

# Complémentarité de personnes partageant des propriétés dans les Réseaux Sociaux

Michel Plantié<sup>1</sup>, Michel Crampes<sup>1</sup>

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

**Résumé** : Nos travaux précédents portaient sur la détection unifiée de communautés dans les réseaux de personnes représentés sous la forme de graphes généralement bipartis : réseaux sociaux, communautés d'acteurs, etc. Dans cet article nous tentons de répondre à des problèmes de complémentarité qui se posent dès que l'on souhaite associer des personnes dans le but de remplir au mieux un objectif. Nous définissons donc la notion de complémentarité entre les sommets d'un graphe biparti. Nous utilisons pour cela les notions d'entropie et d'information mutuelle. Nous montrons l'utilité d'une telle démarche et l'intérêt de l'approche par une expérimentation sur des exemples bien connus.

**Mots-clés** : Détection de Communautés, Communautés recouvrantes, Equilibre de Nash, Complémentarité

## 1 Introduction

Avec le développement d'Internet et l'importance prise par les réseaux sociaux, l'étude de ces derniers fait l'objet de nombreux travaux de recherche. Une des voies est la détection de communautés où le principe général consiste à regrouper les individus de telle manière que les liens qu'ils entretiennent avec ceux présents à l'intérieur de leur communauté soient plus nombreux ou plus forts que les liens qu'ils entretiennent avec ceux présents dans les autres communautés. Les champs d'application dépassent les réseaux sociaux puisqu'on en trouve aussi en biochimie, en phylogénie, dans l'analyse des écosystèmes, ou encore plus récemment en neurologie. Beaucoup d'algorithmes aux approches très variées ont été proposés. Les plus importants sont rapportés dans (Papadopoulos *et al.*, 2011; Yang *et al.*, 2010; Porter *et al.*, 2009) et de manière plus détaillée dans (Fortunato, 2009). En grande majorité ils portent sur la détection de communautés non recouvrantes (on parle aussi de partitionnées) dans des graphes unipartis. Dans ce type de graphes tous les noeuds peuvent éventuellement être reliés. Cependant 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. On parle alors de graphes multimodaux, le cas le plus fréquent étant le graphe bimodal (ou biparti).

Dans cet article nous répondons à des problèmes de complémentarité qui est la suite naturelle des problématiques suggérées par la décomposition 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 en concurrence dans un groupe, alors ce groupe peut devenir instable. Par ailleurs un groupe dont les membres collaborent en fonction de leurs compétences variées à l'obtention d'un but commun est souvent recherché. C'est le cas pour toute entreprise, équipe sportive, etc. Il convient donc

de s'intéresser dans un premier temps à définir ce que nous entendons par complémentarité dans un groupe social. Nous laissons de côté pour l'instant la notion de communauté pour mieux approfondir théoriquement et expérimentalement la complémentarité d'un groupe qui peut être un réseau social ou une communauté plus réduite. Nous définissons la complémentarité puis cherchons à l'appliquer sur des réseaux sociaux. 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 définissons également une méthode pour obtenir une stabilité dans un réseau social en se basant sur la complémentarité. Enfin nous montrons des applications sur un réseau connu afin d'illustrer les concepts et les résultats obtenus.

## 2 État de l'art

Les travaux liés à la complémentarité ont été publiés principalement dans les domaines de l'économie, de l'innovation et du management mais aussi en mathématiques. Globalement le principe 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. Dans ce contexte, (Cassiman & Veugelers, 2006) donnent une définition formelle de la complémentarité entre activités liées à l'innovation. Dans le domaine de l'innovation, elle est étudiée pour comprendre sous quelles conditions des activités liées à l'innovation peuvent être complémentaires. Il est donc primordial d'identifier des variables qui affectent la complémentarité. La définition formelle de la complémentarité de (Cassiman & Veugelers, 2006) entre activités complémentaires liées à l'innovation liste les variables qui l'impactent : force d'exportation, force d'innovation, information publique disponible, propriété intellectuelle, etc. Dans leur approche, la complémentarité est définie par rapport à une performance accrue (corrélation positive). Leur conception de la complémentarité est définie selon la formule suivante :

$$\text{Complémentarité}(a, b) : \Pi(1, 1) - \Pi(0, 1) > \Pi(1, 0) - \Pi(0, 0) \quad (1)$$

