



Organisation de communautés et Équilibre de Nash

Michel Crampes, Michel Plantié

► To cite this version:

Michel Crampes, Michel Plantié. Organisation de communautés et Équilibre de Nash. IC - 25èmes Journées francophones d'Ingénierie des Connaissances, May 2014, Clermont-Ferrand, France. pp.151-162. hal-01015384

HAL Id: hal-01015384

<https://inria.hal.science/hal-01015384>

Submitted on 26 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Organisation de communautés et Equilibre de Nash

Michel Crampes¹, Michel Plantié¹

LABORATOIRE LGI2P, Ecole des Mines d'Ales, Site de Nimes
Parc Georges Besse, 30035 Nimes Cedex, France
{michel.crampes,michel.plantie}@mines-ales.fr

Résumé :

La détection de communautés dans les réseaux sociaux est devenue un enjeu majeur pour extraire de la connaissance, observer l'émergence d'opinions ou de savoirs et les réinjecter de manière ciblée sur des groupes sociaux particuliers. Les nombreux algorithmes de détection déjà proposés fournissent des solutions approchées sur laquelle l'analyste ou même les entités regroupées ne peuvent guère intervenir. Dans cet article nous proposons un renversement de perspective. Nous considérons les entités comme des agents qui cherchent à se regrouper et qui peuvent agir sur leur stratégie de regroupement afin d'en tirer le meilleur bénéfice avec différentes règles partagées. Les communautés sont optimisées via la recherche d'un Equilibre de Nash. Cet article explore plus avant ce paradigme en considérant les différentes possibilités offertes par la recherche de l'Equilibre de Nash en matière de réorganisation pilotée par la sémantique, l'intention pragmatique, ou l'évolution spontanée du réseau.

Mots-clés : Equilibre de Nash, Réseaux sociaux, Détection de Communautés.

1 Introduction

La connaissance est indissociable des réseaux sociaux et des communautés. Les opinions, les savoirs et les actions émergent et se structurent au sein de communautés formellement identifiées ou spontanément regroupées. A l'inverse l'observation et l'analyse de regroupements d'entités (classement) font éminemment partie des méthodes de construction de nouveaux savoirs. Les méthodes de classification classiques comme K-Means, l'Analyse en Composantes Principales, la classification hiérarchique, ou la méthode MDS nécessitent de définir à priori le nombre de classes pour certaines, ou un critère d'arrêt portant sur une fonction de coûts. De plus ces méthodes s'appliquent sur des données vectorielles. Les méthodes récentes de détection de communautés ont pour données de départ un graphe uniparti ou un graphe biparti, voire multipartis, en général non orienté. Nous avons montré dans Crampes & Plantié (2013) que la recherche de communautés dans les graphes unipartis, bipartis, orientés ou non, pourrait se ramener à la recherche de communautés dans les graphes unipartis.

Les algorithmes de détection de communautés sont pilotés par l'optimisation d'une fonction. Celle qui est la plus considérée est la modularité depuis les travaux de Newman (2006). Elle consiste à considérer un maximum de liens à l'intérieur d'une communauté avec le minimum de liens vers l'extérieur de la communauté. Il est intéressant de noter que certaines méthodes, comme celui de Louvain, considèrent des liens pondérés. A l'inverse des fonctions de clustering ces méthodes ne nécessitent pas de donner en préalable le nombre de communautés ou un seuil sur la fonction. Cependant elles présentent certaines limites qui dans nombre de situations peuvent être gênantes pour une analyse plus sûre et plus détaillée. La première tient au fait que, étant donné le caractère NP-Complet de la recherche de communautés, toutes ces méthodes sont des heuristiques qui fournissent une solution approchée, la modularité maximale n'étant

pas assurée. Dans un article précédent Crampes *et al.* (2013) nous avons montré la possibilité d'améliorer ce résultat en considérant les individus comme des agents engagés dans un jeu non coopératif qui cherchent à optimiser une fonction de gain partagée. Celle-ci, qui est associée à la modularité, nous permet d'introduire une 'fonction potentiel' et de conclure qu'il existe un équilibre de Nash. Autrement dit, partant d'un résultat intermédiaire obtenu par un algorithme de détection de communautés comme par exemple celui de Louvain, il est possible de réaffecter les individus de manière à obtenir un véritable optimum de modularité sans toucher au nombre de communautés déjà trouvées. Ces résultats seront rappelés dans le présent article.

La seconde limite des algorithmes de détection de communautés porte sur le résultat unique du nombre de communautés. Il n'est pas possible de déterminer a priori ce nombre de communautés, ou mieux de réviser le nombre de communautés à partir du résultat final pour obtenir une nouvelle organisation en fonction de nouvelles finalités, ou tout simplement pour chercher à améliorer la modularité. Dans cet article nous montrons comment élargir nos résultats sur l'équilibre de Nash soit pour faire varier le nombre de communautés, soit pour diriger l'organisation des communautés en fonction de la sémantique, celle-ci étant donnée par les propriétés des agents dans un graphe biparti.

La section suivante présente un état de l'art de la détection de communautés en mettant en avant les travaux récents qui utilisent l'équilibre de Nash et leur contribution en la matière.

2 État de l'art

La recherche automatique de communautés est devenue un champ de recherche très exploré comme l'atteste l'état de l'art approfondi de Fortunato (2009). La plupart des algorithmes proposent de maximiser une mesure appelée modularité introduite par Newman Newman (2006) et largement appliquée d'abord aux graphes unipartis. Elle a ensuite été étendue aux graphes bipartis dans un premier temps en adaptant sa formulation Murata (2010); Suzuki & Wakita (2009). De nombreuses méthodes de détection de communautés ont été développées à la fois pour les graphes unipartis et bipartis ; le lecteur intéressé trouvera un synthèse dans Crampes & Plantié (2013).

