



HAL
open science

Voisinage et intrication dans les réseaux multiplexes

Benjamin Renoust

► **To cite this version:**

Benjamin Renoust. Voisinage et intrication dans les réseaux multiplexes. Modèles et Analyses Réseau : Approches Mathématiques et Informatiques (MARAMI), 2014, Paris, France. 10.3166/RMPD.0-7 . hal-01454971

HAL Id: hal-01454971

<https://hal.science/hal-01454971>

Submitted on 15 Feb 2017

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.

Voisinage et intrication dans les réseaux multiplexes

Benjamin Renoust¹

Institut National de l'Audiovisuel, Paris, France

JFLI CNRS UMI 3527 & National Institute of Informatics, Tokyo, Japan

LaBRI CNRS UMR 5800, Bordeaux, France

benjamin.renoust@labri.fr

ABSTRACT. L'intérêt du modèle de réseau complexe pour capturer la réalité des interactions sur les données de terrain n'est plus à démontrer. De ces réseaux, nous dérivons souvent des modèles de réseaux bipartis, par association de nœuds et de communautés, ou de manière plus naturelle lorsque plusieurs familles d'entités se croisent. Cependant, certaines données de terrains s'adaptent mal au modèle biparti, qui ne traduit que le cas de l'homophilie entre nœuds d'une même partition. C'est par exemple le cas du réseau de co-auteurs académiques dont on aura enrichi les interactions par des mots-clefs. Dans ce cas de figure, une autre approche, pour le moment peu exploitée, permet de capturer plus finement les interactions entre entités: ce sont les réseaux multiplexes, qui présentent plusieurs couches d'interactions entre les mêmes entités. Les réseaux multiplexes concernent l'union de plusieurs familles d'interactions entre entités, ne se limitant pas au modèle biparti. Ce papier ainsi présente une étude préliminaire qui questionne la notion de voisinage dans les réseaux multiplexes et étend l'approche globale de l'intrication d'arêtes (Renoust et al., 2013), avec deux mesures permettant la caractérisation des nœuds dans leur réseau.

KEYWORDS: réseaux multiplexes, analyse d'intrication, voisinage

1. Introduction

L'analyse et l'exploration d'un réseau social dépend du type de relations en jeu. (Borgatti *et al.*, 2009) ont proposé une taxonomie de ces relations en quatre différentes catégories: *l'homophilie* (ou similarité), *la relation sociale*, *l'interaction directe*, et *le flux*. Lorsque l'on observe un réseau social, nous capturons souvent la superposition de plusieurs de ces types de relations, et notre étude se place sur un seul point de vue en particulier.

Si l'on prend l'exemple du réseau de co-auteurs de publications scientifiques caractérisées par des mots-clefs, il est intéressant d'étudier la similarité entre co-auteurs

au travers des mots-clefs de leurs publications. Nous pouvons donc associer les auteurs à leur mots-clefs, et capturer cette association au travers d'un réseau biparti *auteurs-mots-clefs*. Avec ce modèle, il y a souvent un jeu de nœuds qui est considéré comme *primaire* et responsable de la formation de liens (Opsahl, 2013), c'est-à-dire que lorsque l'on analyse le projeté du biparti sur un réseau à un seule mode, l'autre jeu de nœuds (lui *secondaire*) pourra être considéré comme caractérisant les liens dans ce réseau. Si on analyse le projeté de ce réseau sur les auteurs (*primaires*), il rapproche inmanquablement des auteurs ayant publié sur les mêmes thématiques (*secondaires*), sans qu'ils aient pour autant co-publié. Le modèle de réseau multiplexe permet d'éviter cette erreur en démultipliant la relation de co-publication grâce aux mots-clefs de ces mêmes co-publications, et chaque mot clef représentera une famille de liens dans le réseau. Ainsi si le réseau de co-auteurs se définit par $G = (V, E)$, avec V les auteurs, et E la relation de co-publication, le réseau multiplexe de co-auteurs est définit par: $G = (V, \cup_{t=1}^n E_t)$, avec toujours V les auteurs, mais E_t la relation de co-publication avec le mot-clef t .