où  $\Pi$  est la fonction d'utilité ou de performance du système,  $a$  et  $b$  sont deux éléments du système.  $\Pi(1, 0)$  signifie que  $a$  est présent et  $b$  absent. L'utilité mesure la différence entre  $\Pi$  lorsque  $a$  est là et  $\Pi$  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). L'approche proposée dans (Ma *et al.*, 2006) repose sur des mots-clés structurés. Les auteurs envisagent la recherche d'information complémentaire en particulier en croisant plusieurs média. Les mathématiques ont étudié abondamment la complémentarité et de nombreux travaux comme (Topkis, 1998) définissent la complémentarité comme une fonction mesurant l'effet combiné de deux acteurs simultanés sur un système. Dans la suite nous utiliserons cette notion de complémentarité de (Cassiman & Veugelers, 2006) en utilisant le concept d'entropie et d'information mutuelle que nous développerons plus loin.

La complémentarité entre les personnes est pressentie par les activités collaboratives et interactionnelles développées lors de participation à des réseaux et/ou à des projets. Elle est aussi visible dans la co-édition de documents où les apports de chacun des auteurs peuvent être spécifiés (voir figure 1).

La complémentarité d'information proposée dans (Ma *et al.*, 2006) définit une structure de graphe contenant des sujets et des contenus : topic structure. On peut cependant redouter qu'en

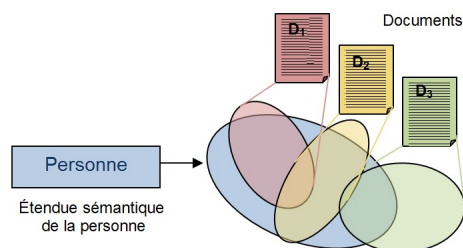


FIGURE 1 – Complémentarité de documents et de personnes.

se basant uniquement sur les termes, l'homonymie dégrade la pertinence des résultats. Les auteurs ambitionnent, dans les perspectives de (Ma *et al.*, 2006) d'utiliser un modèle sémantique (ontologie). Dans sa théorie des hypergraphes C Berge (Berge, 1987) évoque la notion de Transversal : un transversal d'un hypergraphe  $H = (E_1, \dots, E_m)$  sur un ensemble  $X$  est un ensemble  $T \subseteq X$  qui a une intersection non vide avec chaque hyperarête  $T \cap E_i \neq \emptyset$ , ( $i = 1, 2, \dots, m$ ). Les transversaux d'un hypergraphe ne sont pas uniques et l'on peut également parler de transversal minimal. La notion de transversal se rapproche de la complémentarité, si l'on considère des graphes bipartis qui sont équivalents à des hypergraphes. La complémentarité dans ce type de structure consiste en particulier à trouver les éléments du premier sous ensemble de sommets du graphe qui couvrent le maximum de connexions avec les sommets du deuxième sous-ensemble de sommets. La complémentarité se rapproche donc de la problématique de la recherche du transversal minimum dans un hypergraphe. La recherche de transversaux minimum a fait l'objet de nombreux travaux (Kavvadias & Stavropoulos, 2005; Khachiyan *et al.*, 2005; Eiter & Gottlob, 1995). Dans notre cas la complémentarité bien que proche du problème de la recherche de transversaux minimum s'en distingue du fait que nous rajoutons un critère supplémentaire d'utilité. En effet nous souhaitons associer des éléments complémentaires au regard de leur capacité à couvrir l'ensemble des problématiques du système étudié mais aussi à maximiser une fonction d'utilité.

Dans la théorie des ensembles ce thème de la complémentarité a été abondamment étudié et le point de départ est le fait qu'un sous-ensemble  $S_a$  d'un ensemble  $S$  est complémentaire à un autre sous-ensemble  $S_b$  si il est constitué de la différence de l'ensemble  $S$  moins le sous-ensemble  $S_b$ . Cette définition fondamentale a ensuite été étendue et de nombreux travaux mathématiques comme (Topkis, 1998) cité précédemment élargissent la définition de la complémentarité pour se rapprocher de celle de Cassiman.

