

# Une approche centrée graine pour la détection de communautés dans les réseaux multiplexes

Issam Falih \*, Manel Hmimida \*&\*\*, Rushed Kanawati \*

\* LIPN - UMR CNRS 7030,  
SPC, Université Paris 13  
99 Av. J-B. Clément, 93430 Villetaneuse  
prénom.nom@lipn.univ-paris13.fr  
<http://www-lipn.univ-paris13.fr/A3>  
\*\* DICEN-CNAM,  
292 rue Saint Martin, 75003 Paris  
<http://dicen-idf.org>

**Résumé.** Nous nous intéressons dans ce travail au problème de détection de communautés dans les réseaux multiplexes. Le modèle de réseau multiplexe a été récemment introduit afin de faciliter la modélisation des réseaux multi-relationnels, des réseaux dynamiques et/ou des réseaux attribués. Les approches existantes pour la détection de communautés dans ce genre de graphes sont, pour la plupart, basées sur des schémas d'agrégation de couches ou d'agrégation de partitions. Nous proposons ici une nouvelle approche centrée graine qui permet de prendre en compte directement la nature multi-couche d'un réseau multiplexe. Des expérimentations effectuées sur différents réseaux multiplexes montrent que notre approche surpasse les approches de l'état de l'art en termes de qualité des communautés identifiées.

## 1 Introduction

La détection de communautés est une tâche centrale dans le domaine d'analyse des réseaux complexes. La littérature scientifique, concernant cette tâche, est très abondante Fortunato et Castellano (2009), Tang et Liu (2010). La majeure partie des travaux existants s'intéresse au cas de réseaux simples et statiques. Or, dans beaucoup de cas réels les interactions entre acteurs sont mieux modélisées par des réseaux multiplexes. Un multiplexe est un graphe multi-relationnel, il est souvent représenté par un graphe multi-couches. Chaque couche contient le même ensemble de nœuds mais encode une relation différente. Un exemple, est le réseau de transports aériens en Europe Cardillo et al. (2012) : les nœuds représentent des aéroports et chaque couche représente les vols réguliers opérés par une compagnie aérienne (ex. Airfrance, Lufthansa, ... , etc). Nous nous intéressons, dans ce travail, à la problématique de détection de communautés dans des réseaux multiplexes. Nous présentons une nouvelle approche qui généralise l'algorithme Licod Yakoubi et Kanawati (2014) au cas d'un multiplexe. La suite de ce papier est organisée comme suit. Dans la section 2, nous introduisons les notations utilisées ainsi un bref état de l'art. Nous proposons dans la section 3, une approche centrée graine

pour traiter ce problème. Une étude comparative de l'approche proposée avec des approches de l'état de l'art est présentée dans la section 4. Les conclusions de cette étude sont données dans la section 5.

## 2 Détection de communautés dans des graphes multiplexes

Un réseau multiplexe, composé de  $\alpha$  couches, est désigné par :  $G = \langle V, E_1, E_2, \dots, E_\alpha : E_k \subseteq V \times V \forall k \in \{1, \dots, \alpha\} \rangle$ , où  $V$  est l'ensemble des nœuds et  $E_k$  est l'ensemble des liens défini dans la couche  $i$ . Le tableau 1 donne les principales notations que nous utilisons dans la suite de ce travail.

Notation	Description
$A^{[k]}$	Matrice d'adjacence de la couche $k$
$d_i^{[k]}$	Degré du nœud $i$ dans la couche $k$
$d_i^{tot} = \sum_{k=1}^{\alpha} d_i^{[k]}$	Degré total du nœud $i$
$m^{[k]}$	Nombre de liens dans la couche $k$
$\Gamma(v)^{[k]} = \{u \in V : (u, v) \in E_k\}$	Voisinage de $v$ dans la couche $k$
$\Gamma(v)^{tot} = \cup_{s \in \{1, \dots, \alpha\}} \Gamma(v)^{[s]}$	Voisinage de $v$ dans toutes les couches $\alpha$
$SPath^{[k]}(u, v)$	Longueur du plus court chemin entre $u$ et $v$ dans la couche $k$

TAB. 1: Notations utilisées pour les réseaux multiplexes

Nous classifions les approches de détection de communautés dans les réseaux multiplexes en deux grandes familles. La première famille, consiste à ramener le problème de détection de communautés dans un graphe multiplexe en un graphe simple (monoplexe). La deuxième famille consiste à généraliser une méthode de détection de communautés d'un graphe simple pour qu'elle traite simultanément les différentes couches du graphe multiplexe. Deux voies sont empruntées dans la littérature pour la mise en œuvre des approches de premier type : l'agrégation des couches ou l'agrégation des partitions.

