
Introduction d'indices structuraux pour l'analyse de réseaux multiplexes

Application à l'analyse d'un thésaurus

Romain Boulet

UMR ESPACE-DEV

IRD

500 Avenue Jean-François Breton

34093 MONTPELLIER

romain.boulet@ird.fr

RÉSUMÉ. L'analyse de réseaux complexes multiplexes est au coeur de nos préoccupations. Dans cet article nous introduisons de nouveaux indices pour l'analyse structurale de ces réseaux multiplexes. Ces indices portent sur les grandes familles d'indices structuraux : connexité, connectivité locale et globale (transitivité) et les communautés. Le calcul de ces indices a été effectué sur un réseau réel consistant en un thésaurus.

ABSTRACT. The analysis of multiplex complex networks is at the heart of our concerns. In this paper we introduce new indices for the structural analysis of these multiplex networks. These indices refers to the major structural families of indices: connectedness, local and global connectivity (transitivity) and communities. The computation of these indices was carried out on a real network consisting of a thesaurus.

MOTS-CLÉS : Réseaux, graphes, réseaux sémantiques, structure, multiplicité, matrices, thésaurus, environnement

KEYWORDS: Networks, graphs, semantic networks, structure, multiplicity, matrices, thesaurus, environment

1. Introduction

Un réseau est un ensemble de nœuds reliés par des liens. Nous appelons *réseau multiplexe* un réseau possédant des liens de différentes natures. Notre intérêt se porte sur des réseaux réels comme par exemple les réseaux sociaux ou les réseaux de termes. Par exemple les différents liens d'un réseau social peuvent être des liens d'amitiés, de famille, de relations professionnelles.

Un réseau réel peut être orienté, pondéré, multiplexe, dynamique et l'étude d'un tel réseau complexe peut débiter par l'étude d'une modélisation simple de ce dernier. En effet l'étude du graphe statique, simple, avec un seul type de lien est plutôt bien maîtrisée ; différentes catégorisations existent (petit-monde, sans échelle typique) et la structure est appréhendée par une variété d'invariants calculés sur le graphe (diamètre, densité, etc.). Considérer le graphe simple associé à un réseau plus complexe permet d'obtenir un premier aperçu de sa structure et peut être vu comme le squelette sous-jacent de ce réseau. Beaucoup de travaux existants reposent sur l'étude du graphe avec un seul type de lien. Dans cet article nous voulons aller au-delà en prenant en compte l'aspect multiplexe pour analyser la structure d'un réseau. Notre domaine d'application porte sur des réseaux de termes limités aux thésaurus.

Un thésaurus constitue une référence admise consensuellement en terme de langage. Il est constitué par des relations de divers types. Ces thésaurus peuvent ensuite être modélisés en réseaux sémantiques c'est-à-dire en graphe où chaque sommet représente un terme et où il existe des arêtes entre sommets représentant les relations possibles entre termes. Nous avons donc un réseau multiplexe. Dans cet article nous nous focaliserons sur le thésaurus GEMET (GENERAL Multilingual Environmental Thesaurus) qui est un thésaurus multilingue de l'EIONET (European Environment Information and Observation Network) et qui définit une terminologie générale pour l'environnement. Ce thésaurus possède deux types de liens¹ : les liens hiérarchiques que nous dénommerons *narrower/broader* et des liens associatifs appelés *related to*. Il possède 5208 sommets (concepts), 5051 liens *narrower/broader* et 1043 liens *related to*. Une représentation de ce thésaurus ainsi qu'un exemple de liens entre quelques termes sont donnés en figure 1. Nous y voyons que le concept *équipement thermique* est une généralisation (lien *narrower/broader*) du concept *four industriel* qui lui même est une généralisation des trois concepts *chambre de combustion à petite échelle*, *four tournant*, *haut fourneau* ; ces trois ci sont en relation (lien *related to*) avec le concept de *cuisson* qui est généralisé par (lien *narrower/broader*) le concept de *traitement physicochimique*.

Dans la littérature, la multiplicité est souvent abordée via les conséquences qu'elle induit sur la position des sommets, l'étude se focalisant sur les sommets et la multiplicité des arêtes incidentes à ces sommets. Par exemple l'étude de réseaux sociaux multiplexes entraîne l'étude des rôles des individus [BRA 05, Chap.9] de ce réseau ; à titre illustratif citons [KRO 88] qui met en avant la multiplicité en permettant de

1. Un troisième type de liens – la synonymie, aussi appelée *use for* – peut exister dans les thésaurus mais n'est pas présent dans GEMET.