En logique floue, la notion de mesure floue ou capacité se rapproche de la notion de complémentarité que l'on peut interpréter comme le poids de la coalition d'éléments. Si le poids de deux éléments est positif alors leur coalition a un apport positif pour le système. Le poids peut être une mesure floue (caractérisée uniquement par sa monotonie).

dans (Jelassi *et al.*, 2014) les auteurs ont abordé de façon différente la notion de complémentarité. Ils considèrent un graphe monoparti déjà décomposé en communautés non-recouvrantes. Ils introduisent alors la notion de multi-membre : un ensemble de sommets de taille minimum avec au moins un sommet dans chaque communauté et si possible avec plusieurs sommets dans les communautés de plus grande taille.

NAMES OF PARTICIPANTS OF GROUP I	CODE NUMBERS AND DATES OF SOCIAL EVENTS REPORTED IN <i>Old City Herald</i>													
	(01) 6/27	(02) 8/2	(03) 4/12	(04) 5/26	(05) 2/28	(06) 5/19	(07) 3/15	(08) 9/16	(09) 4/8	(10) 6/10	(11) 2/28	(12) 4/7	(13) 11/21	(14) 8/2
1. Mrs. Evelyn Jefferson.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2. Miss Laura Mandeville.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
3. Miss Theresa Anderson.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
4. Miss Brenda Rogers.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
5. Miss Charlotte McDowd.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
6. Miss Frances Anderson.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
7. Miss Eleanor Nye.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
8. Miss Pearl Ogleshorpe.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
9. Miss Ruth DeSand.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
10. Miss Verne Sanderson.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
11. Miss Myra Liddell.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
12. Miss Katherine Rogers.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
13. Mrs. Sylvia Avondale.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
14. Mrs. Nora Fayette.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
15. Mrs. Helen Lloyd.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
16. Mrs. Dorothy Murchison.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
17. Mrs. Olivia Carlston.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X
18. Mrs. Flora Price.....	X	X	X	X	X	X	X	X	X	X	X	X	X	X

FIGURE 2 – Matrice des personnes du réseau WE

### 3 Définition et formules

#### 3.1 Graphe biparti

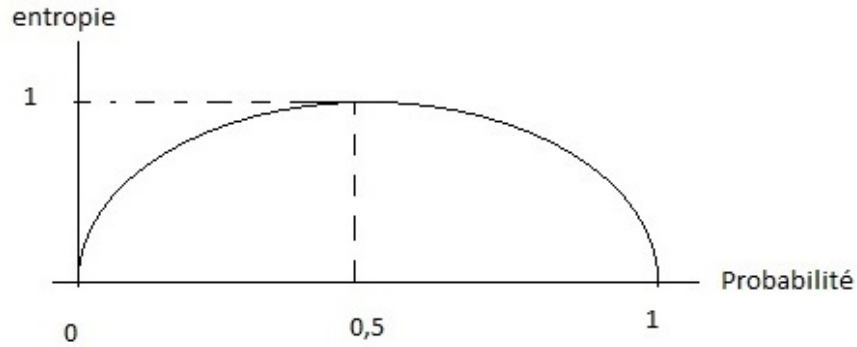
Soit un réseau social de personnes partageant des propriétés. Nous pouvons considérer ce réseau comme un système  $S$  représenté par un graphe biparti  $G = (Pe, Pr, E)$ , avec  $Pe$  un sous-ensemble de  $n_1$  personnes,  $Pr$  un sous-ensemble de  $n_2$  propriétés,  $E$  un ensemble de  $m$  arêtes entre les personnes de  $Pe$  et les propriétés de  $Pr$ , ayant comme matrice d'adjacence :  $A(i, j)$ ,  $i$  représentant l'indice des personnes  $Pe$  et  $j$  l'indice des propriétés  $Pr$ . Nous définissons un système  $S$  comme l'ensemble des propriétés  $Pr_j$  partagées par les personnes  $Pe_i$ . Une extension de ce type de graphe est d'ajouter des poids sur les relations entre personnes et propriétés qui représentent l'importance d'une propriété pour une personne. Bien que plus riche sémantiquement parlant, nous traiterons ce cas dans un article futur.

Un petit exemple de cette configuration est le graphe biparti intensément étudié (voir la méta-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 (voir figure 2). Chaque colonne représente un événement de type "Tea party" et chaque ligne représente les dames y ayant participé. Les propriétés sont donc les évènements représentés en colonnes.