Peu d'approches ont pour le moment été proposées pour étudier ce type d'objets (Bródka, Kazienko, 2012), tout particulièrement concernant la position d'un nœud dans son réseau multiplexe. Il convient malgré tout de rapporter d'intéressants travaux qui se placent dans la même philosophie, telle que la prédiction de lien dans les réseaux multiplexes (Pujari, Kanawati, 2013). Aussi, beaucoup d'approches considèrent une version "agrégée" du réseau multiplexe, comme un graphe pour lequel on ne différencie aucune couche d'arête et dont on étudie la topologie (Zhou *et al.*, 2007 ; Borgatti, 2012), et il a été montré que ces graphes ont alors le défaut de contenir beaucoup de cliques (Guillaume, Latapy, 2005). Ainsi certaines méthodes se sont intéressées au filtrage d'arêtes. C'est le cas pour (Nick *et al.*, 2013) qui recherchent la colonne vertébrale *simélienne* dans la projection d'un réseau à deux modes, mais aussi pour (Neal, 2013) qui lui filtre les arêtes en fonction de leurs propriétés statistiques. Les travaux de (De Domenico *et al.*, 2013) cherchent à apporter une formulation mathématique sur la base des tenseurs qui permet d'englober la définition de chaque famille de réseau comprenant plusieurs couches, sans pour le moment appuyer l'analyse de ces réseaux. La combinaison topologie et attributs a été aussi au centre de récents travaux comme (Rossi, Villa-Vialaneix, 2012 ; Cruz *et al.*, 2011), avec un point de vue souvent exclusivement centré sur les nœuds. Cependant, quelques mesures concernent particulièrement les réseaux multiplexes, bien que parfois se référant à un graphe biparti. On notera par exemple la *multiplicité* (Podolny, Baron, 1997) qui fait référence au nombre d'arêtes parallèles. Se posant des questions similaires à celles de notre approche, la *redondance* (Latapy *et al.*, 2008) est une mesure du recouvrement de voisinage d'un nœud dans un graphe biparti. L'*exposition au réseau* (Fujimoto *et al.*, 2011) d'un évènement mesure l'influence son influence dans le projeté du modèle acteur-évènement sur un modèle acteur-acteur. L'analyse d'intrication (Renoust *et al.*, 2013) mesure l'entremêlement des couches d'arêtes dans un réseau multiplexe, couche par couche, ainsi que de manière globale au réseau. C'est sur la base de cette dernière analyse que nous proposons d'exploiter un indice pour les nœuds d'un réseau multiplexe.

L'établissement de mesures caractérisant la place d'un nœud dans un réseau multiplexe est primordiale. Elles permettent d'identifier des individus saillants de par leur position dans le réseau multiplexe, mais surtout elles peuvent être un appui certain pour la réalisation d'algorithmes de partitionnement propres aux réseaux multiplexes (Kanawati, 2013). Ainsi de récents travaux tels que ceux de (Battiston *et al.*, 2013) présentent une première étude sur un lot de mesures sur la relation entre les nœuds et les couches auxquelles ils contribuent.

Après un bref rappel du concept d'intrication d'arêtes dans un réseau multiplexe, nous introduirons le questionnement de la position d'un nœud dans son voisinage, et présenterons l'intrication d'un nœud que nous appliquerons sur des données tirées des documents de l'INA. Enfin, nous discuterons ces résultats avant de conclure.