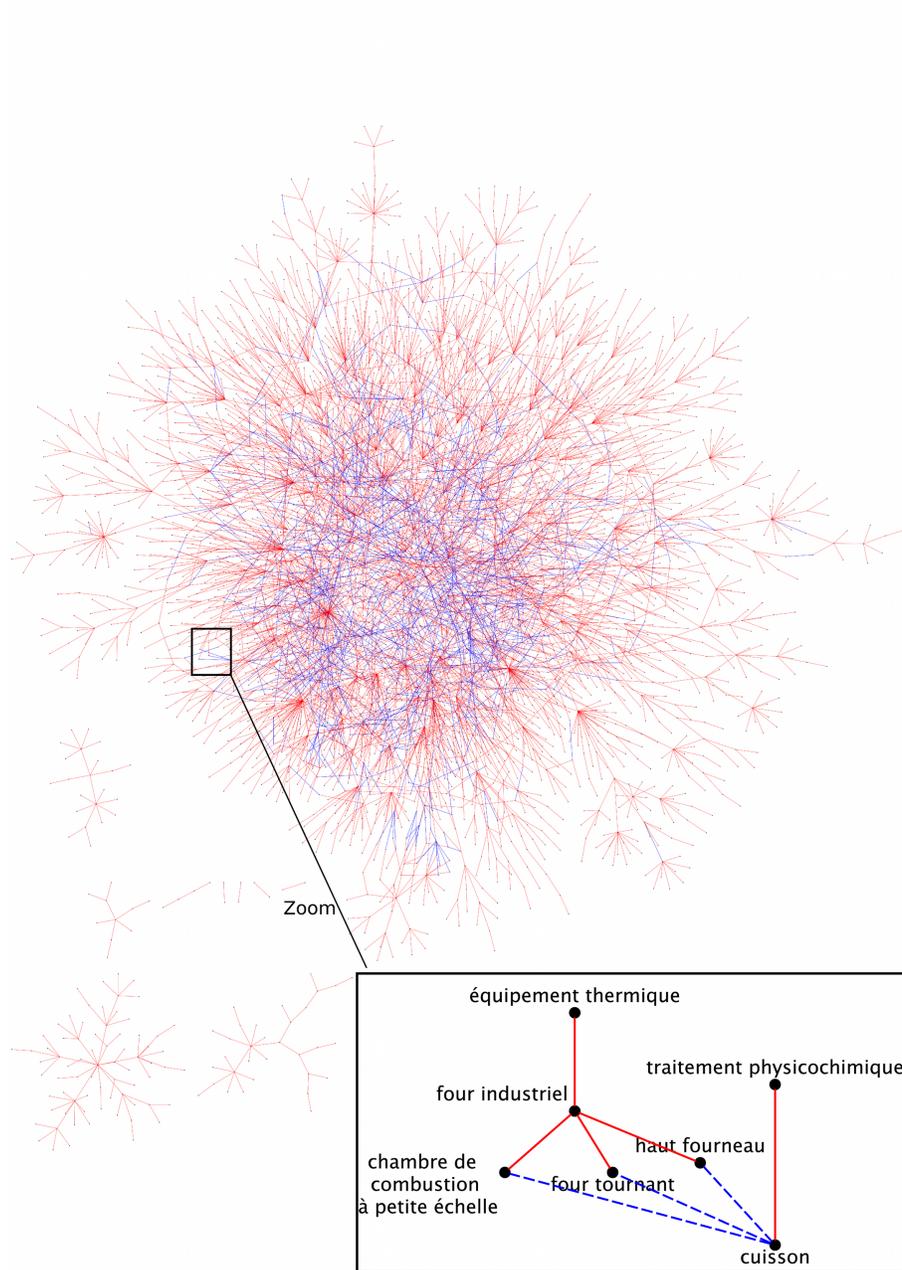


Figure 1. Représentation du thésaurus GEMET, les liens rouges sont les liens narrower/broader, les bleus sont les liens related to. Un focus sur quelques sommets (dans l'encart en bas à droite) est donné à titre d'illustration. Dans ce focus les liens related to sont en trait bleu discontinu, les liens narrower/broader sont en trait rouge continu.

détecter l'importance de certains types de liens incidents à un nœud – représentant un adolescent – dans le tabagisme de celui-ci. L'étude des dyades² (notions de réciprocité) est également au coeur d'analyses de réseaux multiplexes [FIE 85] (par exemple la présence d'un type de liens entre deux sommets implique-t-elle la présence d'un autre type de lien ?).