Dans ce même article récent Crampes & Plantié (2013) nous montrons qu'il est possible d'unifier les graphes unipartis, bipartis, et orientés pour produire des communautés à la fois partitionnées et recouvrantes. Nous mettons en œuvre l'algorithme de Louvain Blondel *et al.* (2008) qui fonctionne en agrégeant itérativement les sommets du graphe afin d'augmenter la modularité au maximum. Cependant dans le cas général cet algorithme, comme toute heuristique, ne produit qu'un résultat approché, c'est-à-dire que l'algorithme s'arrête à partir du moment où la modularité ne peut plus augmenter. Or la fonction de recouvrement montre que certains sommets seraient incités à changer de communauté et que ce changement pourrait améliorer ou inversement abaisser le résultat obtenu. Autrement dit le résultat obtenu n'est pas stable. La recherche de communautés stables a fait l'objet de peu de publications, les auteurs se satisfaisant en général des résultats de l'algorithme qu'ils mettent en œuvre et les comparant aux résultats d'autres auteurs. La recherche de stabilité d'un réseau en termes de théorie des jeux a fait l'objet de publications comme par exemple Nisan *et al.* (2007). Selon cette approche la recherche de stabilité de n agents qui choisissent des stratégies à partir de fonctions de satisfaction suppose l'existence d'un Equilibre de Nash. Appliquée à la détection de communautés le problème consiste à trouver les conditions d'existence d'un Equilibre de Nash tel qu'aucun sommet

ne souhaite au final quitter la communauté à laquelle il a été affecté. L'Equilibre de Nash n'a été que peu utilisé pour détecter des communautés. R Narayanam & Y Narahari (2012) l'appliquent sur des graphes unipartis. Ils utilisent la connectivité des sommets pour atteindre un Equilibre de Nash sans mesurer la modularité du résultat. Chen *et al.* (2011) se focalisent également sur les graphes unipartis avec des communautés recouvrantes, et la recherche de l'Equilibre de Nash est l'unique principe directeur. Mais les résultats expérimentaux ne nous semblent pas meilleurs que ceux trouvés par d'autres algorithmes comme celui de Louvain. De plus ces algorithmes ne sont appliqués que sur des graphes unipartis. Dans Crampes *et al.* (2013) nous nous sommes distingués de ces auteurs en recherchant d'abord une solution approchée à l'aide de l'algorithme de Louvain, puis en effectuant des réaffectations pour converger vers un équilibre local qui est prouvé être un Equilibre de Nash. De plus nous avons appliqué cette méthode avec de bons résultats en matière de modularité sur les trois types de graphes. Les deux sections suivantes reprennent synthétiquement ces résultats avant d'introduire des contributions originales.

3 Détection de communautés partitionnées

3.1 Modularité

Le consensus pour la détection de communautés consiste à rechercher une solution approchée qui maximise la modularité selon Newman (2006). Formellement, étant donné un graphe uniparti $G = (N, E)$ représenté par sa matrice d'adjacence A , la modularité Q d'une partition de graphe est définie :

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (1)$$

où A_{ij} représente le poids de la liaison entre i et j , $k_i = \sum_j A_{ij}$ est la somme des poids des arcs attachés au sommet i , c_i est la communauté à laquelle appartient le sommet i , la fonction de Kronecker $\delta(u, v)$ est égale à 1 si $u = v$ et à 0 sinon, enfin $m = 1/2 \sum_{i,j} A_{ij}$. Pour l'instant nous ne considérons que des graphes binaires et dans ce cas les poids A_{ij} prennent les valeurs 1 ou 0 selon que la liaison existe ou n'existe pas. L'interprétation de cette formule est la suivante : la modularité est la somme pondérée pour toutes les communautés de la différence entre les liaisons observées à l'intérieur de la communauté (terme A_{ij}) et la probabilité de ces liaisons (terme $\frac{k_i k_j}{2m}$ dont le numérateur est le produit des marges correspondant à la cellule i, j). L'application de cette fonction par de nombreux algorithmes donne de bons résultats. Par exemple elle permet de retrouver des communautés sur des graphes construits ad-hoc. Cependant il a été démontré dans Fortunato (2009) que cette fonction a tendance à fusionner les petites communautés et ainsi à masquer une certaine granularité. Pour les graphes bipartis (et les graphes orientés qui peuvent se ramener à des graphes bipartis) d'autres formulations ont été proposées. En particulier la formulation dans Barber & Clark (2009) semble faire consensus. Peu différente de celle de Newman, nos travaux récents montrent qu'elle pourrait aussi s'appliquer aux graphes unipartis. Cependant nous utiliserons la formulation de Newman quel que soit le type de graphe pour rester conforme à nos travaux antérieurs publiés afin d'unifier les trois types de graphes. Ayant défini la fonction à optimiser lors de la détection de communautés, nous regardons quel algorithme de détection appliquer.