#### 3.2 Entropie d'un système

L'entropie d'un système représente le degré d'information que comporte ce système. Elle se calcule généralement comme suit : si un système comporte  $n$  éléments, l'entropie (voir (Yang & Petersen, 1997)) est la quantité :  $H(S) = \frac{-\sum_{i=1}^n P(i) \times \ln(P(i))}{\ln(n)}$ . Afin d'obtenir des propriétés intéressantes de la fonction d'entropie, le facteur de normalisation  $\ln(n)$  est utilisé. Dans ce cas l'entropie varie entre 0 et 1 (voir figure 3) et est maximum quand les probabilités des  $n$  éléments sont à 0,5 soit équirépartis dans le système.

La signification de l'entropie est donc qu'elle permet de représenter la quantité d'information fournie par le système, plus les éléments sont équirépartis dans le système, plus il faut d'information pour les représenter. Si chaque élément est par contre de probabilité soit 0 soit 1 c'est à dire parfaitement défini sans incertitude, alors le système a besoin de moins d'information pour le représenter, ce qui correspond à une valeur d'entropie plus faible. Dans le cas d'un réseau social représenté par un graphe biparti l'entropie d'un système se calcule sur l'ensemble des personnes partageant des propriétés. Chaque propriété a une probabilité d'apparition dans


 FIGURE 3 – courbe entropie :  $-\sum n \times \ln(n)$ 

le système, c'est à dire qu'elle est associée par une arête à une ou plusieurs personnes. Le système défini ici est donc : **un ensemble de personnes partageant des propriétés**. Dans notre exemple, une dame qui participera à un nombre élevé d'événements aura une entropie plus forte.

L'entropie se formule donc comme suit :  $H(S) = \frac{-\sum_{j=1}^{n_2} P(Pr_j) \times \ln(P(Pr_j))}{\ln(n_2)}$ , où la probabilité  $P(Pr_j)$  est définie par  $P(Pr_j) = \frac{\sum_{i=1}^{n_1} A(i,j)}{m}$ .  $P(Pr_j)$  est donc la probabilité d'apparition de la propriété  $Pr_j$  pour toutes les personnes du réseau dans le système  $S$ .

### 3.3 Entropie conditionnelle

L'entropie conditionnelle représente la quantité d'information associée à un élément du système. Dans le cas du réseau social, nous mesurons avec cette entropie la quantité d'information portée par une personne, en tenant compte de la probabilité de présence d'une personne :

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

ou de non-présence d'une personne :

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

Dans notre exemple, c'est l'entropie de l'ensemble des événements connaissant la présence ou l'absence d'une dame aux événements.

### 3.4 Gain d'information

Le gain d'information toujours selon (Yang & Petersen, 1997) représente la quantité d'information spécifique différentielle apportée par un élément, dans notre cas une personne du réseau social. Selon (Yang & Petersen, 1997) c'est la différence des entropies du système et des entropies conditionnelles pour une personne  $Pe_j$  et son complémentaire  $\overline{Pe_j}$  :

$$G(Pe_j) = H(S) - P(Pe_j) \times H(S/Pe_j) - P(\overline{Pe_j}) \times H(S/\overline{Pe_j}) \quad (2)$$

Si un système doté d'une personne a une entropie plus faible que le système tout seul, c'est le signe que la personne en question diminue l'entropie, et donc apporte de l'information. Plus les entropies conditionnelles de la personne  $Pe_j$  sont faibles plus le gain d'information est

élevé, signifiant le fait que la personne  $Pe_j$  contribue fortement à l'apport d'information pour le système ou dit autrement fait baisser l'incertitude du système.

Dans notre exemple, c'est le gain apporté par la connaissance de la participation d'une dame aux événements.

## 4 Complémentarité

### 4.1 Notre définition de la complémentarité

Une problématique souvent évoquée concerne l'utilité des individus dans un groupe. Une des questions se pose alors : les personnes formant un groupe ont-elles un intérêt à être associées au même groupe. Un critère de cet intérêt est la capacité des individus à être complémentaires au sein d'un groupe au sens des propriétés partagées. Dans notre étude, les graphes bipartis sont très bien adaptés pour répondre à ce type de question. La complémentarité a pour but de trouver des éléments complémentaires dans un système. Le principe directeur dans notre cas est donc de trouver les éléments d'un graphe biparti ou réseau social (les lignes de la matrice d'incidence) qui couvre l'ensemble des propriétés (les colonnes de la matrice d'incidence) avec le minimum d'éléments de ce graphe

Dans notre exemple, c'est la mesure d'utilité apportée par la présence simultanée de deux dames aux "tea parties" comparativement à leur présence individuelle.