Dans ce papier nous prenons une autre direction. Au lieu de mettre les sommets, leurs rôles ou les relations simplement dyadiques au coeur de l'analyse, nous focalisons notre étude sur les arêtes, leur agencement et la structure globale qu'elles procurent au réseau. De plus, notons que les deux types d'arêtes *narrower/broader* et *related to* présents dans le thésaurus étudié ne peuvent pas co-exister dans une même dyade, l'étude de leurs positions dans le réseau doit donc dépasser le stade de la dyade.

Le but de cet article est double : il s'agit d'une part de proposer de nouveaux indices structuraux et une méthodologie générale (métrologie) ouvrant la voie à une étude structurelle de réseaux multiplexes ; et d'autre part d'appliquer cette méthodologie à l'étude du thésaurus GEMET en espérant qu'une meilleure connaissance de sa structure permettra de mieux le manipuler et de mieux maîtriser son enrichissement. Ces indices mesurant l'apport des termes *related to* à la structure du réseau, sont reliés aux notions standards de connectivité locale, globale ; la section 2 introduit les nouveaux indices, leur application dans le contexte de GEMET est effectuée en section 3. La dernière partie de ce papier conclut et ouvre les perspectives de ces développements.

2. Indices de structure sur graphes multiplexes

2.0. Formalisme / Méthodologie

Dans le cadre du thésaurus que nous étudions, seuls deux types de liens peuvent cohabiter, les liens *narrower/broader* et les liens *related to*, et il est clair que les premiers constituent la base du thésaurus et que les derniers sont des apports supplémentaires à cette structure. Nous étudierons donc l'apport de ces liens dans la structuration du réseau. Dans la suite, un graphe $G = (V, E)$ est défini par son ensemble V de sommets et son ensemble E d'arêtes. Le graphe G_n (resp. G_r) désigne le graphe dont les sommets est l'ensemble des termes du thésaurus et les arêtes sont de type *narrower/broader* (resp. *related to*) ; le graphe G_t est le graphe *total* possédant les deux types d'arêtes. Pour un graphe G nous noterons \tilde{G} le sous-graphe correspondant à sa plus grande composante connexe. Par souci de concision nous noterons $\tilde{G}_n \cup E_r$ le graphe $(V(\tilde{G}_n), E(G_r) \cap E(\tilde{G}_n))$, c'est à dire le graphe \tilde{G}_n auquel on ajoute les arêtes de G_r reliant deux sommets de \tilde{G}_n .

2. Relations entre deux éléments

Les indices que nous allons définir vont mesurer l'apport des liens *related to* à la structure du réseau. Ils sont reliés à quatre grandes familles d'indices structuraux³ : connexité, connectivité locale et globale (transitivité), communautés. Afin de mieux appréhender la valeur de l'indice proposé, nous le calculons dans deux cas de figure : un réseau réel (GEMET) et un modèle où les liens *related to* sont placés de façon aléatoire (suivant une loi uniforme) sur les non-arêtes du graphe G_n (ou \widetilde{G}_n selon le cas). Dans la suite, les indices présentés sur des graphes aléatoires désigneront un indice moyen obtenu en simulant 1000 graphes aléatoires ainsi construits.

2.1. Multiplicité et connexité

Le fait d'ajouter des liens peut augmenter la connexité du graphe, par exemple en rendant connexe un graphe qui ne l'est pas : les liens de type *related to* peuvent contribuer à connecter différents domaines qui ne l'étaient pas par des liens *narrower/broader*. Nous mesurerons cet apport en analysant le rapport de la croissance de la plus grosse composante connexe observée par l'ajout des liens *related to* à la croissance totale possible

$$\frac{\frac{|V(\widetilde{G}_t)|}{|V(\widetilde{G}_n)|} - 1}{\frac{|V(G_t)|}{|V(G_n)|} - 1} = \frac{|V(\widetilde{G}_t)| - |V(\widetilde{G}_n)|}{|V(G_t)| - |V(G_n)|}$$

De plus, ces liens peuvent également renforcer la connexité par sommets (resp. arcs) qui est le nombre minimal de sommets (resp. arêtes) nécessaire à enlever pour rendre le graphe non-connexe. Les valeurs des connexités par sommets ou par arcs des deux réseaux peuvent ne pas se révéler pertinentes : la présence de feuilles (sommets de degré 1) suffit à faire tomber ces indices de connexité à 1 ; pour que cette valeur augmente dans le graphe \widetilde{G}_t il faudrait que toutes les feuilles deviennent des sommets de degré supérieur ou égal à deux grâce à l'adjonction des liens *related to*. Pour pallier cet inconvénient nous pouvons utiliser la connexité algébrique définie comme étant la deuxième plus petite valeur propre du Laplacien combinatoire du graphe et qui est liée aux différentes notions de connexité sus-citées [FIE 73].