3.2 Algorithme efficace de détection de communautés partitionnées

Il existe de très nombreuses méthodes de détection de communautés, la plupart optimisant dans la mesure du possible la modularité au sens de Newman pour les graphes unipartis et la modularité au sens de Barber pour les graphes bipartis. Toutes ces méthodes ont en commun d'être des heuristiques puisque la recherche de communautés est par essence un problème NP-complet. Elles fournissent donc des solutions approchées tant en nombre de communautés qu'en matière de répartition des individus entre les communautés. La valeur de la modularité finale n'est pas nécessairement optimale. Dans la section suivante nous présenterons une méthode de réaffectation introduite dans Crampes *et al.* (2013) et basée sur l'Equilibre de Nash pour rechercher à coup sûr un optimum de modularité sans changer le nombre de communautés. Cette méthode, pour l'instant limitée aux liaisons non pondérées, peut s'appliquer au résultat de n'importe quel algorithme de détection de communautés. Nous présenterons ensuite la contribution essentielle de cet article qui portera sur l'application de cette méthode de réaffectation pour introduire des contraintes externes à la création de communauté tant d'un point de vue quantitatif (nombre de communautés, ou nombre d'individus dans les communautés) que qualitatif (répartition des individus, ou répartition des propriétés dans les communautés). Notre approche permet ainsi de manière optimale de passer de la détection de communautés à la construction optimale de communautés sous contraintes, et en particulier de contraintes sémantiques.

Dans les expériences que nous avons menées la méthode que nous utiliserons pour effectuer une première recherche de communauté avant réaffectation est l'algorithme de Louvain Blondel *et al.* (2008) que nous avons présenté dans Crampes & Plantié (2013) et dans Plantié & Crampes (2013). Elle est remarquable par son efficacité et la qualité de ses résultats. Cependant, de nature heuristique elle donne un résultat approché pour lequel la modularité n'est pas optimale. Nous faisons maintenant intervenir une fonction de réaffectation pour rechercher cet optimum.

4 Réaffectation et Equilibre de Nash

L'application d'une méthode de détection comme celle de Louvain nous permet d'obtenir un ensemble C de n communautés. Ces communautés peuvent comporter des individus homogènes (cas d'un graphe uniparti) ou appartenant à deux classes (graphes bipartis). La fonction de réaffectation qui est présentée permet de réaffecter les individus (et les propriétés pour les graphes bipartis) sur les communautés déjà trouvées.

4.1 Fonction de réaffectation RM

Afin de définir cette fonction de réaffectation nous utilisons une variante de l'équation 1 pour les graphes non pondérés. Soit C_i une communauté, $|e_i|$ le nombre d'arêtes dans C_i , d_{C_i} la somme des degrés des noeuds dans C_i et m le nombre total de liens dans le graphe. Alors la modularité est :

$$Q = \sum_i \left[\frac{|e_i|}{m} - \frac{(d_{C_i})^2}{(2m)^2} \right] \quad (2)$$

L'interprétation est la suivante : la modularité est la somme pour toutes les communautés du nombre de liens relatifs dans chaque communauté moins la probabilité d'avoir des liens dans cette communauté.

Réaffecter un sommet w de C_1 à C_2 accroît ou décroît la modularité. Nous définissons ce changement comme la mesure de réaffectation de modularité $RM_{w:C_1 \rightarrow C_2} = Q_{w \in C_2} - Q_{w \in C_1}$ où $Q_{w \in C_k}$ est la valeur de la modularité pour $w \in C_k$ et $C_1 \neq C_2$.

Soit $l_{w|i}$ le nombre d'arêtes entre un sommet w et tous les autres sommets w' tels que $w' \in C_i$, et soit d_w le degré de w . Nous considérons que le sommet w appartenant à C_1 est retiré de cette communauté et ensuite réaffecté à une autre communauté C_2 . Nous avons montré dans Crampes *et al.* (2013) que

$$RM_{w:C_1 \rightarrow C_2} = \frac{1}{m}(l_{w|2} - l_{w|1}) - \frac{1}{2m^2}[d_w^2 + d_w(d_{C_2} - d_{C_1})] \quad (3)$$

En appliquant l'équation 3, il est possible de vérifier que si nous réaffectons w de C_1 à C_2 puis à nouveau de C_2 à C_1 , la modularité ne change pas (le calcul se vérifie de différentes manières) et $RM_{w:C_1 \rightarrow C_2 \rightarrow C_1} = 0$. En conséquence : $RM_{w:C_1 \rightarrow C_2} = -RM_{w:C_2 \rightarrow C_1}$.

4.2 Effet de la réaffectation sur les autres sommets

Nous supposons, qu'une première passe de calcul des réaffectation a été effectuée, et à la suite de cela un sommet w du graphe a été déplacé de C_1 à C_2 , suite à une mesure de réaffectation positive. Dans notre cas nous choisissons de réaffecter le sommet le plus instable, c'est à dire celui dont la valeur RM est la plus forte. Soit z un autre sommet du graphe. On peut voir la variation de la valeur de réaffectation pour ce sommet z après le déplacement de w .

Le calcul de la différence de la mesure de réaffectation pour le sommet z de l'étape précédente à l'étape actuelle se calcule comme suit :

On cherche $RM_{z:C_{from} \rightarrow C_{to}}^1 - RM_{z:C_{from} \rightarrow C_{to}}^0$ où $RM_{z:C_{from} \rightarrow C_{to}}^0$ est la mesure de réaffectation de z avant le déplacement de w et $RM_{z:C_{from} \rightarrow C_{to}}^1$ sa mesure après.

Selon les cas de C_{from} et C_{to} on trouve les résultats suivants :

Soit $\Delta R_z = [\{w, z\} - \frac{1}{(2m)}d_z d_w] \frac{1}{m}$, dans lequel $\{w, z\}$ représente le lien entre w et z . S'il n'y a pas de lien cette valeur est nulle.

