170, 2011.
- Matthews T., Whittaker S., Moran T., Yuen S., Judge T., « Productive interrelationships between collaborative groups ease the challenges of dynamic and multi-teaming », *J Computer Supported Cooperative Work*, vol. 21(4-5), p. 371-396, 2012.
- Mazur P., Zmarzkowski K., Orkowski A J., « Genetic Algorithms Approach to Community Detection », *J Act a Physica Polonica A*, vol. 117(4), p. 703-705, 2010.
- McPherson M., Smith-Lovin L., Cook J., « Birds of a feather : Homophily in social networks », *Annual Review of Sociology*, vol. 27(1), p. 415-444, 2001.
- Nardi B., Whittaker S., Isaacs E., Creech M., Johnson J., Hainsworth J., « Integrating Communication and Information Through ContactMap », *Communications of the ACM*, vol. 45(4), p. 89-95, 2002.
- Newman M., Girvan M., « Finding and evaluating community structure in networks », *J Physical Review*, vol. 69(2), p. 26-113, 2004.

- Ogata H., Yano Y., Furugori N., Jin Q., « Computer Supported Social Networking For Augmenting Cooperation », *J Computer Supported Cooperative Work*, vol. 10(2), p. 189-209, 2001.
- Parsopoulos K., Vrahatis M., « Particle Swarm Optimization and Intelligence : Advances and Applications », *Information Science Reference (IGI Global), USA*, 2010.
- Pizzuti C., « Community detection in social networks with genetic algorithms », *In : Int Proc of 10th annual conference on Genetic and evolutionary computation, ACM New York, USA*, p. 1137-1138, 2008.
- Sadi S., Etaner-Uyar S., Gunduz S., « Community detection using ant colony optimization techniques », *In : IntProc of 15th International Conference on Soft Computing*, p. 206-213, 2009.
- Smith M. A., Cadiz J., Burkhalter B., « Conversation Trees and Threaded Chats », *In : IntProc of ACM Conference on Computer-Supported Cooperative Work ,Philadelphia*, vol. , p. 97-105, 2000.
- Wenger E., Snyder W., « Communities of practice : the organizational frontier », *J Harvard Business Review*, vol. 78(1), p. 139-146, 2000.