Trois approches de la Complémentarité de personnes dans les Réseaux Sociaux

Michel Plantié , Mouhamadou Niang , Michel Crampes

LABORATOIRE LGI2P, *Ecole des Mines, Parc Georges Besse, 30035 Nîmes, France*
http:social-networks.mines-ales.fr
michel.{plantie,crampes}@mines-ales.fr

RÉSUMÉ. *La complémentarité dans les réseaux sociaux est une approche intéressante pour identifier la cohésion de groupes de personnes. Nos travaux précédents ont étudié une première approche de la complémentarité dans les réseaux représentés sous la forme de graphes bipartis : réseaux sociaux, communautés d'acteurs, etc. Dans cet article nous comparons plusieurs approches de la complémentarité afin de trouver la technique la plus appropriée pour remplir des objectifs prédéfinis. Notre approche se base toujours sur la notion d'entropie et de gain d'information. Dans certaines conceptions de la complémentarité, la problématique est proche d'une recherche classique : trouver des transversaux dans des hypergraphes, avec toutefois des différences dans les buts recherchés. Afin de valider notre approche, nous appliquons et comparons les différentes méthodes sur des graphes bien connus et des données réelles dont les tailles sont très différentes : des petits graphes à de très grands graphes.*

ABSTRACT. *Complementarity in social networks is an interesting approach to identify cohesion in groups of persons. Our previous works studied a first approach of complementarity in networks represented as bipartite graphs: social networks, communities of actors, etc. In this paper we compare several approaches of complementarity to find the most appropriate technique achieving defined objectives. In some definitions of complementarity, the problem is viewed as close to a classical research: find transversals in hypergraphs, with however differences in final goals. To validate our approach, we apply and compare our methods on well known graphs and real data whose sizes are very different: from small graphs to very large graphs.*

MOTS-CLÉS : *Complementarité, Réseaux Sociaux, graphes bipartis, transversaux, Hypergraphes*

KEYWORDS: *Complementarity, Social Networks, bipartite graphs, transversals, Hypergraphs*

1. Introduction

Beaucoup de travaux de recherche en fouille de données utilisent la notion de similarité ou de distance pour obtenir des corrélations et des analyses sur les données étudiées. La complémentarité se démarque un peu de cette approche en cherchant à associer des éléments qui ne sont pas semblables mais peuvent couvrir un ensemble d'objectifs complémentaires. L'importance prise par les réseaux sociaux, pose des questions en terme de similarité, mais également dans l'identification d'éléments complémentaires. La complémentarité est perçue différemment selon les disciplines. Les mathématiques évidemment ont étudié longuement ce problème à travers de nombreux travaux de recherche. Les problèmes de complémentarité sont la suite naturelle des problématiques suggérées par la décomposition de graphes en communautés. En effet une communauté est d'autant plus viable qu'elle obéit à des règles sémantiques et pragmatiques pilotées par les réalités quotidiennes des individus. Une règle possible et fréquemment appliquée dans la réalité est la complémentarité qui contribue à la stabilité d'un groupe. Si les individus sont trop en concurrence dans un groupe, alors ce groupe peut devenir instable. Prenons un exemple simple : une équipe de rugby. Elle est constituée de joueurs possédant différentes compétences. Tous les joueurs ne peuvent pas être semblables si l'on souhaite que l'équipe soit performante. Si l'on a deux demis de mêlée dans l'équipe, même s'ils jouent à des postes différents, les deux joueurs vont certainement effectuer les mêmes actions et affecter la performance de l'équipe, dans laquelle l'atmosphère va se dégrader et créer de l'instabilité. La complémentarité des joueurs contribuera à la cohésion de l'équipe. Nous étudions donc ici, des réseaux de personnes représentés sous la forme de graphes bipartis. Dans ces graphes, de nombreuses relations sociales sont médiatisées par des propriétés communes entre individus. Il n'y a pas de relations directes entre individus ou entre propriétés mais directement des relations entre individus et propriétés. Nous approfondissons théoriquement et expérimentalement les différentes possibilités de la complémentarité d'un groupe qui peut être un réseau social ou une communauté plus réduite. La complémentarité, comme nous allons le voir peut consister simplement à la couverture d'un espace de propriétés mais peut ensuite être vue comme une propriété plus générale liée à la notion d'utilité d'un ensemble d'individus. Ces deux notions presque duales pouvant d'ailleurs contribuer à des objectifs plus ambitieux.

