



Opérations matricielles et analyse de graphe

Laurent Beauguitte, Pierre Beauguitte

► **To cite this version:**

Laurent Beauguitte, Pierre Beauguitte. Opérations matricielles et analyse de graphe. 2011. <halshs-00632078>

HAL Id: halshs-00632078

<https://halshs.archives-ouvertes.fr/halshs-00632078>

Submitted on 13 Oct 2011

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.

Opérations matricielles et analyse de graphe

Laurent Beauguitte
CNRS, UMR Géographie-cités, bl<at>parisgeo.cnrs.fr
Pierre Beauguitte
Université Paris Diderot Paris 7, pierrebeauguitte<at>gmail.com

Automne 2011 - Version 1



Introduction

Un graphe peut se représenter sous trois formes principales : deux listes (une de sommets et une de liens), un graphe au sens strict et enfin une matrice d'adjacence où, conventionnellement, les origines sont en ligne et les destinations en colonne [4]. Cette dernière notation est nécessaire pour mesurer un grand nombre d'indicateurs utiles en analyse de réseaux.

L'objectif de ce document du groupe fmr est de rappeler les bases du calcul matriciel et de montrer quels indicateurs elles permettent d'obtenir facilement. Toutes les opérations indiquées sont aisément reproductibles avec le logiciel libre et gratuit R (voir les programmes en annexe).

1 Matrices

1.1 Définitions et opérations matricielles de base

Une matrice A de dimension $m \times n$ est un tableau de valeurs à m lignes et n colonnes. On note $a_{i,j}$ la valeur de l'élément situé à la i -ième ligne et j -ième colonne. Il est possible de multiplier une matrice A avec une matrice B si A est de dimension $m \times n$ et B de dimension $n \times p$. Dans ce cas $C = AB$ est la matrice de dimension $m \times p$ obtenue par

$$c_{i,j} = \sum_{1 \leq k \leq n} a_{i,k} b_{k,j}$$

L'élément $c_{i,j}$ est la somme des produits des éléments de la ligne i de A par ceux de la colonne j de B . Le tableau 1 détaille le principe de la multiplication matricielle.

TABLEAU 1 – Multiplication matricielle

$$\begin{array}{cccc}
 \text{A} & \text{B} & \text{AB} & \text{AB} \\
 \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 3 & 2 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} & \begin{bmatrix} 2 \times 1 + 1 \times 2 & 2 \times 0 + 1 \times 3 \\ 1 \times 1 + 0 \times 2 & 1 \times 0 + 0 \times 3 \\ 3 \times 1 + 2 \times 2 & 3 \times 0 + 2 \times 3 \end{bmatrix} & \begin{bmatrix} 4 & 3 \\ 1 & 0 \\ 7 & 6 \end{bmatrix}
 \end{array}$$

TABLEAU 2 – Addition et soustraction de matrices de même dimension

$$\begin{array}{cccc}
 \text{A} & \text{B} & \text{A + B} & \text{A - B} \\
 \begin{bmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 4 & 2 & 0 \end{bmatrix} & \begin{bmatrix} 5 & 0 & 1 \\ 2 & 0 & 0 \\ 1 & 1 & 4 \end{bmatrix} & \begin{bmatrix} 5 & 1 & 3 \\ 5 & 0 & 0 \\ 5 & 3 & 4 \end{bmatrix} & \begin{bmatrix} -5 & 1 & 1 \\ 1 & 0 & 0 \\ 3 & 1 & -4 \end{bmatrix}
 \end{array}$$

Si $m = p$, alors BA existe et n'est en général pas égal à AB (la multiplication matricielle n'est pas commutative).

Si l est un nombre, alors le produit lA est obtenu en multipliant chaque terme de A par l .

L'addition et la soustraction sont définies pour des matrices de même dimension, simplement par l'addition ou la soustraction terme à terme (voir le tableau 2).