**Agrégation des couches :** le principe consiste à transformer le graphe multiplexe en un graphe simple en appliquant une fonction d'agrégation entre les différentes couches puis appliquer un algorithme de détection de communautés sur le réseau résultant. Berlingerio et al. (2011); Suthers et al. (2013)

**Agrégation des partitions :** l'agrégation des partitions consiste à appliquer un algorithme simple de détection de communautés à chaque couche du multiplexe puis utiliser une approche d'ensemble clustering pour regrouper les différentes partitions retrouvées. Seifi et Guillaume (2012); Strehl et Ghosh (2003)

Très peu d'approches ont abordé le problème de détection de communautés dans le multiplexe en explorant simultanément les différentes couches. Dans Tang et Liu (2010), on propose un modèle unifié selon lequel on identifie de nouvelles approches d'agrégations autres que les deux approches triviales d'agrégation des couches et d'agrégation des partitions. Dans un autre travail décrit dans Mucha et al. (2010), on généralise la modularité définie par Girvan

et Newman (2002). Cette nouvelle modularité permet d'utiliser les algorithmes de détection de communautés, fondés sur l'optimisation de la modularité Blondel et al. (2008) au cas des réseaux multiplexes.

### 3 L'approche muxLicod

Nous proposons une nouvelle approche pour la détection de communautés dans les réseaux multiplexes. Notre approche est une généralisation directe de l'algorithme Licod Yakoubi et Kanawati (2014), proposé initialement pour traiter des graphes monoplexes. Le principe est que les communautés se forment autour des nœuds appelées *Leaders*. On identifie en premier temps les *Leaders* puis on calcule les communautés locales égo-centrées sur les nœuds *Leaders*. L'algorithme se fait en quatre étapes :

1. *Identification des leaders* : les nœuds ayant une centralité supérieure à celle de ses voisins sont identifiés comme des leaders. Différentes mesures de centralité peuvent être appliquées. Dans notre approche, nous utilisons la centralité de degré (en utilisant la définition de degré multiplexe). Soit  $\mathcal{L}$  l'ensemble de *Leaders* identifiés dont chaque élément forme une communauté.
2. *Calcul du vecteur de préférences* : chaque nœud  $v \in V$  établit un vecteur de préférences à appartenir à une communauté parmi les communautés identifiées dans  $\mathcal{L}$ . Le degré d'appartenance est estimé par la longueur de *plus court chemin multiplexe* reliant le nœud à chaque *Leader*.
3. *Fusion des vecteurs de préférences* : chaque nœud échange avec ses voisins (selon la définition fournie ci-après) son vecteur de préférences. Les vecteurs de préférences seront fusionnés par chaque nœud. Cette opération est itérée jusqu'à la convergence.
4. *Attribution des communautés* : chaque nœud sera assigné à la communauté placée en tête de son vecteur de préférences après stabilisation.

L'ensemble des métriques classiques dans les réseaux simples sont à redéfinir pour le cas de réseaux multiplexes. Nous redéfinissons ci-après trois métriques classiques dont nous avons besoin pour la mise en œuvre de notre approche :

**Degré d'un nœud** : nous adoptons la définition proposée dans Battiston et al. (2013) qui propose de calculer le degré par une fonction d'entropie :

$$d_i^{multiplexe} = - \sum_{k=1}^{\alpha} \frac{d_i^{[k]}}{d_i^{\sigma t}} \log \left( \frac{d_i^{[k]}}{d_i^{\sigma t}} \right)$$

Le degré d'un nœud  $i$  est nul si tous ses voisins directs sont connectés dans une seule couche. Par contre, le degré du  $i$  sera maximal si ses voisins lui sont connectés dans toutes les couches du multiplexe.