L'article est structuré ainsi : après un état de l'art sur la complémentarité, nous définissons les différentes notions nécessaires à l'établissement d'une mesure de complémentarité, puis nous décrivons plusieurs méthodes différentes de complémentarité. Enfin nous montrons des applications sur des graphes connus ou réels de tailles très variées afin d'illustrer les concepts et les résultats obtenus.

2. État de l'art

Le terme complémentarité dérivé du mot Latin **complere** qui signifie "compléter" [1], est étudié dans les sciences humaines et sociales et dans d'autres disciplines, avec des significations différentes. Elle a été utilisée principalement dans les domaines du biomédical, de l'économie, de l'innovation et du management mais aussi en mathématiques. Dans le domaine biomédical, les auteurs définissent la complémentarité comme la meilleure correspondance spatiale possible dans l'association d'une ou plusieurs protéines pour remplir au mieux une fonction. Pour les molécules, les auteurs (Lawrence, Colman, 1993) proposent une méthode statistique pour mesurer la complémentarité de formes entre deux protéines. On parle alors de **complémentarité géométrique**. En économie, (Milgrom, Roberts, 1994) donnent une définition de la complémentarité qui représente la logique de marché. Deux produits intermédiaires entrant dans un processus de production sont des compléments si la diminution du prix de l'un augmente la croissance de la demande de l'autre.Par exemple, si l'on considère que le lait et le sucre sont deux produits complémentaires, le fait de diminuer le prix du lait, va faire exploser la demande d'achat sur le sucre, montrant ainsi que les deux produits dépendent l'un de l'autre au sens de la complémentarité. Pour mieux employer ce concept de complémentarité dans l'étude des choix concernant les niveaux d'activités internes d'une organisation ou d'une structure ainsi que les niveaux d'achat de biens intermédiaires, les deux auteurs adoptent une définition plus générale de la complémentarité : des activités sont complémentaires si l'augmentation de la rentabilité de l'une d'entre elles nommée A augmente la rentabilité de toutes les autres activités d'une quantité supérieure à leur augmentation lorsque elles sont utilisées sans l'activité A et réciproquement. Par exemple l'étude faite dans (Ichniowski *et al.*, 1997) des effets des stratégies innovantes sur la productivité auprès de 36 lignes de production d'acier homogène appartenant à 17 compagnies montre que certaines pratiques de travail (stratégies) telles que la formation des employés, les primes de travail, affectations flexibles des tâches ont eu un effet plus important sur la productivité quand elles étaient couplées à un système de Ressources Humaines bien organisé que quand elles ont été utilisées séparément. Cette notion de complémentarité est formalisée dans (Milgrom, Roberts, 1995) pour comprendre les liens existants entre les éléments de stratégie d'une entreprise manufacturière et sa structure. Dans leur définition un élément x est complémentaire à un élément y si et seulement si la condition de supermodularité suivante est vérifiée :

$$\prod(x \wedge y) + \prod(x \vee y) \geq \prod(x) + \prod(y)(Milgrom, Roberts, 1990) \qquad (1)$$

$\prod$ est une fonction d'utilité.

Dans le domaine de l'innovation, la complémentarité peut être étudiée pour comprendre comment les innovations technologiques et organisationnelles évo-

1. http://lexicon.ft.com/Term?term=complementarity

luent et s'influencent dans le temps mais aussi sous quelles conditions des activités liées à l'innovation peuvent être complémentaires. (Ayerbe, 2006) identifie trois niveaux de complémentarité entre innovations technologiques et organisationnelles :

- coexistence entre innovations technologiques et organisationnelles: le simple fait de leur présence simultanée (sans que l'une n'influence l'autre) améliore la performance du système.

- l'innovation technologique et l'innovation organisationnelle s'influencent mutuellement pour améliorer la performance du système.