Dans le tableau ci dessous, nous montrons les différentes valeurs calculées individuellement, de la correction de réaffectation pour z . Dans ce tableau C_3 et C_4 sont des communautés autres que C_1 et C_2 .

to\from	C_1	C_2	C_3	C_4
C_1	0	$-2\Delta R_z$	$-\Delta R_z$	$-\Delta R_z$
C_2	$2\Delta R_z$	0	ΔR_z	ΔR_z
C_3	ΔR_z	$-\Delta R_z$	0	0
C_4	ΔR_z	$-\Delta R_z$	0	0

On observe que la matrice est antisymétrique, avec $RM_{z:C_{from} \rightarrow C_{to}}^1 - RM_{z:C_{from} \rightarrow C_{to}}^0 = -[RM_{z:C_{to} \rightarrow C_{from}}^1 - RM_{z:C_{to} \rightarrow C_{from}}^0]$, ce qui est compatible avec les propriétés de la réaffectation présentées en 4.1.

Ce tableau permet de simplifier les calculs de réaffectation. Il est aussi un moyen intéressant d'étudier l'impact sémantique de la réaffectation. Sans détailler les conclusions dans les limites de cet article, synthétiquement ces propriétés montrent qu'un noeud a tendance à suivre un noeud voisin, ou bien à quitter une communauté dans laquelle arrive un noeud avec lequel il n'a pas de liaison.

4.3 Réaffectation par l'Equilibre de Nash (EN)

Nous montrons dans cette section que l'application de la réaffectation conduit à un Equilibre de Nash, une situation stable où aucun sommet n'a intérêt à quitter la communauté à laquelle il a été finalement affecté. Dans ce but nous ramenons le problème de la réaffectation des sommets à un problème de théorie des jeux non coopératifs. Un lecteur soucieux d'une expression formelle pourra se référer à Crampes *et al.* (2013). En effet les n sommets peuvent être considérés comme n joueurs qui cherchent à optimiser leur gain en jouant des stratégies. Une stratégie est un déplacement accessible à un joueur. Dans notre cas une stratégie est le choix d'une communauté et chaque agent a le choix parmi les m différentes communautés. Un profil de stratégie s à un moment donné est la combinaison des choix stratégiques des agents, c'est-à-dire dans notre cas l'état des affectations à un moment donné.

Une fonction de gain est ce qu'est en droit d'attendre un joueur i en jouant une stratégie compte tenu des stratégies que jouent les autres joueurs. Prenant en compte cette fonction de gain pour chaque joueur, un profil de stratégie s^* est un Equilibre de Nash (EN) si aucun joueur n'a intérêt à changer de stratégie étant donné la stratégie jouée par chacun des autres joueurs. Trois questions se posent : (i) quelles sont les conditions pour qu'il existe au moins un EN, (ii) s'il en existe un comment l'atteindre, et (iii) est-il possible de l'atteindre dans un temps raisonnable polynomial.

Il est possible de donner une réponse à l'existence d'un EN et à la possibilité de converger vers cet équilibre si on peut définir une "fonction potentiel" qui permet d'atteindre un optimum global en recherchant des optima locaux pour les agents. Si de plus la recherche d'un optimum local pour un agent se fait en un temps polynomial, alors le problème entre dans la classe des problèmes PLSComplets (Polynomial Local Search) (E. Tardos & T. Wexler Nisan *et al.* (2007)).

Pour tout jeu fini, une fonction potentiel exacte Φ est une fonction qui fait correspondre à chaque vecteur stratégie s une valeur réelle $\Phi(s)$ avec la condition qu'à une évolution du gain pour un agent correspond la même évolution du gain pour la fonction potentiel. Autrement dit les joueurs partagent un intérêt commun lorsqu'ils choisissent individuellement leur stratégie, même s'ils ne coopèrent pas. L'intérêt de chacun rejoint l'intérêt du groupe. Ainsi armé d'une fonction potentiel il est possible de trouver sûrement un Equilibre de Nash en convergeant à partir de la recherche d'optima locaux pour chaque joueur. Pour trouver de manière convergente un Equilibre de Nash, c'est-à-dire une partition stable des communautés qui satisfasse tout le monde, il nous faut disposer d'une fonction potentiel $\Phi(s)$. La fonction $RM_{w:C_1 \rightarrow C_2}$ qui représente le gain f_w attendu par le sommet w quand il est réaffecté est en fait le gain de modularité pour toutes les communautés. Cette fonction $RM_{w:C_1 \rightarrow C_2}$ peut en conséquence jouer le rôle de fonction potentiel.

Ce choix de gain local assure donc que l'algorithme local de réaffectation des noeuds présenté dans la section précédente converge bien vers un Equilibre de Nash. Dans la mesure où le calcul de RM est fait à chaque étape de réaffectation uniquement pour les noeuds des deux communautés concernées et que ce calcul est polynomial, l'algorithme de convergence vers l'Equilibre de Nash est PLSComplet (Papadimitriou Nisan *et al.* (2007)).