La *transposée* d'une matrice $m \times n$ est la matrice $n \times m$ obtenue en permutant lignes et colonnes. On note A^T la transposée de A . On dit qu'une matrice A est *symétrique* si $A^T = A$.

Soit la matrice A suivante

$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Sa transposée, notée A^T , est

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Une matrice diagonale est une matrice carrée où toutes les cases $a_{i,j} = 0$ si $i \neq j$. Formulée autrement, toutes les cases contiennent des 0, exceptée la diagonale.

1.2 Vecteurs et valeurs propres

Une matrice carrée $n \times n$ peut être vue comme une application linéaire sur les vecteurs de longueur n : on peut multiplier une telle matrice par un vecteur colonne, *i.e.* une matrice $n \times 1$, et obtenir un nouveau vecteur. Une

matrice A possède toujours des *vecteurs propres* (*eigenvectors*), qui sont les vecteurs v_i vérifiant

$$Av_i = \lambda_i v_i, \quad \text{où } \lambda_i \text{ est un nombre complexe}^1$$

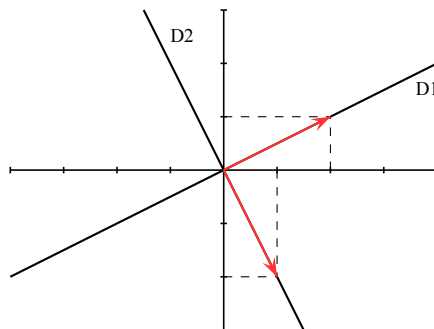
Un vecteur propre est un vecteur dont la direction n'est pas affectée par A . Ces vecteurs définissent en fait des axes invariants pour A . Le nombre λ_i est la *valeur propre* (*eigenvalue*) associée à v_i . Si la matrice est symétrique, alors les valeurs propres sont réelles. Une valeur propre peut être nulle, ce qui est un cas problématique : tout l'axe du vecteur v_i est « aplati » sur 0 par A , c'est-à-dire que A « écrase » une dimension de l'espace².

Si une même valeur propre λ est associée à plusieurs vecteurs propres indépendants (*i.e.* qui ne définissent pas les mêmes axes), le nombre de tels vecteurs est la *multiplicité* de λ , qu'on note m . Cependant, lorsqu'on énumère les valeurs propres d'une matrice, il est d'usage de faire apparaître m fois λ . À une matrice $n \times n$, on associe donc les n nombres $\lambda_i, 1 \leq i \leq n$, qui ne sont pas nécessairement distincts deux à deux.

Dans l'exemple de la figure 1, les vecteurs propres v_1 et v_2 définissent les droites $D1$ et $D2$, toutes deux invariantes pour A . Le fait que λ_2 soit nulle indique que l'image de tout vecteur de $D2$ est le point $(0, 0)$. Tout vecteur a son image sur la droite $D1$: A transforme le plan (2 dimensions) en une droite (une dimension).

FIGURE 1 – Vecteurs et valeurs propres

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \quad (v_1, \lambda_1) = \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, 5 \right) \quad (v_2, \lambda_2) = \left(\begin{bmatrix} 1 \\ -2 \end{bmatrix}, 0 \right)$$



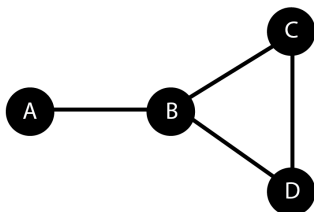
La somme des valeurs propres est égale à la *trace* de la matrice, c'est-à-dire la somme des éléments de sa diagonale.

1. \mathbb{C} désigne l'ensemble des nombres complexes, c'est-à-dire l'ensemble des sommes et produits de nombres réels et du nombre imaginaire i (tel que $i^2 = -1$).