2.2. Multiplicité et connectivité globale

Nous abordons désormais l'effet de l'ajout de liens de type *related to* sur les indices structuraux relatifs aux graphes en petit-monde. Un graphe en *petit-monde* possède deux propriétés structurales : une connectivité globale resserrée – qui va être traitée ici – et une connectivité locale forte qui sera traitée en sous-section suivante. La

3. Dans ce papier nous n'avons pas abordé les indices structuraux de manière exhaustive, nous pouvons mentionner d'autres mesures telles la distribution des degrés [BAR 99] ou la centralisation [FRE 79].

connectivité globale peut être mesurée, entre autres, par la moyenne des longueurs des plus courts chemins du graphe. L'ajout de nouveaux liens dans le réseau ne peut que faire diminuer cette valeur. Le diamètre (la longueur du plus grand des plus courts chemins du graphe) est aussi une mesure de la connectivité globale bien que peu robuste, sa valeur va permettre d'évaluer l'apport des liens *related to* sur le rapprochement des sommets situés en périphérie du réseau.

La position stratégique sur des géodésiques⁴ du graphe des nouveaux liens peut être évaluée via le calcul de la centralité d'intermédiarité des arêtes [BRA 08] définie, pour une arête e , par

$$\sum_{i,j \in V(G), i \neq j} \frac{g_{i,e,j}}{g_{i,j}}$$

où $g_{i,e,j}$ est le nombre de géodésiques liant les sommets i et j et passant par l'arête e et $g_{i,j}$ est le nombre total de géodésiques liant i et j .

La comparaison s'effectuera avec un modèle aléatoire présenté précédemment. Notons que pour la connectivité globale, il est connu [WAT 03] que le fait de placer des arêtes de façon aléatoire dans un réseau a pour conséquence de créer de nombreux raccourcis dans ce réseau et donc de faire chuter la moyenne des plus courts chemins et le diamètre.

2.3. Multiplicité et connectivité locale

Dans une étude structurale menée sur un graphe simple (non multiplexe) la connectivité locale peut être mesurée par le deuxième coefficient de clustering défini comme étant le rapport entre le nombre de triangles du graphe et le nombre de triplets connexes de ce graphe [BOR 03]. Il correspond à la probabilité qu'un lien soit créé entre deux sommets sachant que ces deux sommets possèdent un voisin commun ou bien comme un taux de transitivité du réseau. L'apport des liens *related to* à cette transitivité peut alors être mesuré par le nombre de fermetures supplémentaires de triangles engendrées par l'ajout de ces liens. Nous allons comparer le nombre de triangles hétérogènes avec deux liens *narrower/broader* et un lien *related to* avec le nombre de triplets connexes homogènes *narrower/broader* afin d'évaluer si l'ajout des liens *related to* crée des fermetures de triangles (i.e. crée de la transitivité, crée du clustering).

Comme précédemment, la comparaison avec un modèle aléatoire prend son sens : si les modèles aléatoires binomiaux produisent de bons résultats quand à la moyenne des plus courts chemins, leur comportement vis à vis de la connectivité locale est loin de suivre un modèle de petit-monde.

Une méthode de dénombrement de ces triangles hétérogènes basée sur le spectre du graphe est exposée en annexe 5.

4. une géodésique est un plus court chemin

2.4. Multiplexité et communautés

La question sous-jacente à cette sous-section est la suivante : les liens de type *related to* sont-ils plutôt des liens inter-communautés ou bien des liens intra-communautés ?

Pour apporter des éléments de réponse, nous partitionnons le graphe \widetilde{G}_t avec des algorithmes existants⁵ et on calcule le nombre de liens de chaque type dans et entre les communautés. Comme précédemment la comparaison est effectuée avec un modèle aléatoire (des liens sont ajoutés aléatoirement et uniformément sur les non-arêtes de \widetilde{G}_n) partitionné avec le même algorithme.

3. Application au thésaurus GEMET

3.1. Multiplexité et connexité

Le tableau 1 répertorie les indices introduits en section 2.1 mesurant l'apport des liens *related to* à la connexité du graphe. Le gain de connexité de GEMET est proche du gain obtenu par l'aléatoire. Notons que la plupart des réseaux réels possèdent une composante connexe géante. Le graphe G_n ne possédait pas cette propriété puisque seuls 38% de ses sommets appartenaient à la plus grande composante connexe ; en revanche l'ajout des liens *related to* permet d'atteindre le pourcentage de 97%. Les mesures de connexités par sommets et par arcs ne sont pas pertinentes à cause de la présence de nombreux sommets de degré 1 dans les réseaux étudiés. Quant à la connexité algébrique de GEMET, même si sa progression est importante avec l'ajout des liens *related to*, elle n'atteint pas celle d'un modèle aléatoire.