- la dimension temporelle des relations entre innovations (ordre d'apparition des innovations dans le temps) que nous n'étudierons pas dans cet article.

La définition formelle de la complémentarité de (Cassiman, Veugelers, 2006) entre activités complémentaires liées à l'innovation est que deux entités sont complémentaires **si leur apport au système est plus grand quand ils sont présents ensemble dans le système que leur apport séparé**. Cette définition a été formalisée comme suit: :

$$Complémentarité(a, b) : \prod(a, b) - \prod(\neg a, b) > \prod(a, \neg b) - \prod(\neg a, \neg b) \quad (2)$$

où a et b sont deux éléments du système. $\prod(a, \neg b)$ signifie que a est présent et b absent. L'utilité mesure la différence entre $\prod$ lorsque a est présent et $\prod$ lorsque a est absent dans les deux cas en présence de b par rapport aux mêmes situations lorsque b est absent.

La complémentarité d'information sur un sujet a été abordée dans la communauté Recherche d'Information (RI). Dans (Ma *et al.*, 2006), les auteurs proposent d'abord une nouvelle méthode de représentation des informations extraites dans différentes médias appelée **topic-structure**. Puis à partir d'un sujet ou d'une information intéressant(e) extrait(e) de la diffusion d'une émission de télévision (source primaire) cherche des pages web (source secondaire) qui sont complémentaires au sujet qui nous intéresse. Un " topic-structure " consiste en un couple (sujets, contenus) où sujets représentent les termes dominants dans les médias (thème) et contenus représentent les termes qui ont des relations de co-occurrence fortes avec les sujets diffusés. Cependant le problème de la polysémie émerge très rapidement que les auteurs ont prévu de résoudre dans leurs futurs travaux.

Par ailleurs la complémentarité se rapproche indirectement de la notion d'hypergraphe transversal. En effet dans (Jelassi *et al.*, 2014), les auteurs démontrent que la recherche de multi-membres est semblable au problème de détermination d'un sous ensemble de transversaux minimaux. Un multi-membre est un sous ensemble de sommets (acteurs) qui représentent le mieux possible l'ensemble des communautés, c'est à dire un ensemble de sommets qui couvrent un nombre maximum ou l'ensemble des communautés.

Cependant notre approche se distingue de leur méthode. Dans (Jelassi *et al.*, 2014), les auteurs travaillent sur des communautés d'acteurs. Ils supposent que les communautés existent ou peuvent être préalablement détectées par des algorithmes adéquats. Dans notre méthode nous rajoutons un critère supplémentaire : l'utilité d'un acteur qui n'apparait pas de façon évidente dans l'appartenance à une communauté.

3. COMPLÉMENTARITÉ : DÉFINTIONS ET FORMULES

Dans cette section nous définissons et formalisons notre notion de complémentarité. Auparavant nous présentons les concepts que nous allons utiliser. Les réseaux sur lesquels nous étudions la complémentarité sont représentés sous la forme de graphes bipartis.

3.1. Entropie d'un système

L'entropie est la mesure d'information apportée par un élément dans un ensemble de classe d'éléments (Yang, Petersen, 1997). Dans notre méthode l'entropie système sera calculé sur l'espace des propriétés. En effet nous cherchons à déterminer la quantité d'information portée par une personne qui peut se matérialiser par l'espace des propriétés que la personne couvre. Soit S un système représenté par un graphe biparti $G = (Pe, Pr, E)$, avec Pe un sous-ensemble de n_1 personnes, Pr un sous-ensemble de n_2 propriétés, E un ensemble de m arêtes entre les personnes de Pe et les propriétés de Pr. L'entropie est formulée comme suit :
$H(S) = \frac{-\sum_{j=1}^{n_2} P(Pr_j) \times ln(P(Pr_j))}{ln(n_2)}$, $n_2 > 0$.
$P(Pr_j) = \frac{\sum_{i=1}^{n_1} A(i,j)}{m}$ est la probabilité d'apparition de la propriété Pr_j pour toutes les personnes du réseau dans le système S. $A(i, j)$ est la matrice d'adjacence de G, i représentant l'indice des personnes Pe et j l'indice des propriétés Pr (voir (Crampes, Plantié, 2015) pour une explication plus détaillée).