2. L'intrication dans les réseaux multiplexes

Nous avons présenté le concept d'intrication dans (Renoust *et al.*, 2013). Celui-ci se place dans le cas d'un réseau multiplexe où chaque couche représente une famille d'arêtes: $G = (V, \cup_{t=1}^n E_t)$. Le concept d'intrication cherche à mesurer combien et comment les familles d'arêtes s'entremêlent dans un réseau multiplexe. Grâce à l'analyse des co-occurrences de familles d'arêtes, nous pouvons réduire le problème de l'analyse d'un réseau multiplexe (gourmand par la combinatoire que celui-ci introduit) à l'analyse coordonnée de deux réseaux séparés: le réseau d'interaction des nœuds (c'est-à-dire la version "aplatie" du réseau multiplexe) et le réseau d'interaction des familles d'arêtes. Dans ce second réseau, une famille représente un nœud, et deux nœuds sont liés lorsqu'au moins deux familles d'arêtes co-occurrent entre eux. L'analyse de ces objets va au delà des approches classiques considérant des poids sur des arêtes multiples comme le montre la figure 1.

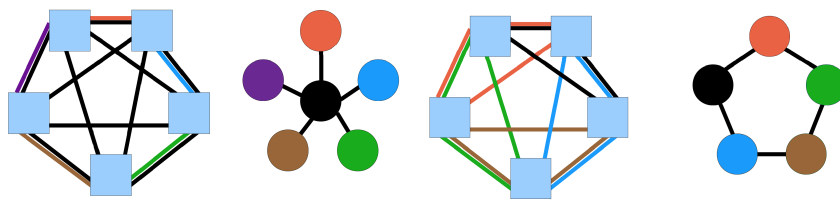


FIGURE 1. Ces exemples soulignent les nuances que prend en compte l'analyse d'intrication dans les réseaux multiplexes. Les deux réseaux multiplexes (à gauche) se projettent sous la même forme de clique-5, et si l'on assigne des poids aux arêtes correspondant au nombre de familles d'arêtes (en couleur) les liant, les deux réseaux serait exactement les mêmes. C'est la forme de l'interaction de ces familles d'arêtes (à droite) qui permet de différencier ces deux cas.

L'analyse d'intrication nous permet ainsi de dégager un indice d'intrication g_s qui permet de mesurer l'influence d'une couche d'arête s dans l'entremêlement avec les autres couches. Le calcul du vecteur d'indices d'intrication ($g_{s'} \cdot \lambda = \sum_s c_{s,s'} \gamma_s$) nécessite l'analyse algébrique d'une matrice d'intrication C , à partir de laquelle, nous pouvons évaluer deux autres mesures propres au réseau dans son ensemble: l'intensité d'intrication $\mathcal{I}(G) = \frac{\lambda_C}{|C|}$, et l'homogénéité d'intrication $\mathcal{H}(G) = \frac{\langle 1_C, \gamma \rangle}{1_C \cdot \gamma}$. La matrice C et les différentes mesures sont introduites et formalisées dans (Renoust *et al.*, 2013).

L'intensité d'intrication mesure la quantité d'intrication relative à un maximum théorique, c'est-à-dire relative au cas où toutes les arêtes du réseau sont saturées par toutes les différentes familles présentes dans le réseau. L'homogénéité d'intrication mesure la répartition de cette intrication, c'est-à-dire relative à un cas optimal où chaque famille d'arête apporte autant à l'intrication globale du groupe.

Il est à noter que cette analyse ne concerne pas la topologie du réseau multiplexe "aggrégé" en lui-même, mais se concentre sur la co-occurrence des familles d'arêtes. C'est pour cela que nous avons combiné les deux réseaux (celui "aggrégé" et celui des familles) afin de permettre la compréhension de la structure du réseau multiplexe. Enfin, à travers divers exemples, nous avons démontré le lien entre les mesures d'intensité et d'homogénéité d'intrication et la notion interprétative de *cohésion* d'un ensemble d'entités dans un réseau.

C'est inspiré par cette notion de cohésion que nous voulons questionner la place d'un nœud dans son voisinage, en nous basant sur les mesure d'intrication.