**Plus court chemin** : la taille du plus court chemin entre deux nœuds est simplement donnée par la moyenne des plus courts chemins dans chacune des couches. Si dans une couche donnée, les deux nœuds sont dans deux composantes connexes différentes, la couche n'est pas prise en compte.  $SPath(u, v) = \sum_{i=1}^{\alpha} \frac{path_i(v, u)}{\alpha}$

**Voisinage d'un nœud :** différentes approches pour calculer le voisinage d'un nœud sont proposées dans la littérature Brodka et al. (2011). Nous proposons ici une définition simple qui définit les voisins multiplexe d'un nœud comme l'union de tous ses voisins dans les différentes couches du multiplexe.  $\Gamma(v) = \cup_{s \in \{1, \dots, \alpha\}} \Gamma(v)^{[s]}$

## 4 Résultats & discussions

Nous avons utilisé, pour nos expérimentations, les réseaux multiplexes suivantes :

**Réseau des transports aériens en Europe Nicosia et Latora (2014) :** Comporte 449 nœuds et il est composé de 37 couches dont chacune représente les vols réguliers entre les aéroports Européen.

**Réseau d'innovations des médecins Coleman et al. (1957) :** Ce réseau comporte trois couches/rerelations (conseil, discssion, amis) entre médecins (246). Le but est de savoir l'impact des relations entre les médecins sur l'adoption d'un nouveau médicament.

**Réseau des cabinets d'avocats Lazega (2001) :** Comporte trois types de relations (travaille-avec, amis, conseil) entre les partenaires et les associés des cabinets d'avocats (71).

**Réseau de Vickers :** Les données ont été collectées par Vickers auprès de 29 élèves, qui étaient amenés à désigner leurs camarades de classe sur un certain nombre de relations (meilleurs-amis, travaille-avec, sympathiser).

Très peu de travaux ont abordé le problème d'évaluation pour le cas des graphes multiplexes. A notre connaissance, nous n'avons pas de graphes réels ni de générateurs de graphes artificiels qui peuvent nous donner des réseaux multiplexes de benchmark. Au niveau des indicateurs topologiques, on trouve la modularité multiplexe et le critère de redondance.

**Modularité multiplexe Mucha et al. (2010) :** est utilisé pour mesurer la proportion de liens internes aux communautés et la même quantité dans un modèle nul où aucune structure communautaire n'est attendue. Le modèle nul est un graphe aléatoire ayant le même nombre de nœuds, de couches et de liens et la même distribution de degré.

**Mesure de redondance Berlingerio et al. (2011) :** mesure la qualité d'une partition. Elle calcule la moyenne de redondance de chaque lien intra-communauté dans l'ensemble des couches du multiplexe. L'intuition est que les liens intra-communauté doivent être des liens redondants dans les différentes couches.

Nous avons comparé notre approche avec les deux approches classiques d'agrégation de couches et d'agrégation des partitions en utilisant différents algorithmes de base reconnus dans l'état de l'art à savoir : *Edge Betweenness* Girvan et Newman (2002), *Walktrap* Pons et Latapy (2006), *Louvain* Blondel et al. (2008) et *Infomap* Rosvall et al. (2009). La figure 1 présente les performances des approches comparées sur les quatre réseaux, en termes de redondance et de modularité. Le préfixe LA (resp. PA) désigne l'application d'une approche d'agrégation de couches (resp. agrégation des partitions). Notre approche a surpassé les approches de bases en terme de qualité de partitionnement. En effet, les liens intra-communauté sont plus redondants dans les différentes couches du multiplexe avec notre approche que ceux retrouvées avec les deux approches de bases. En terme de la modularité, *muxLicod* est un peu moins compétitive que l'algorithme Louvain Blondel et al. (2008), basé sur l'optimisation de la modularité. En prenant les deux critères d'évaluation en compte, notre approche se situe clairement dans le

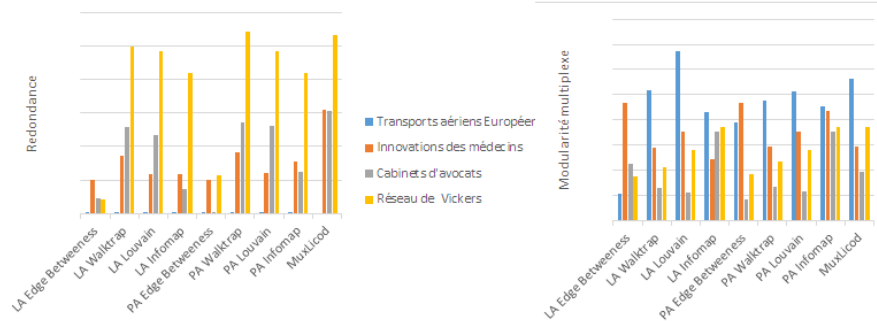


FIG. 1: Mesure de redondance(à droite), mesure de modularité (à gauche)

front de Pareto. Une étude de paramétrage de l'approche (choix de fonction de centralité pour la sélection de leaders, choix de l'approche de fusion des vecteurs de préférences, etc.) est en cours.