3.2. Entropie conditionnelle

Nous considérons toujours le graphe biparti $G = (Pe, Pr, E)$. L'entropie conditionnelle est définie et expliquée dans (Crampes, Plantié, 2015) comme l'entropie portée par une personne Pe_j . Elle est formalisée comme suit :

$$H(S/Pe_j) = \frac{-\sum_{i=1}^{n_2} P(Pr_i/Pe_j) \times ln(P(Pr_i/Pe_j))}{ln(n_2)}$$

ou de non-présence d'une personne :

$$H(S/\overline{Pe_j}) = \frac{-\sum_{i=1}^{n_2} P(Pr_i/\overline{Pe_j}) \times ln(P(Pr_i/\overline{Pe_j}))}{ln(n_2)}$$

$\mathrm{P}(\mathrm{P}_{r_i} \backslash \mathrm{P}_{e_j})$ est la probabilité d'avoir la propriété $\mathrm{P}(\mathrm{P}_{r_j})$ sachant que la personne P_{e_j} soit présente. Elle se calcule comme suit : $P(P_{r_i} \backslash P_{e_j}) = \begin{cases} 0 & si\ A(i,j) = \end{cases}$ $0, 1 \setminus nb$ sinon. nb est le nombre de fois ou la personne P_{e_j} a participé aux évènements.

3.3. Fonction d'utilité, Gain d'information

Une fonction d'utilité permet de mesurer le niveau de satisfaction, l'importance ou l'impact d'un élément dans un ensemble d'éléments. Dans notre cas nous utiliserons le gain d'information pour mesurer l'importance d'une personne (la quantité d'information qu'elle porte) dans le système S défini plus haut. Plus l'information portée par une personne est grande plus la personne est importante dans le système. Le gain d'information représente la différence des entropies du système et des entropies conditionnelles pour une personne Pe_j et son complémentaire $\overline{Pe_j}$. Elle est formulée comme suit :

$$G(Pe_j) = H(S) - P(Pe_j) \times H(S/Pe_j) - P(\overline{Pe_j}) \times H(S/\overline{Pe_j}) \qquad (3)$$

3.4. Définition générique de la complémentarité

L'état de l'art montre que la complémentarité est utilisée dans plusieurs domaines avec des significations différentes. De ce constat, nous donnons une définition générique qui peut englober l'ensemble de ces domaines. Toutes les conceptions de la complémentarité reposent sur le fait que deux éléments sont complémentaires, si leur apport au système est plus grand quand ils sont présents ensemble que leur apport quand ils sont pris séparément. Cette notion peut se décliner selon les quatre variantes suivantes :

– les éléments sont complémentaires par influence réciproque,

– des éléments du système influencent d'autres éléments (sans réciprocité),

– l'objectif unique est la couverture par les éléments de l'ensemble des propriétés (la formule 1 de Milgrom en est une implémentation),

– l'objectif est aussi la couverture par les éléments de l'ensemble des propriétés, avec en plus des propriétés communes entre ces éléments (Cassiman).

Les deux premiers cas ne correspondent pas à notre approche de la complémentarité fondée sur la notion d'utilité conjointe. Nous cherchons à trouver un ensemble minimal d'éléments qui couvre toutes les propriétés Pr d'un graphe biparti, et éventuellement que cet ensemble d'éléments apporte un surplus d'utilité au système. Ce qui correspond aux deux dernières variantes ci-dessus que nous traiterons. Nous déclinons ces deux variantes en trois méthodes différentes (les formules détaillées seront présentées dans le paragraphe suivant) de calcul de la complémentarité :

1 méthode **simple** : les éléments sont sélectionnés au hasard, par association successive, pour trouver un ensemble d'éléments couvrant toutes les propriétés du graphe.

2 méthode inspirée de la définition de la complémentarité par **Cassiman** avec 3 variations possibles :

- variante 1 : formule de Cassiman avec fonction **OU** d'appariement : les éléments sélectionnés couvrent toutes les propriétés et peuvent avoir en plus des propriétés communes.