Dans un graphe biparti, les personnes partagent des propriétés. Ces propriétés sont par exemple : l'appartenance à une photo, la coécriture d'un article, l'appartenance à une thématique, un concept associé à une personne, un "<j'aime"> de facebook, un "<suivre"> sur twitter, etc. Ainsi dans le cas de concepts ou thématiques partagés par exemple, il est intéressant de connaître le nombre minimum de personnes couvrant l'ensemble des thématiques.

Nous adoptons la notion de complémentarité de Cassiman qui permet mieux que les autres visions de la complémentarité de représenter l'apport supplémentaire de l'association de personnes, ce qui est l'essence même d'un réseau social. en y incorporant le concept d'entropie et d'information mutuelle. La notion de multi-membre telle qu'évoquée dans l'état de l'art ne prend pas en compte la notion d'utilité. La complémentarité entre deux personnes, conformément à la formule 1, est donc basée sur le fait suivant : la présence simultanée de deux personnes apporte une plus grande utilité ("fonction" d'utilité  $\Pi$ ) au système que chacun des deux éléments séparément.

La fonction d'utilité  $\Pi$ , c.f. formule 1, peut être définie de différentes façons, et dans notre cas nous utiliserons le concept d'entropie et d'information mutuelle.

### 4.2 Le gain d'information comme fonction d'utilité

Nous utiliserons la mesure d'entropie et ses dérivés pour représenter l'utilité car c'est la mesure d'information apportée par un élément dans un ensemble de classe d'éléments (Yang & Petersen, 1997), et elle est très pertinente dans le contexte du partage d'information et de propriétés que porte un graphe biparti. Dans notre cas nous pouvons considérer la fonction d'utilité comme le gain d'information défini plus haut. Plus le gain d'information apporté par deux éléments au système est important, plus ces deux éléments en question sont utiles ensemble pour

le système. La complémentarité se traduit donc par :

$$\text{complémentarité}(Pe_i, Pe_j) = G(Pe_i, Pe_j) - G(\overline{Pe_i}, Pe_j) - (G(Pe_i, \overline{Pe_j}) - G(\overline{Pe_i}, \overline{Pe_j})), \quad (3)$$

si l'on développe cette formule cela donne :

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

soit après simplification :

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

ou sous une autre forme :

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

le but sera donc d'évaluer si cette complémentarité est positive, qui signifie que les deux éléments  $Pe_i$  et  $Pe_j$  ont intérêt à être associés dans le système.

### 4.3 Algorithme de la complémentarité

Ainsi nous cherchons à trouver l'ensemble minimum des personnes qui couvrent au mieux l'ensemble des propriétés du système. Pour cela, nous effectuons plusieurs étapes :

1. Calculer l'entropie du système
2. Chercher la personne  $Pe_{max}$  ayant le maximum de gain d'information (formule 2). Si plusieurs sommets présentent la même valeur maximum, alors prendre la personne offrant le maximum de couverture des événements
3. Chercher la personne  $Pe_{comp\_max}$  ayant la complémentarité la plus forte avec  $Pe_{max}$  (formule 3)
4. **condition d'arrêt** : est-ce que toutes les propriétés ont été couvertes au moins une fois ?  
si oui : arrêt,  
si non : continuer au point suivant

personnes	Gain	personnes	Gain
1	<b>0,2590131</b>	10	0,09085268
2	0,21040042	11	0,09085268
3	<b>0,2590131</b>	12	0,16715123
4	0,21040042	13	0,21040042
5	0,09085268	14	<b>0,2590131</b>
6	0,09085268	15	0,12768425
7	0,09085268	16	0,02141921
8	0,05572812	17	0,02141921
9	0,09085268	18	0,02141921