Il est intéressant de noter que récemment quelques auteurs ont fait appel plus largement à l'Equilibre de Nash pour rechercher les communautés partitionnées R Narayanam & Y Narahari (2012); Chen *et al.* (2011). La méthode que nous avons présentée ici de détection de

3	RM	-25,38		-25,38						12,50	5,18	-1,39	-12,62	-30,93	-29,04	-4,80	-10,48	0	0	0
2		-47,09	-47,78	-47,09	-47,78			-16,92	-16,92	4,36	-5,68	0	0	0	0	0	0	-5,68		
1		0	0	0	0	0	0	0	0	0	0	-18,81			-59,53	-39,64	-32,26			
Women		W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18	
Events		E1	E2	E3	E4	E5	E6	E7			E8	E10	E12	E13	E14			E9	E11	
1	RM	0	0	0	0	0	0	0			-1,64							-50,12		
2							-58,33	-12,37			0	0	0	0	0			-19,19	-11,87	
3											-19,32							0	0	

FIGURE 1 – Communautés et mesures de réaffectation pour : Women Events

communautés s'éloigne de ces auteurs parce qu'elle accepte dans un premier temps d'autres heuristiques efficaces bien établies pour rechercher une solution approchée et trouver le nombre de communautés, puis elle finalise la recherche pour atteindre un Equilibre de Nash qui assure la stabilité. A l'inverse les autres auteurs utilisent la recherche de l'Equilibre de Nash pour construire les communautés et réaliser les affectations. De plus leur fonction potentiel diffère de celle que nous avons présentée ici. Nous avons montré dans Crampes *et al.* (2013) que la combinaison Louvain-Réaffectation donne de meilleurs résultats que la recherche de communautés avec uniquement l'Equilibre de Nash. Il est cependant un point essentiel à faire observer. La fonction potentiel que nous utilisons fait que chaque sommet cherche à améliorer la modularité de tout le graphe. Autrement dit nous avons là une fonction socialement très consensuelle dans laquelle l'intérêt de chacun s'identifie à l'intérêt de tous, un peu comme dans un essaim d'abeilles. Cette hypothèse très restrictive sera discutée à la fin de l'article.

4.4 Expérimentations

Des expérimentations de réaffectation ont été réalisées d'abord sur des graphes unipartis, et surtout bipartis de taille petite (de 40 à 1000 noeuds) Crampes *et al.* (2013). L'algorithme de Louvain est appliqué au graphe et quand il s'arrête les RM de tous les noeuds pour toutes les communautés sont calculés. Comme on verra sur l'exemple SW ci-dessous, notre algorithme de réaffectation entre alors en jeu, observe les individus ou les événements qui ont une propension à changer de communautés (RM positif), réaffecte le plus instable, et recalcule les RM pour les seuls noeuds de la communauté de départ et de la communauté d'arrivée. Ce calcul est réitéré jusqu'à ne plus obtenir de RM positifs. Un équilibre est effectivement atteint assez rapidement dans presque tous les cas expérimentés. Il s'agit d'un Equilibre de Nash où tous les noeuds sont stables (RM négatifs ou nuls). L'algorithme conserve le nombre de communautés trouvées par Louvain et la réaffectation permet d'améliorer la modularité. Sur ce type de graphes les temps de calcul qui dépendent du nombre de communautés sont de l'ordre de quelques secondes. Nous avons réalisé récemment des expérimentations sur des graphes bipartis moyens (environ 10.000 noeuds). Le temps de calcul va jusqu'à 1 minute pour 90 itérations (réaffectations). Pour un graphe de plus grande taille de 120.000 noeuds et 180 communautés trouvées par Louvain qui représente des chercheurs reliés aux articles dont ils sont auteurs le temps d'obtention de l'équilibre est de plus d'une heure dû à un nombre très important de réaffectations, l'heuristique

Individus :																		
Noeud	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
1 (7)	0	0	0	0	0	0	0	-10,983	-3,661	-18,811	-41,283	-68,299	-59,525	-39,641	-32,256	-18,432	-29,668	-29,668
2 (6)	-50,625	-50,877	-50,625	-50,877	-41,156	-18,685	-18,685	-7,385	-10,1	0	0	0	0	0	0	-4,797	-27,269	-27,269
3 (2)	-32,445	-46,206	-32,445	-46,206	-32,067	-20,831	-20,831	0	0	-1,389	-12,625	-30,93	-29,037	-4,797	-10,478	0	0	0

Evènements :														
Noeud	E1	E2	E3	E4	E5	E6	E7	E8	E10	E12	E13	E14	E9	E11
1 (7)	0	0	0	0	0	0	0	-1,641	-65,964	-79,535	-39,2	-39,2	-84,459	-37,369
2 (5)	-30,678	-30,678	-62,492	-41,156	-73,097	-50,625	-5,555	0	0	0	0	0	-36,359	-10,1
3 (0)	-23,861	-23,861	-48,857	-32,067	-43,681	-32,445	-16,538	-19,316	-44,186	-53,402	-26,133	-26,133	0	0

FIGURE 2 – Equilibre de Nash après réaffectation pour : Women Events

mise en oeuvre par Louvain produisant beaucoup d'instabilités.

Parmi les petits graphes nous détaillons ici celui qui permet de mieux cerner par la suite comment pourront s'appliquer des contraintes. Connu sous le nom Southwern Women (SW) il consiste en un relevé par cinq ethnologues de la participation différenciée de 18 dames à 14 événements sociaux de type "tea party", travail coopératif, etc. à Natchez, Mississipi en 1930. Le but des ethnologues était d'étudier les comportements raciaux au travers des relations individuelles. Formellement il s'agit d'un graphe biparti qui a été intensément étudié et pour lequel beaucoup d'auteurs ont proposé des partitionnements généralement en deux, parfois en trois communautés (voir la meta-analyse de Freeman (2003)).

Nous avons détecté trois communautés regroupant des dames et des événements avec Louvain. Elles sont montrées dans la figure 1 en couleur rouge, bleu et jaune (les individus sont numérotés de I1 à I18, les événements de E1 à E14). Ce résultat donne une modularité supérieure à tous les autres résultats trouvés dans la littérature et il est donc plus précis, en particulier comparé à ceux présentés dans Freeman (2003). La mesure de RM montre avec acuité la propension à être affecté à une autre communauté. La valeur 0 correspond à la réaffectation dans sa propre communauté. Concrètement deux individus (I8 et I9) sont instables. La figure 2 montre le résultat après réaffectation et obtention de l'Equilibre de Nash. La modularité passe de 0,309 à 0,325. Il est intéressant de noter que les deux individus réaffectés le sont dans la troisième communauté qui se trouve renforcée alors qu'aucun autre auteur n'avait identifié cette communauté. Pourtant c'est cette configuration qui donne la meilleure valeur de modularité.