## 5 Conclusion

Nous avons proposé une nouvelle approche centrée graine pour la détection de communauté dans les réseaux multiplexes. Cette approche, qui explore simultanément les couches d'un multiplexe présente des performances intéressantes comparées aux approches classiques basées sur l'agrégation de couches ou l'agrégation des partitions. L'approche, étant une approche centrée graine, est basée sur des calculs locales ce qui permet d'envisager son application sur des grands graphes. Cette étude doit être étendue pour inclure une comparaison avec d'autres approches, notamment celles basées sur l'optimisation de la modularité multiplexe Mucha et al. (2010). Une réflexion plus approfondie doit aussi être menée pour le choix des critères d'évaluation des communautés dans les réseaux multiplexes. Ce problème d'évaluation de communautés reste un problème ouvert, même pour les réseaux monoplexes. Des approches orientées tâches, à l'instar du travail présenté dans Yakoubi et Kanawati (2014) sont aussi à l'étude.

## Références

- Battiston, F., V. Nicosia, et V. Latora (2013). Metrics for the analysis of multiplex networks. *CoRR abs/1308.3182*.
- Berlingerio, M., M. Coscia, et F. Giannotti (2011). Finding and characterizing communities in multidimensional networks. In *International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2011, Kaohsiung, Taiwan, 25-27 July 2011*, pp. 490–494.
- Blondel, V. D., J.-I. Guillaume, et E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics : Theory and Experiment 2008*, P10008.

- Brodka, P., K. Skibicki, P. Kazienko, et K. Musial (2011). A degree centrality in multi-layered social network. In *CASoN*, pp. 237–242.
- Cardillo, A., J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. del Pozo, et S. Boccaletti (2012). Emergence of network features from multiplexity. *CoRR abs/1212.2153*.
- Coleman, J., E. Katz, et H. Menzel (1957). The diffusion of an innovation among physicians. *Sociometry* 20(4), 253–270.
- Fortunato, S. et C. Castellano (2009). Community structure in graphs. pp. 1141–1163.
- Girvan, M. et M. E. J. Newman (2002). Community structure in social and biological networks. *PNAS* 99(12), 7821–7826.
- Lazega, E. (2001). *The collegial phenomenon : the social mechanisms of cooperation among peers in a corporate law partnership*. Oxford : Oxford university press.
- Mucha, P. J., T. Richardson, K. Macon, M. A. Porter, et J.-P. Onnela (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science* (5980), 876–878.
- Nicosia, V. et V. Latora (2014). Measuring and modelling correlations in multiplex networks. *CoRR abs/1403.1546*.
- Pons, P. et M. Latapy (2006). Computing communities in large networks using random walks. *J. Graph Algorithms Appl.* 10(2), 191–218.
- Rosvall, M., D. Axelsson, et C. T. Bergstrom (2009). The map equation. *Eur. Phys. J. Special Topics* 13, 178.
- Seifi, M. et J.-L. Guillaume (2012). Community cores in evolving networks. In A. Mille, F. L. Gandon, J. Misselis, M. Rabinovich, et S. Staab (Eds.), *WWW (Companion Volume)*, pp. 1173–1180. ACM.
- Strehl, A. et J. Ghosh (2003). Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research* 3, 583–617.
- Suthers, D. D., J. Fusco, P. K. Schank, K.-H. Chu, et M. S. Schlager (2013). Discovery of community structures in a heterogeneous professional online network. In *HICSS*, pp. 3262–3271. IEEE.
- Tang, L. et H. Liu (2010). *Community Detection and Mining in Social Media*. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool Publishers.
- Yakoubi, Z. et R. Kanawati (2014). Licod : Leader-driven approaches for community detection. *Vietnam Journal of Computer Science - Springer* 1, 241–256.

## Summary

The multiplex network concept has recently been introduced to ease the modeling of multi-relational, dynamic or attributed networks. Approaches for community detection in such networks are mostly based on layers aggregation or partitions aggregation schemas. In this paper, we propose a new approach that consists on adapting a seed-centric algorithm to the multiplex case. First experiments on different real multiplexes networks show the relevance of the approach compared to state-of-the art approaches.