Indice	GEMET	Aléatoire
$ V(G_t) $	5208	5208
$ V(\widetilde{G}_t) $	5059	5198
$ V(\widetilde{G}_n) $	1958	1958
gain	0.95	0.99
$v(\widetilde{G}_n)$	1	1
$v(\widetilde{G}_n \cup E_r)$	1	1
$e(\widetilde{G}_n)$	1	1
$e(\widetilde{G}_n \cup E_r)$	1	1
$\lambda_2(\widetilde{G}_n)$	$7.739 \cdot 10^{-4}$	$7.739 \cdot 10^{-4}$
$\lambda_2(\widetilde{G}_n \cup E_r)$	$9.315 \cdot 10^{-3}$	$3.07 \cdot 10^{-2}$

Tableau 1. Gain de connexité obtenu par la prise en compte des liens *related to* dans le thésaurus GEMET et comparaison avec un modèle aléatoire.

5. Le choix de l'algorithme de partitionnement n'est pas le but de cet article.

3.2. Multiplexité et connectivité globale

Indice	GEMET	Aléatoire
Moyenne des plus courts chemins de \widetilde{G}_n	15.09	15.09
Moyenne des plus courts chemins de $\widetilde{G}_n \cup E_r$	9.59	8.58
Diamètre de \widetilde{G}_n	41	41
Diamètre de $\widetilde{G}_n \cup E_r$	23	18.70

Tableau 2. Influence des liens *related to* sur la connectivité globale du thésaurus GEMET et comparaison avec un modèle aléatoire de même ordre et taille.

La connectivité globale, ici mesurée par la moyenne des plus courts chemins et par le diamètre (tableau 2), diminue nettement lors de l'ajout des liens *related to* en la faisant tendre vers celle d'un graphe aléatoire, ces liens créent donc des raccourcis dans le réseau et ont en ce sens un rôle stratégique dans la navigation au sein du réseau. De plus, la courbe en figure 2 représente la part d'arêtes de type *related to* parmi les $x\%$ arêtes de plus forte intermédiarité ; cette proportion est à comparer avec la part globale d'arêtes *related to* dans le graphe (représentée en pointillés) et la part de telles arêtes dans un graphe où les liens *related to* sont disposés de manière uniformément aléatoire (représentée en trait discontinu). Ce résultat vient confirmer la position stratégique des liens *related to*.

3.3. Multiplexité et connectivité locale

Comme détaillé en section 2.3, la connectivité locale est ici mesurée par le deuxième coefficient de clustering, basé sur le nombre de triangles dans le réseau, calculé sur GEMET et sur un modèle aléatoire. Dans le cas du modèle aléatoire, si nous négligeons le cas où une seule arête peut générer la fermeture de plusieurs triangles comme par exemple la diagonale d'un carré (ceci est justifié car notre thésaurus ne possède que 42 carrés⁶), la valeur théorique est approchée par l'espérance d'une loi hypergéométrique de paramètres $N = |E(\widetilde{G}_n)|$, $m = |\text{triplets connexes narrower différent d'un triangle}|$ et $n = |E(G_r)|$ (tirage aléatoire sans remise de n arêtes dans un ensemble de N arêtes dont m arêtes créent une fermeture de triangles) qui est $\frac{nm}{N}$.

Nous constatons (tableau 3) une transitivité marquée dans GEMET. Les résultats des sections 3.2 et 3.3 montrent d'une part que les liens *related to* contribuent à diminuer la connectivité locale et à augmenter le clustering ; les liens *related to* transforment donc l'arbre induit par les liens *narrower/broader* en un graphe de petit-monde.

6. le nombre de carrés d'un graphe G pouvant être calculé par la formule suivante : $\sum_i \lambda_i^4 - 2|E(G)| - 4 \times (\text{nb de triplets connexes})$ où les λ_i sont les valeurs propres de la matrice d'adjacence du graphe.