5 Variation des communautés

Les algorithmes de détection tels que celui de Louvain donnent une solution approchée. De plus le nombre de communautés et l'affectation des noeuds sont des résultats non contrôlés. A l'inverse les méthodes de clustering comme K-means ou l'ACP permettent de contrôler le nombre de clusters, mais ne permettent pas de faire varier les paramètres autour d'un optimum. Nous montrons ici comment avec la méthode de réaffectation RM-Nash il est possible de faire varier différents paramètres et d'observer les tentatives de retour à l'équilibre. Nous soulignons l'intérêt pratique de ces transformations en particulier au plan sémantique et les champs de recherche qui s'ouvrent.

Individus

Noeud	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
1 (9)	0	0	0	0	0	0	0	0	0	-24,492	-46,964	-71,203	-61,04	-51,004	-36,548	-23,229	-23,229	-23,229
2 (9)	-46,964	-57,505	-46,964	-57,505	-44,944	-22,472	-22,472	11,425	0	0	0	0	0	0	0	0	0	0

Evènements

Avant Equilibre

Noeud	E1	E2	E3	E4	E5	E6	E7	E8	E10	E12	E13	E14	E9	E11
1 (7)	0	0	0	0	0	0	0	6,565	-59,02	-71,203	-35,033	-35,033	-57,064	-46,964
2 (7)	-33,518	-33,518	-68,173	-44,944	-91,908	-69,436	-26,259	0	0	0	0	0	0	0

Individus

Noeud	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
1 (8)	0	0	0	0	0	0	0	-11,425	0	-22,977	-45,449	-68,931	-58,389	-47,974	-34,655	-22,472	-22,472	-22,472
2 (10)	-49,994	-60,157	-49,994	-60,157	-46,459	-23,987	-23,987	0	-1,515	0	0	0	0	0	0	0	0	0

Evènements

Après Equilibre

Noeud	E1	E2	E3	E4	E5	E6	E7	E8	E10	E12	E13	E14	E9	E11
1 (7)	0	0	0	0	0	0	0	-10,605	-57,127	-68,931	-33,897	-33,897	-74,991	-45,449
2 (7)	-34,655	-34,655	-70,446	-46,459	-94,938	-49,994	-30,047	0	0	0	0	0	0	0

FIGURE 3 – Réduction de communautés, réaffectation et Equilibre de Nash pour : Women Events

5.1 Réduction du nombre de communautés

La première transformation intéressante est de réduire le nombre de communautés à partir d'une structuration intermédiaire, que ce soit le résultat non optimal obtenu par Louvain sur SW ou le résultat optimal obtenu après retraitement par RM-Nash. Ce problème a de nombreuses applications pratiques. C'est en effet celui rencontré par un organe de décision qui souhaite réduire le nombre de sous-structures. Par exemple un laboratoire de recherche comprend trois équipes et on souhaite une meilleure visibilité extérieure en passant à deux équipes plus importantes en taille (le lecteur reconnaîtra un problème familier en général difficile à gérer). L'exemple SW simule ce problème, les chercheurs étant représentés par les dames et les thèmes de recherche, ou les papiers cosignés étant représentés par les évènements.

Pour effectuer une réduction nous opérons de la manière suivante : nous affectons arbitrairement les membres de la communauté qui semble la plus instable (par exemple celle qui présente la plus forte moyenne des RM, ou bien la plus faible numériquement si le but est la visibilité) à l'une des autres communautés. Dans le cas de SW il ne reste plus que deux communautés pour lesquelles tous les RM sont ensuite calculés. Certains noeuds présentent alors des valeurs de RM positives ce qui témoigne de leur instabilité. Nous appliquons ensuite RM-Nash jusqu'à l'équilibre que nous savons exister pour obtenir une affectation sur un nombre de communautés réduit.

Expérimentalement la figure 3 montre le résultat obtenu suite au traitement suivant : on part du résultat non optimal de Louvain qui donne trois communautés, nous fusionnons les communautés 2 et 3 pour ne laisser que deux communautés. Tous les RM sont recalculés sur ces deux communautés. Les noeuds I8 et E8 apparaissent instables avec un RM positif. RM-Nash est ensuite appliqué pour obtenir un équilibre à deux communautés. On peut observer que les deux communautés trouvées sont celles qui sont observées par la majorité des sociologues dans Freeman (2003). Le résultat calculé rejoint donc le consensus. Mais nous pouvons aussi observer que la modularité avec deux communautés est plus faible qu'avec les trois communautés trouvées par Louvain, et a fortiori avec l'optimisation effectuée par RM-Nash comme montré dans

la section précédente. Ainsi nous obtenons bien un équilibre à deux communautés, mais moins bon qu'à trois communautés. En pratique, les agents ainsi réorganisés devraient tisser de nouveaux liens dans leur nouvelle structure pour améliorer l'organisation de l'ensemble. Un point intéressant à noter est que nous obtenons le même résultat expérimental à deux communautés si nous fusionnons au départ les communautés 1 et 2 ou 1 et 3. On peut se poser la question théorique de savoir si c'est toujours vrai dans le cas général. Nous conjecturons que ce n'est pas le cas car rien n'indique qu'il n'y a qu'un seul équilibre de Nash comme nous allons le voir dans la section suivante.