3. Le voisinage d'un nœud dans un réseau multiplexe

Nous devons considérer un certain nombre de faits avant de se lancer dans une quelconque mesure. Tout d'abord, est-ce que ce sous-graphe ego-centré nécessite l'inclusion d'arêtes entre les voisins de u ? Autrement dit, devons-nous considérer le sous-graphe induit (ici noté $N_G(u)$)? ou bien devons-nous considérer seulement les arêtes adjacentes à u (ici noté $E_G(u)$)? Cette différence, illustrée par la figure 2 dépend de la question que nous nous posons. Si l'on veut questionner la légitimité de l'appartenance d'un nœud à son voisinage, $N_G(u)$ semble plus pertinent. Si on s'intéresse à la cohésion des relations d'un nœud, c'est $E_G(u)$ qui se révèle plus pertinent.

Une autre différence subtile peut aussi diriger notre choix. En effet, si le projeté "aggrégé" de sa composante connexe que nous étudions forme une clique, alors chacun des nœuds u de la clique présentera exactement le même voisinage $N_G(u)$, ce qui ne sera pas nécessairement le cas $E_G(u)$. Nous avons ainsi une méthode pour discriminer les différents nœuds d'une même clique.

4. L'intrication d'un nœud dans un réseau multiplexe

Lorsque nous nous intéressons à l'intrication d'arête au niveau d'un nœud, la question est quelle est la cohésion de ce nœud dans son voisinage? De manière directe,

nous pouvons appliquer l’analyse de l’intrication du nœud dans un sous-graphe “ego-centré” à chaque nœud u se composant du nœud et de son voisinage direct. Ainsi nous avons deux mesures, l’intensité d’intrication du nœud $\mathcal{I}_{N_G(u)}$ ou $\mathcal{I}_{E_G(u)}$, et l’homogénéité d’intrication du nœud $\mathcal{H}_{N_G(u)}$ ou $\mathcal{H}_{E_G(u)}$ dans son voisinage.

Nous avons aussi étudié la corrélation de nos mesures d’intrication avec d’autres mesures connues sur un jeu de données comprenant 10 000 sujets de JT de l’INA. Ces documents sont annotés avec un codage ethnographique (des mots-clefs) et près de 300 groupes ont été préalablement formés. Pour chacun des groupes, nous avons créé un réseau de documents où un lien est formé par proximité sémantique (plus de détails dans (Renoust, 2014)) et multiplexé ces liens avec les mots-clefs partagés par les deux documents proches. Pour chacun de ces documents, nous avons mesuré

Tableau 1. Comparaison du coefficient de corrélation de Pearson entre l’intensité et l’homogénéité d’intrication avec neuf autres mesures en utilisant le sous-graph “ego-centré” des arêtes adjacentes.

	<i>Ent. intensity</i>	<i>Ent. homogeneity</i>
Betweenness centrality	−0.130	−0.104
Closeness centrality	−0.083	−0.030
Clustering coefficient	0.403	0.295
Degree	−0.108	−0.060
Eccentricity	0.032	0.010
K-cores	0.152	0.104
PageRank	−0.266	−0.170
Strength	0.321	0.251
# Familles de liens	−0.117	−0.078
Multiplexity	−0.095	−0.039

$\mathcal{I}_{N_G(u)}$, $\mathcal{H}_{N_G(u)}$, ainsi que neuf autres mesures connues, et renseigné le tableau 1 (nous utilisons le coefficient de corrélation de (Pearson, 1901), le détail se trouve en annexe A de (Renoust, 2014)).