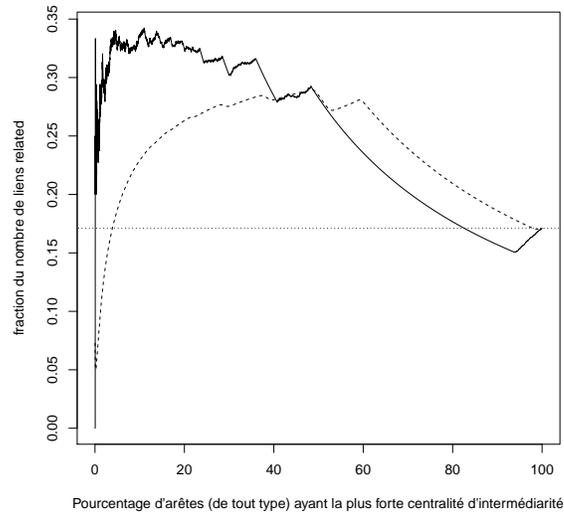


Figure 2. Fraction d'arêtes de types related to parmi les $x\%$ d'arêtes à plus forte centralité d'intermédiarité pour le thésaurus GEMET (trait plein) et pour un graphe aléatoire (trait discontinu). La ligne en pointillés représente la fraction d'arêtes de type related to dans l'ensemble du graphe.

Indice	GEMET	Aléatoire
nb de triangles narrower/broader	16	16
nb de triangles avec 2 narrower/broader et un related to	38	2.27 (2.30)
nb de triplets connexes narrower/broader	28145	28145

Tableau 3. Nombre de triangles homogènes de type narrower/broader, nombre de triangles hétérogènes possédant deux liens narrower/broader et un lien related to et nombre de triplets connexes narrower/broader (trois sommets reliés par au moins deux liens narrower/broader) du graphe représentant le thésaurus GEMET et pour un graphe où les liens related to sont disposés de façon aléatoire (entre parenthèses la valeur théorique).

3.4. Multiplicité et communautés

Pour effectuer le partitionnement nous avons choisi l'algorithme *fast-greedy* [CLA 04] pour sa renommée et popularité dans la littérature, pour sa rapidité et les bons résultats donnés via le critère de modularité. De plus une approche comparative de plusieurs algorithmes de partitionnement sur des thésaurus a été menée dans [SAY 11] faisant émerger *fast-greedy* comme étant le mieux adapté à ces données issues de thésaurus en considérant le graphe simple associé aux thésaurus. Nous souhaitons déterminer si les relations *related to* font plutôt le lien entre communautés (au vu des résultats sur leur position stratégique, section 3.2) ou bien s'ils renforcent la structure intra-communauté (puisqu'ils sont ajoutés entre termes sémantiquement proches).

Indice	GEMET	Aléatoire
nb liens <i>related to</i> intra-communauté	879	946
nb liens <i>related to</i> inter-communauté	164	97
nb liens <i>narrower/broader</i> intra-communauté	4635	4195
nb liens <i>narrower/broader</i> inter-communauté	416	856

Tableau 4. Répartition (intra-communauté ou inter-communauté) des différents types de liens lors d'un partitionnement du réseau.

Les résultats sont présentés en tableau 4 et montrent que les liens *related to* sont en proportion plus présents entre les communautés que les liens *narrower/broader* (16% des liens contre 8%) alors que cette proportion est inversée dans le cas de simulations aléatoires. Cependant la comparaison de ce résultat avec les simulations aléatoires est rendue difficile par le fait que le partitionnement, et le nombre de communauté de ce partitionnement, n'est pas figé pour chaque graphe.

4. Conclusion

4.1. Synthèse

Notre propos fut d'introduire de nouveaux indices pour mesurer la structure sous-jacente d'un réseau construits en mettant en avant la position des arêtes dans le réseau. Ils mesurent l'apport de l'ajout de liens de différente nature dans la structuration d'un réseau qui devient alors multiplexe, viennent enrichir la métrologie existante pour l'analyse de grands graphes et incitent à prendre en compte l'aspect multiplexe de ces derniers.

Pour le graphe réel ici étudié (le thésaurus GEMET), les arêtes de type *related to* augmentent la connexité du réseau, diminuent leur étendue (resserrent la connectivité locale en diminuant la moyenne des plus courts chemins) et augmentent la transitivité, transformant le réseau initial en réseau de petit-monde.

Ces nouveaux indices ont été exploités dans le champ applicatif du thésaurus GEMET. L'intérêt de ces indices peut être exploité par la communauté éditrice du thésaurus puisqu'ils permettent d'évaluer l'impact des liens *related to* dans l'enrichissement du thésaurus et également de mieux contrôler l'ajout de ces liens en des zones stratégiques selon la volonté d'améliorer la connexité/connectivité du thésaurus.