5.2 Essaimage de communautés et flexibilité structurelle

L'essaimage consiste à rajouter une ou plusieurs nouvelles communautés à celles déjà existantes. Dans une première étape nous partons de n communautés et nous passons à $n + 1$. Parmi les applications pratiques on peut citer le désengorgement d'une structure par essaimage, la diversification, etc. Ici aussi les algorithmes de clustering comme K-means nécessitent de recommencer le calcul à partir de zéro sans pouvoir expliciter le phénomène ni justifier un équilibre. Bien évidemment les algorithmes de détection de communautés ne peuvent opérer ce genre de révision puisqu'ils donnent une seule solution. Pour notre part nous partons de n communautés à l'équilibre (cette condition n'est pas cependant nécessaire) et forçons dans une communauté nouvellement créée un sous ensemble d'éléments provenant des autres communautés. En pratique, nous réaffectons l'élément le moins stable même s'il est à l'équilibre (RM négative mais maximale). Cet élément est bloqué dans sa nouvelle communauté et tous les autres RM sont recalculés. L'équilibre est recherché de manière itérative par l'application de notre algorithme RM-Nash. Les résultats expérimentaux sur SW sont montrés dans la figure 4. Nous sommes partis de l'équilibre précédent à deux communautés et avons construit une troisième communauté. L'élément réaffecté en premier est $I9$. Il entraîne avec lui $I7$ et $E7$. L'équilibre est atteint alors avec 3 éléments dans la nouvelle communauté. On se retrouve en final avec une structure à trois communautés très différente de la structure à trois communautés d'où on était parti initialement mais avec une modularité plus faible. Sémantiquement on voit que la nouvelle communauté contient les éléments (les événements) les plus partagés par les deux communautés initiales. Nous avons réitéré l'expérience en changeant le premier élément forcé et le résultat final varie avec des structures différentes.

Différentes expériences ont aussi été menées pour passer de trois communautés à 4 selon le principe ci-dessus. Plusieurs équilibres ont été trouvés selon le sommet que nous avons privilégié pour ensemençer la communauté rajoutée. Les modularités sont variables dont une encore meilleure que la meilleure à trois communautés. Avec ces expériences il apparaît qu'il est possible de faire varier les affectations et le nombre de communautés autour de divers équilibres. Le choix et le nombre des sommets initialement forcés interviennent dans le résultat final et permettent donc de jouer sur la sémantique ou la pragmatique de l'organisation. Notre approche se différencie donc de toutes les autres puisqu'elle permet non seulement de trouver des optima, mais aussi de générer différentes organisations à l'équilibre selon les objectifs de l'expérimentateur.

Nous avons appliqué les mêmes procédures à un autre jeu de données plus conséquent : un graphe biparti de partage de 1000 photos entre environ 300 personnes provenant d'un compte Facebook. Nous avons supprimé une communauté sur la structure trouvée par Louvain. Puis nous avons recherché l'Equilibre de Nash. Enfin comme précédemment nous avons créé une

Individus

Noeud	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
1 (7)	0	0	0	0	0	0	0	-8,774	-16,16	-30,678	-41,914	-63,628	-63,439	-52,14	-41,472	-20,704	-20,704	-20,704
2 (10)	-57,064	-55,107	-45,828	-55,107	-38,758	-27,522	-16,286	0	-8,964	0	0	0	0	0	0	0	0	0
3 (1)	-41,156	-21,525	-18,685	-21,525	-8,332	-19,568	2,904	-8,648	0	-4,292	-26,764	-40,904	-25,691	-21,841	-11,299	-13,13	-13,13	-13,13

Evènements

Noeud	E1	E2	E3	E4	E5	E6	E7	E8	E10	E12	E13	E14	E9	E11
1 (6)	0	0	0	0	0	0	0,757	-9,469	-52,708	-63,628	-31,246	-31,246	-75,622	-41,914
2 (7)	-37,306	-37,306	-75,748	-49,994	-90,771	-57,064	-20,578	0	0	0	0	0	0	0
3 (1)	-22,914	-22,914	-46,964	-30,804	-41,156	-41,156	0	-12,625	-33,771	-40,904	-19,884	-19,884	-41,409	-26,764

Avant Equilibre

Individus

Noeud	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
1 (6)	0	0	0	0	0	0	-2,904	-8,774	-14,14	-30,678	-41,914	-63,628	-63,439	-52,14	-41,472	-20,704	-20,704	-20,704
2 (10)	-59,083	-56,874	-47,847	-56,874	-39,768	-28,532	-19,189	0	-7,954	0	0	0	0	0	0	0	0	0
3 (2)	-45,196	-25,06	-22,724	-25,06	-10,352	-21,588	0	-8,648	0	-4,292	-26,764	-40,904	-25,691	-21,841	-11,299	-13,13	-13,13	-13,13

Evènements

Noeud	E1	E2	E3	E4	E5	E6	E7	E8	E10	E12	E13	E14	E9	E11
1 (6)	0	0	0	0	0	0	-16,665	-9,469	-52,708	-63,628	-31,246	-31,246	-75,622	-41,914
2 (7)	-38,063	-38,063	-77,263	-51,004	-81,555	-47,847	-29,289	0	0	0	0	0	0	0
3 (1)	-24,429	-24,429	-49,994	-32,824	-22,724	-22,724	0	-12,625	-33,771	-40,904	-19,884	-19,884	-41,409	-26,764

Après Equilibre

FIGURE 4 – Réduction de communautés, réaffectation et Equilibre de Nash puis ajout de communauté et atteinte de l'Equilibre de Nash pour : Women Events

nouvelle communauté à partir de l'élément le plus instable, et nous avons calculé le nouvel Equilibre de Nash. Le premier équilibre est trouvé après 18 itérations. le deuxième équilibre est trouvé après 10 itérations supplémentaires. La modularité sur le deuxième Equilibre de Nash est inférieure à la modularité du premier. Nous retrouvons là les mêmes constatations que ci-dessus.