- variante 2 : formule de Cassiman avec fonction **OU Exclusif** d'appariement : les éléments sélectionnés partagent le moins de propriétés possible. On cherche ainsi à trouver des éléments les plus disjoints au sens des propriétés partagées.

- variante 3 : formule de Cassiman avec fonction **ET** d'appariement : certains éléments sélectionnés ont de fortes connections entre eux (redondance élevée).

3 méthode utilisant la formule de supermodularité de **Milgrom**

La figure 1 suivante illustre sur un exemple la spécificité des méthodes 2 et 3 décrites ci-dessus. Il s'agit d'un graphe de 4 personnes nommées de A à D et 4 colonnes. À droite un tableau montre le résultat de complémentarité de chaque méthode. Les sommets sélectionnés assurent une couverture complète.

A	1	1	0	0
B	0	1	1	1
C	0	0	1	1
D	1	0	0	0

	CASSIMAN			Milgrom
Méthode	OU	ET	Ou Exclusif	
Résultats	A et B	A,D et C	A et C	A et C

FIGURE *1. Complémentarité sur un graphe de 4 personnes et 4 propriétés*

3.5. Formalisation de la Complémentarité

Dans cette section nous donnons les formules des cinq méthodes ou variantes de calcul de complémentarité décrites ci dessus. La fonction d'utilité est le gain d'information.

3.5.1. Méthode simple

Elle consiste à prendre les éléments du graphe au hasard puis à poursuivre jusqu'à couverture du graphe :

– Choix d'un premier sommet au hasard et arrêt si couverture totale,

– Sinon, prendre l'élément suivant au hasard et recommencer jusqu'à l'obtention d'une couverture de toutes les propriétés.

Cette méthode simple donne une référence de comparaison pour les autres méthodes. La notion d'entropie et de gain d'information n'interviennent pas dans cette méthode.

3.5.2. Méthode utilisant la formule de Cassiman

Nous appliquons la formule d'utilité 2 couplée à la formule 3 du gain.

variante 1, méthode Cassiman utilisant le **OU** :

$$compl\acute{e}mentarit\acute{e}(Pe_i, Pe_j) =$$
$$G(Pe_i \vee Pe_j) - G(\overline{Pe_i} \vee Pe_j) - (G(Pe_i \vee \overline{Pe_j}) - G(\overline{Pe_i} \vee \overline{Pe_j})), \quad (4)$$

variante 2, méthode Cassiman utilisant le **ET** :

$$complémentarité(Pe_i, Pe_j) = G(Pe_i \wedge Pe_j) - G(\overline{Pe_i} \wedge Pe_j) - (G(Pe_i \wedge \overline{Pe_j}) - G(\overline{Pe_i} \wedge \overline{Pe_j})), \quad (5)$$

variante 3, méthode Cassiman utilisant le **OU Exclusif** :

$$complémentarité(Pe_i, Pe_j) = G(Pe_i \oplus Pe_j) - G(\overline{Pe_i} \oplus Pe_j) - (G(Pe_i \oplus \overline{Pe_j}) - G(\overline{Pe_i} \oplus \overline{Pe_j})), \quad (6)$$

Le développement de la variante 1 donne la formule finale suivante :

$$\begin{aligned} complémentarité(Pe_i, Pe_j) = {} & \frac{1}{ln(n_2)} \times [-P(Pe_i \vee Pe_j) \times H(S/Pe_i \vee Pe_j) \\ & - P(\overline{Pe_i \vee Pe_j}) \times H(S/\overline{Pe_i \vee Pe_j}) + P(\overline{Pe_i} \vee Pe_j) \times H(S/\overline{Pe_i} \vee Pe_j) \\ & + P(\overline{\overline{Pe_i} \vee Pe_j}) \times H(S/\overline{\overline{Pe_i} \vee Pe_j}) + P(Pe_i \vee \overline{Pe_j}) \times H(S/Pe_i \vee \overline{Pe_j}) \\ & + P(\overline{Pe_i \vee \overline{Pe_j}}) \times H(S/\overline{Pe_i \vee \overline{Pe_j}}) - P(\overline{Pe_i} \vee \overline{Pe_j}) \times H(S/\overline{Pe_i} \vee \overline{Pe_j}) \\ & - P(\overline{\overline{Pe_i} \vee \overline{Pe_j}}) \times H(S/\overline{\overline{Pe_i} \vee \overline{Pe_j}})] \quad (7) \end{aligned}$$

Les formules finales des autres variantes, s'obtiennent en remplaçant respectivement $\vee$ par $\wedge$ et $\oplus$ dans la formule 7.