4.2. Perspectives

Afin de compléter notre réflexion et d'ouvrir la voie à de futurs travaux qui porteront sur l'analyse de nouveaux indices structuraux permettant de catégoriser les réseaux réels étudiés, nous proposons une liste possible de classes de réseaux multiplexes :

- Transformation en petit-monde. Un type d'arêtes fait tendre le graphe vers un modèle de petit monde, comme exposé précédemment c'est le cas de GEMET.
- Localisation du multiplexage. Les arêtes sont situées principalement dans une sous-partie du graphe. Ce n'est pas le cas du thésaurus GEMET ici étudié : les liens *related to* ne sont pas localisés dans une partie précise du réseau, ils affectent des propriétés structurales globales du graphe, cet aspect peut également être évalué de visu grâce à la figure 1.
- Renforcement de la structure existante. L'ajout de nouveaux types de liens ne modifie pas fondamentalement la structure, les dyades peuvent être renforcées en possédant plusieurs arêtes de différents types.
- Transformation en monde concentré. Les arêtes vont transformer le réseau en monde concentré [MAZ 09] : forte densité des liens et les centralités (les pouvoirs) sont concentrés au sein d'un groupe de sommets.
- Centralisation du réseau. Les arêtes vont contribuer à donner une position centrale à certains sommets. Ceci rejoint les études sur la multiplicité et les rôles abordée en section 1.

Ces modèles ne sont pas disjoints, nous pouvons avoir une localisation du multiplexage et une centralisation du réseau par exemple.

Cette étude préliminaire sur l'impact de la multiplicité sur la structure réticulaire a vocation à être poursuivie et ouvre d'autres questions comme l'extension de ces développements à des réseaux multiplexes possédant plus de deux types de liens.

Enfin, du côté applicatif, au-delà de GEMET nous souhaitons élargir l'application de cette méthodologie à d'autres thésaurus afin de comparer leur structure et en établir une typologie et, pour mieux mettre en avant l'organisation structurale, nous désirons mettre en place des algorithmes de partitionnement prenant en compte l'aspect multiplexe de ces réseaux.

5. Annexe : dénombrement des triangles via le spectre

Dans un graphe simple le nombre de triangles est donné par la somme des valeurs propres au cube de la matrice d'adjacence [BIG 93], nous introduisons ici une approche similaire basée sur des méthodes algébriques pour calculer le nombre de triangles de différents types (homogènes où les trois arêtes sont d'un seul type, hétérogènes où le triangle est composé d'arêtes de types différents) dans un graphe ayant deux types de liens.

Soit $\zeta = e^{\frac{2i\pi}{5}}$ une racine primitive cinquième de l'unité et considérons la matrice $A = A_n + \zeta A_r$ où A_n (resp. A_r) est la matrice d'adjacence de G_n (resp. G_r). On a :

$$\begin{aligned} (A_n + \zeta A_r)^3 &= A_n^3 + \zeta(A_n^2 A_r + A_r A_n^2 + A_n A_r A_n) \\ &\quad + \zeta^2(A_r^2 A_n + A_n A_r^2 + A_r A_n A_r) + \zeta^3 A_r^3 \end{aligned}$$

Rappelons [LAN 02] que l'extension cyclotomique $\mathbb{Q}(\zeta)$ peut être vue comme un \mathbb{Q} -espace vectoriel de dimension 4 dont une base est $(1, \zeta, \zeta^2, \zeta^3)$, c'est-à-dire que tout élément de $u \in \mathbb{Q}(\zeta)$ se décompose de façon unique $u = \alpha_0 + \alpha_1 \zeta + \alpha_2 \zeta^2 + \alpha_3 \zeta^3$ avec $\alpha_i \in \mathbb{Q}$, nous avons donc

$$\exists! (\alpha_0, \alpha_1, \alpha_2, \alpha_3) \in \mathbb{N}^4 : \text{tr}(A^3) = \alpha_0 + \alpha_1 \zeta + \alpha_2 \zeta^2 + \alpha_3 \zeta^3$$

De plus, il est facile de voir que si A_1, A_2 , et A_3 sont trois matrices d'adjacences de trois graphes sur le même ensemble de sommets et ayant respectivement E_1, E_2 et E_3 comme ensemble d'arêtes alors l'entrée (i, j) de la matrice $A_1 A_2 A_3$ donne le nombre de marches de longueur 3 allant de i vers j tels que la $k^{\text{ième}}$ arête appartienne à E_k ⁷. Nous avons donc :

- α_0 est égal à six fois le nombre de triangles dont les trois liens sont des *narrower/broader*
- α_1 est égal à six fois le nombre de triangles avec deux liens *narrower/broader* et un lien *related to*
- α_2 est égal à six fois le nombre de triangles avec deux liens *related to* et un lien *narrower/broader*
- α_3 est égal à six fois le nombre de triangles dont les trois liens sont des *related to*

Le spectre de $A = A_n + \zeta A_r$ nous donne donc les nombres des différents types de triangles (la trace de A^3 est donnée par la somme des valeurs propres au cube de A).