5.3 Axes de recherche ouverts : sémantique et évolution

Nous décrivons ici quelques pistes exploratoires qu'offre la méthode de variation des équilibres. Nous définissons dans un graphe biparti la sémantique d'une organisation en communautés d'un ensemble d'individus par l'association des éléments de l'autre ensemble. Dans l'exemple SW, les évènements regroupés avec les dames donnent la sémantique des regroupements. Pour un ensemble de chercheurs, les thèmes de recherche donnent la sémantique des équipes auxquelles sont rattachés les chercheurs et les thèmes. La méthode de réduction ou d'extension des communautés présentée ci-dessus permet de rechercher une organisation la plus consensuelle et éventuellement d'effectuer des réorganisations dirigées par la sémantique. En effet en forçant certains éléments sémantiques à être présents dans une même communauté, l'organisation sera sémantiquement forcée. D'autres contraintes peuvent être appliquées comme le nombre minimal ou maximal d'individus dans une communautés. Si ces manipulations conduisent à des équilibres dans le cas de la réduction du nombre de communautés, ce n'est pas toujours le cas pour l'extension pour lequel un équilibre peut ne s'obtenir qu'en vidant une communauté. Les conditions d'obtention d'équilibre sont un axe de recherche prometteur, en particulier en considérant la variation sémantique des communautés : par exemple peut-on rassembler une communauté en équilibre autour d'un sous ensemble de caractérisation des individus ?

Un autre axe de recherche intéressant est celui de l'évolution d'un réseau et de son impact sur l'organisation communautaire. Il fait l'objet de travaux très poussés aujourd'hui comme par exemple dans Roth (2008) parce qu'il permet de mettre en lumière des mouvements sociaux ou

écologiques. La réaffectation peut apporter la possibilité de faire varier certains éléments structuraux du réseau (rajout/suppression de sommets, de liaisons) et d'observer l'évolution marginale de l'organisation, voire l'impact de renforcement ou d'affaiblissement de l'organisation.

6 Conclusion

La recherche de communautés dans les réseaux sociaux, et au delà dans les grands graphes a fait l'objet de nombreuses recherches, les plus récentes étant basées sur l'optimisation d'une fonction de modularité représentative d'un découpage 'optimal'. Mais le problème est NP-complet et les algorithmes les plus efficaces sont des heuristiques qui donnent des résultats approchés. En transformant le problème en celui d'un jeu non coopératif, nous avons pu montrer comment améliorer les résultats en recherchant un Equilibre de Nash par réaffectation des sommets. Dans le présent article nous avons poussé plus loin cette logique en donnant un statut d'agents aux sommets et en guidant le découpage dans plusieurs directions (réduction ou augmentation du nombre de communautés, guidage sémantique, évolutivité). Il est ainsi possible de conduire et de moduler l'organisation selon des critères sémantiques ou organisationnels tout en garantissant un optimum. De nombreuses applications sont possibles quand il est question d'analyser des graphes tant dans la sphère sociale que dans des domaines scientifiques (marketing, biologie, médical, etc.) et de faire évoluer des structures.

Références

- BARBER M. J. & CLARK J. W. (2009). Detecting network communities by propagating labels under constraints. *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, **80**(2 Pt 2), 026129.
- BLONDEL V. D., GUILLAUME J.-L., LAMBIOTTE R. & LEFEBVRE E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics : Theory and Experiment*, **2008**(10).
- CHEN W., LIU Z., SUN X. & WANG Y. (2011). Community detection in social networks through community formation games. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Three, IJCAI'11*, p. 2576–2581 : AAAI Press.
- CRAMPES M. & PLANTIÉ M. (2013). Partition et recouvrement de communautés dans les graphes bipartis, unipartis et orientés. In *IC 2013 Ingénierie des connaissances*.
- CRAMPES M., PLANTIÉ M. & LOPEZ M. (2013). Optimisation dans la détection de communautés recouvrantes et équilibre de Nash. In *MARAMI conférence sur les Modèles et l'Analyse des Réseaux : Approches Mathématiques et Informatique*, Saint-Etienne.
- 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.
- MURATA T. (2010). Detecting communities from tripartite networks. *WWW*, p. 0–1.
- NEWMAN M. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, **74**(3 Pt 2), 036104.
- NISAN N., ROUGHGARDEN T., TARDOS E. & VAZIRANI V. V. (2007). *[BOOK] Algorithmic game theory*. Cambridge University Press.
- PLANTIÉ M. & CRAMPES M. (2013). *Survey on Social Community Detection*. Springer.
- R NARAYANAM & Y NARAHARI (2012). A game theory inspired decentralized local information based algorithm for community detection in social graphs . In *ICPR 21st International Conference on Pattern Recognition*, Vienna.
- ROTH C. (2008). Coévolution des auteurs et des concepts dans les réseaux épistémiques : le cas de la communauté Â« zebrafish Â». *Revue Française de Sociologie*, **49**(3), 523–553.
- SUZUKI K. & WAKITA K. (2009). Extracting Multi-facet Community Structure from Bipartite Networks. *2009 International Conference on Computational Science and Engineering*, **4**, 312–319.