3.5.3. Méthode utlisant la formule de supermodularité de Milgrom

Nous appliquons la définition de la supermodularité donnée à la formule 1 couplée à la formule 3 du gain.

$$complémentarité(Pe_i, Pe_j) = G(Pe_i \vee Pe_j) - G(Pe_j) - (G(Pe_i) - G(Pe_i \wedge Pe_j)), \quad (8)$$

Cette formule se développe de façon similaire à la formule 7.

3.6. Algorithme de complémentarité

L'algorithme de recherche de complémentarité est effectué en plusieurs étapes :

Algorithm 1 Algorithme de recherche de complémentarité

donnée initiale : Liste des candidats V, une liste vide L qui va contenir les personnes complémentaires
Calcul de l'entropie système
Trouver la personne avec Gain d'information maximum et l'ajouter à L
$nbcandidat \leftarrow longueur(V)$
$couverturetotale \leftarrow false$
while $((nbcandidat \neq 0)$ and $(\overline{arretCouvertureTotale()}))$ **do**
  $trouverLaPersonneComplementaire()$ avec les personnes déjà sélectionnées
  if $((nbcandidat \neq 0)$ and $(arretCouvertureTotale() == true))$ **then**
    $(couverturetotale \leftarrow true)$
  end if
end while
if $couverturetotale == false$ **then**
  ▷ nous avons une couverture partielle
end if

Procédure trouverLaPersonneComplementaire() : calcule pour chaque candidat sa complémentarité avec les personnes déjà sélectionnées.
Sortie : personne ayant la valeur maximale de complémentarité

Ainsi l'ensemble des Pe_{comp_max} trouvé constitue un groupe de sommets complémentaires couvrant partiellement ou totalement les propriétés du graphe.

La complexité est de $O(n_1^2 \times n_2)$, n_1 étant le nombre de personnes, n_2 étant le nombre de propriétés.

4. Expérimentation

Nous montrons des expérimentations des cinq méthodes de traitement de la complémentarité sur plusieurs graphes de taille et de contexte différents. Tout d'abord, nos expériences se font sur un graphe biparti intensément étudié (voir la meta-analyse de (Freeman, 2003)) "Southern Women" (SW) ou "'Women Events"' (WE). Il consiste en un relevé de la participation différenciée de 18 dames à 14 évènements sociaux. Cette expérimentation simple permet de comprendre l'effet de nos méthodes. Nous expérimentons ensuite nos méthodes sur un graphe biparti issu du monde du rugby sportif. Cette expérimentation a pour objet de vérifier le bien fondé de chaque méthode. Nous utilisons ensuite un graphe biparti issu de Facebook, de taille intermédiaire, et enfin nous testons nos algorithmes sur un grand graphe issu du monde des publications scientifiques, afin de mesurer l'efficacité et le passage à l'échelle. Nos mesures de performances sont le pourcentage de couverture, nombre de personnes sélectionnées et le temps d'exécution.

4.1. graphe biparti WE

Les résultats sur le graphe Women-Events sont présentés dans la figure 2. Nous constatons que les deux algorithmes les plus efficaces sont celui de Cassiman/Ou_Exclusif et Milgrom. L'algorithme simple donne un nombre variable de personnes en fonction de la sélection du premier élément qui s'effectue au hasard. L'utilisation du ou exclusif permet d'obtenir la couverture maximum des évènements avec le minimum de personnes.