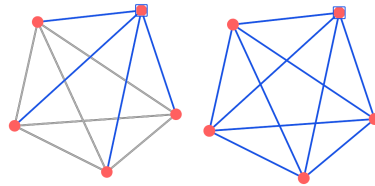
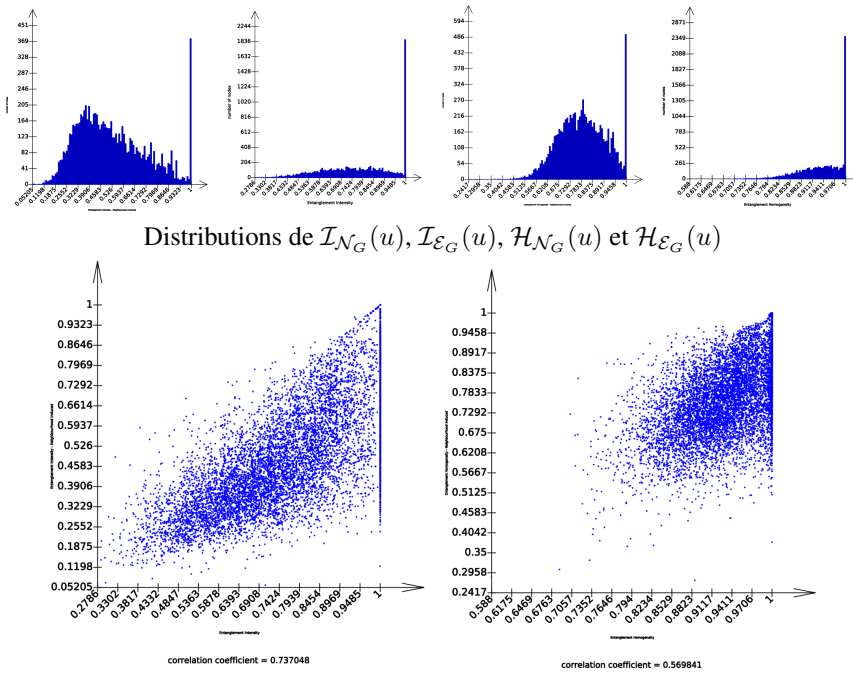


FIGURE 2. Les deux manières de concevoir le voisinage(en bleu) dans un réseau multiplexe: les arêtes adjacentes à un nœud $E_G(u)$ (gauche), et le sous-graphe induit à au voisinage d’un nœud $N_G(u)$ (droite).

5. Discussion et conclusion

La figure 3 suggère qu’il y a de sérieuses différences entre les différents types de voisinage dans les réseaux multiplexes. C’est sans surprise que l’on constate beaucoup plus de cas d’intensité et d’homogénéité d’intrication maximales pour les réseaux multiplexes ego-centrés incluant uniquement les arêtes adjacentes. Le principal avantage que l’on retiendra dans ce cas de figure est la discrimination des différents individus dans le cas où la topologie “agrégée” du réseau forme une clique.

L’intensité et l’homogénéité d’intrication d’un nœud ne semblent pas corrélées (ni anti-corrélées) avec aucune autre mesure classique de réseau. Bien sûr ce papier court rapporte des travaux préliminaires et de plus amples études sont nécessaires avant d’établir toute conclusion. A l’exception de la *multiplicité* et du nombre de familles de liens, ces autres mesures s’appliquent sur la version “agrégée” du réseau multiplexe. Pour attester pleinement l’originalité des mesures d’intrication, il faudra aussi les comparer aux mesures présentées dans (Battiston *et al.*, 2013 ; Pujari, Kanawati,



Comparaison des deux distributions $\mathcal{I}_{\mathcal{E}_G}(u) \times \mathcal{I}_{\mathcal{N}_G}(u)$, et $\mathcal{H}_{\mathcal{E}_G}(u) \times \mathcal{H}_{\mathcal{N}_G}(u)$

FIGURE 3. Distribution des mesures d’intrication parmi 10 000 sujets de JT. La comparaison des distributions de \mathcal{I} et \mathcal{H} entre voisinages montre des différences significatives, \mathcal{E}_G présentant plus de valeurs optimales.

2013 ; Cozzo *et al.*, 2013 ; Berlingiero *et al.*, 2013), à la *redondance*, à l'*exposition au réseau* et aux autres mesures présentées en section 1. Cette correspondance devra aussi être établie sur des modèles de graphes multiplexes aléatoires qui n'ont pour le moment pas été produits. Ces résultats préliminaires semblent déjà encourager une solide piste dans l'originalité de cette approche, ouvrant de nouvelles perspectives pour l'analyse des réseaux multiplexes, et notamment dans la recherche de communautés pour les réseaux multiplexes.