Notons que la recherche des entiers α_i peut être simplifiée en écrivant $\zeta^k = \cos\left(\frac{2k\pi}{5}\right) + i \sin\left(\frac{2k\pi}{5}\right)$ et $\cos\left(\frac{2\pi}{5}\right) = \frac{\sqrt{5}-1}{4}$, $\cos\left(\frac{4\pi}{5}\right) = \cos\left(\frac{6\pi}{5}\right) = \frac{-\sqrt{5}-1}{4}$, $\sin\left(\frac{2\pi}{5}\right) = \frac{\sqrt{10+2\sqrt{5}}}{4}$, $\sin\left(\frac{4\pi}{5}\right) = -\sin\left(\frac{6\pi}{5}\right) = \frac{\sqrt{10-2\sqrt{5}}}{4}$,

7. Ce résultat est valable pour plus de trois matrices et pour des marches de longueur supérieure à trois.

ce qui nous donne

$$\begin{cases} 4\operatorname{Re}(u) = (4\alpha_0 - \alpha_1 - \alpha_2 - \alpha_3) + (\alpha_1 - \alpha_2 - \alpha_3)\sqrt{5} \\ 2\operatorname{Im}(u)\sqrt{10 + 2\sqrt{5}} = 5\alpha_1 + (\alpha_1 + 2\alpha_2 - 2\alpha_3)\sqrt{5} \\ -2\operatorname{Re}(u) + \operatorname{Im}(u)\left(\sqrt{10 + 2\sqrt{5}} + \sqrt{10 - 2\sqrt{5}}\right) = \\ \quad (-2\alpha_0 + 3\alpha_1 + 3\alpha_2 - 2\alpha_3) + (\alpha_1 + \alpha_2)\sqrt{5} \end{cases}$$

Quant au nombre de triplets connexes homogènes de type *narrower/broader*, il est égal au nombre de chemins de longueur 2 dans G_n c'est-à-dire $\sum_{u \in V(G_n)} C_2^{d(u)}$ où $d(u)$ est le degré de u dans G_n (chaque sommet u est le centre de $C_2^{d(u)}$ chemins de longueur 2).

6. Bibliographie

- [BAR 99] BARABÁSI A.-L., ALBERT R., « Emergence of Scaling in Random Networks », *Science*, vol. 286, n° 5439, 1999, p. 509-512.
- [BIG 93] BIGGS N., *Algebraic graph theory*, Cambridge mathematical library, Cambridge University Press, 1993.
- [BOR 03] BORNHOLDT S., SCHUSTER H. G., Eds., *Handbook of Graphs and Networks - From the Genome to the Internet*, Wiley-VCH, 2003.
- [BRA 05] BRANDES U., ERLEBACH T., *Network analysis : methodological foundations*, Lecture notes in computer science : Tutorial, Springer, 2005.
- [BRA 08] BRANDES U., « On variants of shortest-path betweenness centrality and their generic computation », *Social Networks*, vol. 30, n° 2, 2008, p. 136 - 145.
- [CLA 04] CLAUSET A., NEWMAN M. E. J., MOORE C., « Finding community structure in very large networks », *Physical Review E*, vol. 70, n° 6, 2004, page 066111.
- [FIE 73] FIEDLER M., « Algebraic connectivity of graphs », *Czechoslovak Mathematical Journal*, vol. 23, n° 98, 1973, p. 298–305.
- [FIE 85] FIENBERG S. E., MEYER M. M., WASSERMAN S. S., « Statistical Analysis of Multiple Sociometric Relations », *Journal of the American Statistical Association*, vol. 80, n° 389, 1985, p. 51–67, American Statistical Association.
- [FRE 79] FREEMAN L. C., « Centrality in social networks : Conceptual clarification », *Social Networks*, vol. 1, n° 3, 1979, p. 215–239.
- [KRO 88] KROHN M., J.L. M., M. Z., « Role Overlap, Network Multiplexity, and Adolescent Deviant Behavior », *Social Psychology Quarterly*, vol. 51, n° 4, 1988, p. 346-356.
- [LAN 02] LANG S., *Algebra*, Graduate texts in mathematics, Springer, 2002.
- [MAZ 09] MAZZEGA P., BOURCIER D., BOULET R., « The network of French legal codes », *Proceedings of the 12th International Conference on Artificial Intelligence and Law, ICAIL '09*, ACM, 2009, p. 236–237.
- [SAY 11] SAYAH H., « Composant sémantique pour l'amélioration de la recherche de données environnementales », Master's thesis, Université Montpellier II, 2011.
- [WAT 03] WATTS D. J., *Small Worlds : The Dynamics of Networks between Order and Randomness*, Princeton University Press, November 2003.