2. En termes mathématiques, l'image de tout l'axe du vecteur v_i par A est le point 0.

TABLEAU 3 – Du graphe à la matrice laplacienne

Graphe



Matrice d'adjacence

$$\begin{array}{c}
 A \quad B \quad C \quad D \\
 A \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}
 \end{array}$$

Matrice des degrés

$$\begin{array}{c}
 A \quad B \quad C \quad D \\
 A \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}
 \end{array}$$

Matrice laplacienne

$$\begin{array}{c}
 A \quad B \quad C \quad D \\
 A \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}
 \end{array}$$

L'existence de ces vecteurs et valeurs propres est un fait mathématique, mais les déterminer est un réel problème informatique quand les matrices sont de grandes dimensions [12].

2 Matrices associées à un graphe

Un graphe non valué $G = (V, E)$ d'ordre n (nombre de sommets) peut être représenté par sa *matrice d'adjacence* A , une matrice carrée $n \times n$ où $a_{i,j} = 1$ si $(i, j) \in E$, 0 sinon. Si G est simple, la diagonale est nulle. S'il est non orienté, A est symétrique.

Soit G un graphe simple non valué, orienté ou non. On définit la *matrice laplacienne* (*Laplacian matrix*) $L = D - A$, où D est la matrice diagonale des degrés, telle que $d_{i,i}$ est égal au degré (nombre de liens adjacents) du sommet i , noté δ_i . Donc

$$l_{i,j} = \begin{cases} \delta_i & \text{si } i = j \\ -1 & \text{si } i \neq j, (i, j) \in E \\ 0 & \text{sinon} \end{cases}$$

Dans le tableau 3, les étapes menant au calcul d'une matrice laplacienne d'un graphe simple, non orienté et non valué, sont détaillées.

3 Applications en analyse de réseau

3.1 Opérations élémentaires

Soit A la matrice d'adjacence d'un graphe orienté ou non, éventuellement avec boucles. Alors $a_{i,j}^{(k)}$, élément de la matrice $A^k = A \times \dots \times A$, est égal au nombre de chemins de longueur k (pouvant passer plusieurs fois par un même point) entre les sommets i et j du graphe. Ainsi, la longueur du plus court chemin entre i et j est le plus petit k tel que $a_{i,j}^{(k)} \neq 0$.

Dans le cas d'un graphe biparti, où V est partitionné en X (à m éléments) et Y (à n éléments), plutôt que de considérer la matrice d'adjacence de dimension $(n+m) \times (n+m)$, on peut s'intéresser à la matrice A de dimension $m \times n$, où $a_{i,j} = 1$ si $(x_i, y_j) \in E$. Cela revient à ne considérer que les liens existants entre les deux ensembles X et Y . Alors AA^T est la matrice des cooccurrences en lignes (relatives à X), et $A^T A$ la matrice des cooccurrences en colonnes (relatives à Y).

Soit le graphe biparti G suivant

$$\begin{array}{ccc} & 1 & 2 & 3 \\ A & \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \\ B & \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \end{array}$$

où est indiquée la participation de deux géographes A et B à trois colloques de géographie 1, 2 et 3.

En multipliant G par sa transposée G^T , on obtient la matrice suivante

$$\begin{array}{cc} A & B \\ A & \begin{bmatrix} 1 & 1 \end{bmatrix} \\ B & \begin{bmatrix} 1 & 2 \end{bmatrix} \end{array}$$

et en multipliant G^T par G , on obtient

$$\begin{array}{ccc} & 1 & 2 & 3 \\ 1 & \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \\ 2 & \begin{bmatrix} 1 & 2 & 0 \end{bmatrix} \\ 3 & \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \end{array}$$

Dans le premier cas, on obtient les données relatives aux géographes (A a participé à un colloque, B à 2, et ils étaient présents à un même colloque); dans le second, les données relatives aux colloques³.

3.2 Corrélation et régression de matrices

La corrélation permet de mesurer la ressemblance entre deux matrices de même dimension⁴, la régression explique (au sens statistique du terme) une matrice par une autre. Ainsi, des tests effectués sur des matrices de similarité concernant les positions de vote à l'ONU durant la guerre froide montrent

3. L'exemple donné ici est volontairement de taille minimale... mais le principe reste évidemment le même quel que soit l'ordre du graphe.