Méthode :	Simple	CASSIMAN			Milgrom
		OU	ET	Ou Exclusif	
nombre de personnes sélectionnées	variable 2 à 5/18	16/18	11/18	2/18	2/18
personnes sélectionnées	---	---	---	1-14	1-14
% de couverture	100	100	100	100	100
temps exécution (s)	0	0	0	0	0
temps préparation (s)	0,016	0	0,009	0	0,005

FIGURE *2. Résultats de complémentarité sur le graphe Women-Events*

4.2. La meilleure équipe de Rugby pour la coupe du monde...

Nous avons appliqué les mêmes calculs sur un jeu de données réel : il s'agit de trouver la meilleure équipe française de rugby pour participer à la prochaine coupe du monde. Les données sont donc représentée sous la forme d'un graphe biparti de 91 joueurs pour lesquels 11281 évènements ont été pris en compte : participation aux matches, blessures, points marqués etc sur les cinq dernières années. Les évènements jouent le rôle de propriétés partagées et les personnes le même rôle que dans le graphe Women Events. Les résultats sur le graphe rugby sont présentés dans la figure 3 nous constatons que l'algorithme le plus efficace est celui de Milgrom qui trouve un collectif de 31 joueurs permettant de couvrir l'ensemble des considérations de jeu présentes dans le graphe. Précisons que les contraintes d'un maximum de joueur et autres (nationalités, etc.) n'ont pas été prise en compte dans cette expérimentation.

Méthode :	Simple	CASSIMAN			Milgrom
		OU	ET	Ou Exclusif	
nombre de personnes sélectionnées	variable	91/91	91/91	42/91	31/91
personnes sélectionnées	---	Toutes	Toutes	---	---
% de couverture	100	100	100	100	100
temps exécution (s)	2,06	17,7	0,579	5,3	0,376
temps préparation (s)	0,272	11,5	0,449	0,264	0,375

FIGURE *3. Résultats de complémentarité sur le graphe Rugby*

4.3. graphe Facebook

Nous avons appliqué les mêmes calculs sur un jeu de données plus conséquent : un graphe biparti de partage de 700 photos entre environ 274 personnes provenant d'un compte Facebook. Les photos jouent ici le rôle de propriétés partagées et les personnes le même rôle que dans le graphe Women Events. Les résultats sur le graphe Facebook sont présentés dans la figure 4. Nous constatons à nouveau que les deux algorithmes les plus efficaces sont ceux de Cassiman/Ou_Exclusif et Milgrom.

4.4. graphe Auteur-Article

Pour évaluer le passage à l'échelle de notre méthode, nous avons utilisé un jeu de données bien plus grand : un graphe biparti de relations co-auteurs sur

Méthode :	Simple	CASSIMAN			Milgrom
		OU	ET	Ou Exclusif	
nombre de personnes sélectionnées	variable ~112/273	273/273	273/273	112/273	111/273
personnes sélectionnées	---	Toutes	Toutes	---	---
% de couverture	100	100	100	100	100
temps exécution (s)	0,056	48	0,439	0,078	0,302
temps préparation (s)	0,142	0,18	0,556	0,094	0,423

FIGURE *4. Résultats de complémentarité sur le graphe Facebook*

des articles scientifiques. Les données proviennent de la banque de données d'articles scientifiques bien connue "Pubmed" dans le domaine biomédical [2]. Le jeu de données contient plus de 82000 individus et plus de 40000 articles scientifiques. Les résultats sur ce graphe sont présentés dans la figure 5. Nous constatons à nouveau que les deux algorithmes les plus efficaces sont ceux de Cassiman/Ou_Exclusif et Milgrom. Le temps de traitement de l'algorithme est relativement court (environ 40 secondes), cependant le temps de préparation (constitution des structures de données et stockages) est assez important de l'ordre de 25 minutes. Ce dernier traitement pourra être bien mieux optimisé dans le futur.