FIGURE 4 – Mesures du gain d’information pour les personnes du réseau WE

- Chercher la personne ayant la complémentarité la plus forte avec  $Pe_{comp\_max}$  et  $Pe_{max}$  qui s’exprime ainsi :  

$$complémentarité(Pe_i, Pe_{comp\_max} \wedge Pe_{max}) = G(Pe_i, Pe_{comp\_max} \wedge Pe_{max}) - G(\overline{Pe_i}, Pe_{comp\_max} \wedge Pe_{max}) - (G(Pe_i, \overline{Pe_j}) + G(\overline{Pe_i}, \overline{Pe_{comp\_max} \wedge Pe_{max}})),$$
- continuer sur tous les sommets jusqu’à ce que la complémentarité de tous les sommets du graphe soit négative
- si l’ensemble des sommets a été parcouru et toutes les complémentarités sont positives, s’arrêter sinon reprendre au 4.

l’ensemble des sommets ainsi identifiés constitue le groupe de sommets présentant une complémentarité maximale pour l’ensemble du graphe. Dans le cas où l’on s’arrête avant d’avoir couvert l’ensemble complet des personnes, la couverture (connexion avec toutes les propriétés) est totale. Dans le second cas la couverture est partielle.

#### 4.4 Complexité du calcul de complémentarité

La complexité est de  $O(m \times n^2)$ ,  $n$  étant le nombre de personnes,  $m$  étant le nombre de propriétés. En effet chaque calcul d’entropie s’effectue sur l’ensemble des propriétés, et le calcul de complémentarité requiert au maximum  $\frac{n \times (n-1)}{2}$  opérations.

### 5 Expérimentation

Nous utilisons 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 (voir figure 2).

#### 5.1 Mesure d’information mutuelle et de complémentarité sur le graphe biparti WE

Le premier calcul est selon les étapes 1 et 2 de la liste du paragraphe 4.3 celui du gain d’information de chaque sommet ou personne selon la formule 2. la figure 4 montre les valeurs pour chaque sommet. Ce calcul donne trois sommets avec une entropie maximum, nous choisissons le premier de ces sommets, le sommet 1, qui offre une couverture maximale des évènements.

Ensuite selon toujours la liste du paragraphe 4.3 nous calculons la complémentarité de chaque sommet/personne avec le sommet 1 que nous montrons dans la figure 5.1. Ce calcul



personnes	complémentarité	personnes	complémentarité	personnes	complémentarité	personnes	complémentarité
		10	0,032024786			10	-0,205307041
2	-0,241284557	11	0,095461542	2	-0,056074702	11	-0,171822694
3	-0,249998904	12	0,198456905	3	-0,126066811	12	-0,153876567
4	-0,175434381	13	<b>0,267897921</b>	4	-0,074081708	13	
5	-0,126538143	14	0,206555642	5	-0,011852276	14	<b>0,187685538</b>
6	-0,172270819	15	0,090862253	6	-0,094459499	15	-0,142052887
7	-0,113088838	16	-0,092933968	7	-0,137395416	16	-0,202092756
8	-0,146138666	17	0,016246178	8	-0,207179604	17	-0,098646124
9	-0,051317974	18	0,162265216	9	-0,205307041	18	0,166238366

FIGURE 5 – Mesures de complémentarité avec la personne 1 puis 1 et 13 du réseau WE

donne le sommet 13 comportant la complémentarité maximum avec le sommet 1. On constate à ce point que la majorité des sommets comportent une complémentarité négative, ce qui signifie que l'on a atteint un quasi équilibre en choisissant les sommets 1 et 13 qui représentent donc la grande majorité des sommets du graphe en terme de représentativité.

Puis selon toujours la liste du paragraphe 4.3 nous calculons la complémentarité de chaque sommet/personne avec les sommet 1 et 13 et que nous montrons dans la figure 5.1. Ce calcul donne le sommet 14 comportant la complémentarité maximum avec les sommets 1 et 13. On constate à ce point que tous les événements du graphe sont couverts, c'est à dire la condition d'arrêt 4 de l'algorithme est vérifiée. Nous obtenons donc un premier résultat, la couverture complète par trois personnes sur le graphe des Women Events. Les trois personnes sont considérées comme un des meilleurs arrangements de couverture complémentaire pour ce graphe.