4. Lorsqu'il s'agit de matrices de similarités, on applique le test de Mantel.

que plus de 70% de l'information de la matrice au temps $t + 1$ s'explique par la structure de la matrice au temps t (Beauguitte, 2011[3]).

3.3 Spectre d'un graphe (*graph spectra*)

La *théorie spectrale des graphes* s'intéresse aux spectres des matrices décrivant les graphes, qui sont simplement l'ensemble de leurs valeurs propres. On ne s'intéressera qu'aux matrices symétriques, donc aux graphes non orientés, afin de manipuler des nombres réels. On ordonne les valeurs propres λ_i par ordre croissant, *i.e.* $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. Le spectre est un invariant du graphe : si la numérotation des sommets change, la matrice d'adjacence va changer, mais pas ses valeurs propres.

L'utilisation des spectres des matrices en analyse de réseau est récente (excepté l'indice de Bonacich, voir *infra*) et en plein développement. Les définitions proposées et les pistes jugées les plus prometteuses varient d'un auteur à l'autre⁵. De plus, les applications sur des données empiriques restent essentiellement exploratoires. À notre connaissance, seuls les physiciens ont à ce jour mobilisé cet outil [1] [5]. Ceci s'explique sans doute par le caractère essentiellement non orienté des graphes étudiés par ces derniers : le spectre d'un graphe orienté est complexe, ce qui rend son utilisation difficile. Seules quelques propriétés élémentaires sont présentées ici.

Spectre de A

L'*eigenvector centrality* est une mesure proposée par Bonacich [6]. Il s'agit d'une mesure plus fine que le degré, pour laquelle la centralité x_i du sommet v_i est proportionnelle à la centralité des sommets adjacents. Cela peut s'écrire

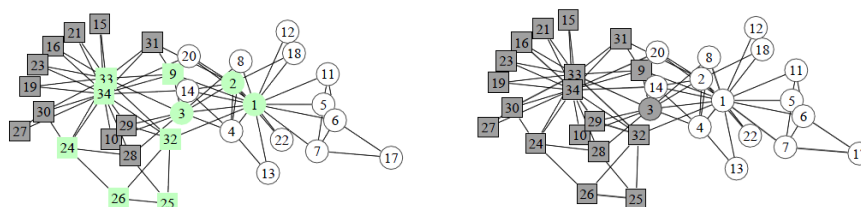
$$\lambda x_i = \sum_{1 \leq j \leq n} a_{i,j} x_j \quad \text{ou encore} \quad \lambda x = Ax, \text{ où } x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Donc les centralités sont données par un vecteur propre de A , précisément celui associé à la plus grande valeur propre λ_n . Quand G est connexe, cette valeur propre est de multiplicité 1, elle est donc associée à un unique vecteur propre.

Selon Bonacich, l'avantage principal de cette mesure de centralité est de prendre en compte la structure générale du graphe. Dans un article plus récent, il a proposé une extension aux graphes signés et valués [7]. Dans la littérature, cet indice reste peu utilisé comparé aux indicateurs « classiques » (degré, intermédiarité...).

5. Ainsi, l'appellation de graphe laplacien désigne des objets différents en fonction des auteurs (voir [14], p.3). Pour simplifier, il est possible d'utiliser trois matrices laplaciennes différentes, deux d'entre elles étant normalisées.

FIGURE 2 – Spectre du graphe et partition



Sur les deux figures, les carrés et les ronds représentent les deux groupes créés suite à la scission du club de karaté Zachari (une base de données fréquemment utilisée en analyse de réseaux sociaux). À gauche, on voit la partition obtenue grâce à une classification hiérarchique. Les sommets en vert clair ne sont assignés à aucun des deux blocs trouvés. À droite, la partition est obtenue grâce à un algorithme utilisant le spectre du graphe : seuls deux sommets se trouvent placés dans le « mauvais » bloc (figures tirées de [11]).