Méthode :	Simple	CASSIMAN			Milgrom
		OU	ET	Ou Exclusif	
nombre de personnes sélectionnées	variable ~938/82457	57645/82457	82456/82457	938/82457	938/82457
personnes sélectionnées	---	---	---	---	---
% de couverture	100	100	100	100	100
temps exécution (s)	22,3	1461,53	2434,98	33,1	41,9
temps préparation (s)	1380,5	1401,61	1351,21	1378,6	1388,8

FIGURE *5. Résultats de complémentarité sur le graphe Auteurs-Articles*

discussion :

Nous cherchons à couvrir de façon minimale un espace de propriétés, avec cependant un point additionnel : les éléments sont choisis par rapport à leur utilité, apportant ainsi une différence par rapport aux travaux de recherche des transversaux dans des hypergraphes. Notre algorithme est calculable avec une complexité $O(n_1^2 \times n_2)$. Nos résultats montrent que les algorithmes Cassiman/Ou_Exclusif et Milgrom sont les plus efficaces pour atteindre l'objectif de couverture minimale. "Cassiman/variante OU exclusif" donne une couverture minimale tout en favorisant le partage de propriétés. La méthode "Milgrom" favorise les personnes ayant des propriétés les plus disjointes. Sur l'exemple Rugby la méthode "Cassiman/variante OU exclusif" sélectionne plus de joueurs, mais ils peuvent couvrir une variété plus grande de postes tout en favorisant une couverture maximale. Cette méthode est plus intéressante en terme de complémentarité. La méthode simple est intéressante mais n'est pas optimisée. Elle constitue une base de référence pour vérifier l'efficacité de notre mesure d'utilité et l'amélioration des algorithmes suivants sur les résultats. La méthode "Cassiman/variante OU" ne permet pas d'obtenir une couverture minimale. La méthode "Cassiman/variante ET" se rapproche trop de la notion de similarité.

2. http://www.ncbi.nlm.nih.gov/pubmed

5. Conclusion

Dans cet article nous avons étudié différentes approches de la complémentarité dans les réseaux représentés sous forme de graphes bipartis. L'approche fondée sur "Cassiman/Ou exclusif" semble remplir les objectifs définis au départ. Dans le futur, nous allons optimiser nos algorithmes notamment dans la partie préparation de données qui prend un temps assez long (plus de la moitié du temps d'exécution), puis les étendrons aux graphes pondérés pour apporter plus de sémantique, et nous approfondirons les définitions statistiques dans l'utilisation de l'entropie pour une meilleure sémantique de l'utilité.

Bibliographie

Ayerbe C. (2006). Innovations technologique et organisationnelle au sein de pme innovantes: complémentarité des processus, analyse comparative des mécanismes de diffusion. *Revue internationale PME: Économie et gestion de la petite et moyenne entreprise*, vol. 19, n° 1, p. 9–34.

Cassiman B., Veugelers R. (2006, janvier). In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition. *Management Science*, vol. 52, n° 1, p. 68–82.

Crampes M., Plantié M. (2015). Complémentarité de personnes partageant des propriétés dans les Réseaux Sociaux. In *Ic 2015 ingénierie des connaissance.* Rennes,France.

Freeman L. C. (2003). Finding social groups: A meta-analysis of the southern women data. In *Dynamic social network modeling and analysis. the national academies*, p. 39—-97. Press.

Ichniowski C., Shaw K., Prennushi G. (1997). The effects of human resource management practices on productivity: A study of steel finishing lines. *The American Economic Review*, vol. 87, n° 3, p. pp. 291-313.

Jelassi M. N., Largeron C., Yahia S. B. (2014). Efficient unveiling of multi-members in a social network. *Journal of Systems and Software*, vol. 94, n° 0, p. 30 - 38.

Lawrence M. C., Colman P. M. (1993). Shape complementarity at protein/protein interfaces. *Journal of molecular biology*, vol. 234, n° 4, p. 946–950.

Ma Q., Nadamoto A., Tanaka K. (2006). Complementary information retrieval for cross-media news content. *Information Systems*, vol. 31, n° 7, p. 659–678.

Milgrom P., Roberts J. (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. econometrica. In *Journal of the econometric society.*

Milgrom P., Roberts J. (1994). *Economics, organization and management.* Prentice-Hall.

Milgrom P., Roberts J. (1995). Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of accounting and economics*, vol. 19, n° 2, p. 179–208.

Yang Y., Petersen J. O. (1997). A comparative Study on Feature Selection in Text Categorisation. In *Fourteenth international conference on machine learning, icml'97.*