Bibliographie

- Battiston F., Nicosia V., Latora V. (2013). Metrics for the analysis of multiplex networks. *arXiv preprint arXiv:1308.3182*.
- Berlingiero M., Pinelli F., Calabrese F. (2013). Abacus: frequent pattern mining-based community discovery in multidimensional networks. *Data Mining and Knowledge Discovery*, vol. 27, n° 3, p. 294–320.
- Borgatti S. P. (2012). Two-mode concepts in social network analysis. In R. A. Meyers (Ed.), *Computational complexity - theory, techniques, and applications*, p. 2912–2924. Springer.
- Borgatti S. P., Mehra A., Brass D. J., Labianca G. (2009). Network analysis in the social sciences. *Science*, vol. 323, n° 5916, p. 892–895.
- Bródka P., Kazienko P. (2012). Multi-layered social networks. *arXiv preprint arXiv:1212.2425*.
- Cozzo E., Kivelä M., De Domenico M., Solé A., Arenas A., Gómez S. *et al.* (2013). *Clustering coefficients in multiplex networks*. *arXiv preprint arXiv:1307.6780*.
- Cruz J. D., Bothorel C., Poulet F. (2011). *Entropy based community detection in augmented social networks*. In Computational aspects of social networks (cason), 2011 international conference on, p. 163–168.
- De Domenico M., Solè-Ribalta A., Cozzo E., Kivelä M., Moreno Y., Porter M. A. *et al.* (2013). Mathematical formulation of multi-layer networks. *arXiv preprint arXiv:1307.4977 physics.soc-ph*.
- Fujimoto K., Chou C.-P., Valente T. W. (2011). The network autocorrelation model using two-mode data: Affiliation exposure and potential bias in the autocorrelation parameter. *Social networks*, vol. 33, n° 3, p. 231–243.
- Guillaume J.-L., Latapy M. (2005). Bipartite graphs as models of complex networks. In *Combinatorial and algorithmic aspects of networking*, vol. 3405, p. 127–139. Springer.
- Kanawati R. (2013). Détection de communautés dans les grands graphes d'interactions (multiplexes): état de l'art. *HAL-00881668*.
- Latapy M., Magnien C., Vecchio N. D. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks*, vol. 30, n° 1, p. 31–48.
- Neal Z. (2013). Identifying statistically significant edges in one-mode projections. *Social Network Analysis and Mining*, p. 1–10.
- Nick B., Lee C., Cunningham P., Brandes U. (2013). Simmelian backbones: Amplifying hidden homophily in facebook networks. In *2013 international conference on advances in social networks analysis and mining (asonam)*, p. 525–532.

- Opsahl T. (2013). Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, vol. 35, n° 2, p. 159–167.
- Pearson K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, n° 11, p. 559–572.
- Podolny J. M., Baron J. N. (1997). Resources and relationships: Social networks and mobility in the workplace. *American sociological review*, vol. 62, n° 5, p. 673–693.
- Pujari M., Kanawati R. (2013). Link prediction in multiplex bibliographical networks. *International Journal of Complex Systems in Science*, vol. 3, n° 1, p. 77–82.
- Renoust B. (2014). Analysis and visualisation of edge entanglement in multiplex networks (Thèse de doctorat, University of Bordeaux). HAL.
- Renoust B., Melançon G., Viaud M.-L. (2013). Assessing group cohesion in homophily networks. In *Advances in social network analysis and mining (asonam) 2013*, p. 149-155. ACM/IEEE.
- Rossi F., Villa-Vialaneix N. (2012). Représentation d'un grand réseau à partir d'une classification hiérarchique de ses sommets. *Journal de la Société Française de Statistique*, vol. 152, n° 3, p. 34–65.
- Zhou T., Ren J., Medo M., Zhang Y.-C. (2007). Bipartite network projection and personal recommendation. *Physical Review E*, vol. 76, n° 4, p. 046115.