Spectre de L

λ_1 est toujours nulle et la multiplicité de λ_1 est le nombre de composantes connexes de G (voir notamment [13] p.74). Donc si une seule valeur propre est égale à 0, le graphe est connexe.

Clustering spectral

Les vecteurs propres d'une matrice définissent des axes invariants pour l'application définie par la matrice, donc un espace où cette application est « plus simple ». Un algorithme de *clustering* spectral est en fait un algorithme de *clustering* standard, *k-means* par exemple, mais appliqué aux vecteurs propres et non aux données elles-mêmes. Comme la structure de A est plus simple dans ce nouvel espace, les résultats sont meilleurs [2] [14].

Newman a ainsi testé un algorithme basé sur le spectre du graphe concernant les données du club de karaté Zachari [11]. Comme le montre la figure 2, la partition obtenue est plus fidèle aux données empiriques que celle provenant d'une classification hiérarchique.

Limites

L'utilisation du spectre du graphe est prometteuse mais certains inconvénients peuvent être soulignés. Il s'agit d'une méthode semble-t-il peu adaptée pour comparer des graphes : plusieurs graphes peuvent partager le même spectre ; et une variation minimale de la structure du graphe peut entraîner de très fortes variations du spectre correspondant (ces propriétés sont discutées dans [15]).

Le choix de la matrice laplacienne à utiliser est discuté, sachant que plus la distribution des degrés est hiérarchisée, plus les matrices laplaciennes diffèrent [14].

Conclusion

Les logiciels utilisés en analyse de réseau ont un aspect boîte noire qu'il est aisé de dissiper si l'on se rappelle quelques unes des propriétés ci-dessus. Ainsi, il n'est nul besoin d'un logiciel spécialisé pour déterminer des graphes de coappartenances (i.e. matrices de cooccurrences) à partir de graphes bipartis. Le nombre de composantes connexes d'un graphe peut également être déterminé sans problème.

La littérature récente en analyse des réseaux est marquée par une complexité croissante, qu'il s'agisse des méthodes de partitionnement utilisant le spectre du graphe ou des méthodes d'analyses statistiques des graphes. Maîtriser les quelques bases présentées ici peut faciliter la lecture d'ouvrages et d'articles récents, du moins l'espérons-nous.

A Calcul matriciel et analyse de graphe avec R

Créer une matrice directement dans R se fait avec la fonction `matrix`. Elle s'utilise de la façon suivante : `matrix(vec, ncol=k, byrow=TRUE)` où `vec` est le vecteur de départ, `k` le nombre de colonnes et l'option `byrow` permet de savoir si la matrice est rangée par lignes ou par colonnes.

Soit les 3 lignes de code suivantes :

```
m<-c(1,2,3,4,5,6)
mat1<-matrix(m, ncol=3, byrow=TRUE)
mat2<-matrix(m, ncol=3, byrow=FALSE)
```

On obtient $\text{mat1} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ et $\text{mat2} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$

Bien entendu, dans le cas de données à visée non pédagogique, la démarche la plus courante consiste à importer ses fichiers dans R et à les transformer à l'aide de la commande `as.matrix`.

Le produit matriciel s'effectue à l'aide de la commande `%*%`. Lorsque les deux matrices ont la même dimension, la commande `*` permet d'obtenir le produit terme à terme. Si le produit matriciel est impossible, R le signale (tableaux de tailles inadéquates).

La transposée d'une matrice se calcule à l'aide de la commande `t(x)` où `x` est la matrice de départ.