## 5.2 Mesure de complémentarité sur d'autres jeux de données plus importants

Nous avons expérimenté notre méthode sur le graphe du club de karaté Zachary (1977) un jeu de données bien connu dans l'analyse de graphes. Ce graphe représente les relations d'affinités entre les personnes d'un club de karaté. C'est un graphe monoparti, cependant nous pouvons le considérer comme un graphe biparti, dans lequel les personnes partagent des propriétés (les propriétés étant dans ce cas les personnes liées par des liens d'amitié). Nous obtenons des résultats similaires au graphes des Women Events. La couverture totale est obtenue avec 25 personnes sur 34.

Nous avons appliqué les mêmes procédures à un autre 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. C'est un graphe biparti. Sur ce jeu de données le nombre d'itérations nécessaires pour obtenir la couverture complète est égal au nombre de personnes. Après analyse du jeu de données, on constate qu'aucune combinaison des personnes inférieure au nombre total ne permet d'obtenir une couverture complète. Les photos considérées correspondent à des périodes différentes dans le temps et donc les personnes appartiennent souvent à des groupes disjoints. En effet le jeu de photos est effectué avec des groupes de personnes pratiquement disjoints correspondant aux différentes périodes de vie du propriétaire du compte facebook. Nous pouvons en déduire que pour obtenir une complémentarité avec un nombre de personnes réduit, les personnes doivent partager un minimum les propriétés considérées. Ce qui n'est pas le cas dans ce jeu de données. Notre méthode met donc en lumière sur ce jeu de données la disparité des personnes et leur difficulté à devenir complémentaires.

Nota : ces travaux ont été réalisés grâce au concours d'un groupe de trois élèves de niveau Master de l'Ecole des Mines d'Alès.

## 6 Conclusion

Dans les réseaux sociaux trouver des groupes de personnes stables, peut s'avérer intéressant pour différents objectifs. Après avoir exploré les différentes voies, nous utilisons la notion de complémentarité pour exprimer cette stabilité. Après avoir défini la notion de complémentarité, nous montrons comment utiliser la notion d'Information Mutuelle pour exprimer cette complémentarité. Nous montrons à travers plusieurs jeux de données que cette mesure est prometteuse pour évaluer la stabilité d'un ensemble de personnes socialement complémentaires.

## Références

- BERGE C. (1987). Hypergraphes, Combinatoires des ensembles finis. *Gauthier-Villars*.
- CASSIMAN B. & VEUGELERS R. (2006). In Search of Complementarity in Innovation Strategy : Internal R&D and External Knowledge Acquisition. *Management Science*, **52**(1), 68–82.
- EITER T. & GOTTLOB G. (1995). Identifying the minimal transversals of a hypergraph and related problems. *SIAM Journal on Computing*, **24**, 1278–1304.
- FORTUNATO S. (2009). Community detection in graphs. *Physics Reports*, **486**(3-5), 103.
- 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.
- JELASSI M. N., LARGERON C. & YAHIA S. B. (2014). Efficient unveiling of multi-members in a social network. *Journal of Systems and Software*, **94**(0), 30 – 38.
- KAVVADIAS J. & STAVROPOULOS E. (2005). An efficient algorithm for the transversal hypergraph generation. *Journal of Graph Algorithms and Applications*, **9**, 239–264.
- KHACHIYAN L., BOROS E., ELBASSIONI K. & GURVICH V. (2005). Computing and combinatorics. *Lecture Notes in Computer Science*, **3595**, 767–776.
- MA Q., NADAMOTO A. & TANAKA K. (2006). Complementary information retrieval for cross-media news content. *Inf. Syst.*, **31**(7), 659–678.
- PAPADOPOULOS S., KOMPATSIARIS Y., VAKALI A. & SPYRIDONOS P. (2011). Community detection in Social Media. *Data Mining and Knowledge Discovery*, **1**(June), 1–40.
- PORTER M. A., ONNELA J.-P. & MUCHA P. J. (2009). Communities in Networks.
- TOPKIS D. (1998). *Supermodularity and Complementarity*. Frontiers of Economic Research. Princeton University Press.
- YANG B., LIU D., LIU J. & FURHT B. (2010). *Discovering communities from Social Networks : Methodologies and Applications*. Boston, MA : Springer US.
- YANG Y. & PETERSEN J. O. (1997). A comparative Study on Feature Selection in Text Categorisation. In *Fourteenth International Conference on Machine Learning, ICML'97*.
- ZACHARY W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, **33**(4), 452–473.