Les scripts suivants permettent de reproduire toutes les opérations de ce document `fmr`.

```
#initialisation
rm(list=ls())
#création de la matrice
m<-c(0,1,1,0,0,
      1,0,1,0,0,
      1,1,0,0,0,
      0,0,0,0,1,
      0,0,0,1,0)
m<-matrix(m, ncol=5, byrow=TRUE)
#matrice diagonale des degrés
m1<-diag(colSums(m))
#matrice laplacienne
m2<-m1-m
#valeurs et vecteurs propres
eigen(m2)
```

La fonction `eigen` renvoie un liste constituée du vecteur des valeurs propres (`$values`) et de la matrice des vecteurs propres (`$vectors`) rangés en co-

lonnes. La fonction appliquée la matrice $m2$ ci-dessus fournit les résultats suivants :

```
eigen(m2)
$values
[1] 3.000000e+00 3.000000e+00 2.000000e+00
     1.776357e-15 1.110223e-15

$vectors
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.0000000 0.8164966 0.0000000 -0.5773503 0.0000000
[2,] -0.7071068 -0.4082483 0.0000000 -0.5773503 0.0000000
[3,] 0.7071068 -0.4082483 0.0000000 -0.5773503 0.0000000
[4,] 0.0000000 0.0000000 -0.7071068 0.0000000 0.7071068
[5,] 0.0000000 0.0000000 0.7071068 0.0000000 0.7071068
```

L'écriture $x \text{ e-}k$ signifie $x \cdot 10^{-k}$, $1.776357\text{e-}15$ et $1.110223\text{e-}15$ désignent en réalité 0 : les algorithmes utilisés ne peuvent donner que des valeurs approchées. Le graphe m est donc constitué de deux composantes connexes.

```
#transformation d'un graphe biparti
m2<-c(0,1,0,1,1,0)
m2<-matrix(m2, ncol=3, byrow=TRUE)
#co-appartenance en lignes
affilrow<-m2%*%t(m2)
affilrow
#co-appartenance en colonnes
affilcol<-t(m2)%*%m2
affilcol
```

Références

- [1] R. ALBERT et A.L. BARABÁSI : Statistical mechanics of complex networks. *Review of Modern Physics*, 74(1):47–97, 2002.
- [2] B. AUFFARTH : Spectral graph clustering. Rapport technique, 2007.
- [3] L. BEAUGUITTE : *L'Assemblée générale de l'ONU de 1985 à nos jours : acteur et reflet du Système-Monde. Essai de géographie politique quantitative*. Thèse de doctorat, Université Paris Diderot Paris 7, 2011.
- [4] L. BEAUGUITTE : Graphes, réseaux, réseaux sociaux : vocabulaire et notation. *Groupe fmr*, 7p., 2010 (<http://halshs.archives-ouvertes.fr/FMR/fr/>).
- [5] S. BOCCALETTI, V. LATORA, Y. MORENO, M. CHAVEZ et D.U. HWANG : Complex networks : Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006.
- [6] P. BONACICH : Power and Centrality : A Family of Measures. *The American Journal of Sociology*, 92(5):1170–11852, 1987.
- [7] P. BONACICH : Some unique properties of eigenvector centrality. *Social Networks*, 29(4):555–564, 2007.
- [8] A.E. BROUWER et W.H. HAEMERS : *Spectra of graphs*. Springer, 2011.
- [9] F.R.K. CHUNG : *Spectral graph theory*. Numéro 92. American Mathematical Society, 1997.
- [10] E.M. HAGOS : Some results on graph spectra. *Linear algebra and its applications*, 356(1-3):103–111, 2002.
- [11] M.E.J. NEWMAN : Detecting community structure in networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):321–330, 2004.
- [12] W. RICHARDS et A. SEARY : Eigen Analysis of Networks. *Journal of Social Structures*, 1(2), 2000.
- [13] P. VAN MIEGHEM : *Graph Spectra for Complex Networks*. Cambridge University Press, 2011.
- [14] U. VON LUXBURG : A Tutorial on Spectral Clustering. *Statistics and Computing*, 17(4):395–416, 2007.
- [15] R.C. WILSON et P. ZHU : A study of graph spectra for comparing graphs and trees. *Pattern Recognition*, 41(9):2833–2841, 2008.