



Sur la similarité spectrale des graphes par mesure de corrélation

Tristan Averty, Delphine Dare-Emzivat, Abdel-Ouahab Boudraa

► To cite this version:

Tristan Averty, Delphine Dare-Emzivat, Abdel-Ouahab Boudraa. Sur la similarité spectrale des graphes par mesure de corrélation. XXIXème Colloque Francophone de Traitement du Signal et des Images (GRETSI'23), Aug 2023, Grenoble, France. hal-04197112

HAL Id: hal-04197112

<https://hal.science/hal-04197112>

Submitted on 5 Sep 2023

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.

Sur la similarité spectrale des graphes par mesure de corrélation

Tristan AVERTY Delphine DARÉ-EMZIVAT Abdel-Ouahab BOUDRAA

Institut de Recherche de l'École Navale, IREnav EA 3634, BRCM Brest, École Navale CC600, 29240 Brest Cedex 9, France

Résumé – Dans cet article, nous présentons une mesure de similarité spectrale entre deux graphes basée sur un calcul de corrélation entre les spectres de leurs matrices de représentation $\mathbf{T}_\alpha := \alpha\mathbf{D} + (1 - 2\alpha)\mathbf{A}$, paramétrée par $0 \leq \alpha \leq 1$, où $\mathbf{A}$ et $\mathbf{D}$ sont respectivement la matrice d'adjacence et la matrice des degrés. Nous montrons par ailleurs que $\mathbf{T}_\alpha$ est semi-définie positive pour $\alpha \geq 1/2$. Ce travail tend à montrer la pertinence de cette mesure, qui, introduit dans un noyau de type Gaussien d'un SVM permet une classification performante de bases de données de graphes connues de la littérature et de classification de signaux réels transformés en graphe grâce à la méthode dite de visibilité. Les résultats obtenus en termes de mesure d'exactitude sont similaires voire meilleurs à ceux obtenus avec des noyaux structurels pour un temps de calcul bien moindre et ce, en ne calculant qu'un seul spectre pour chaque graphe. De plus, nous montrons l'apport de $\mathbf{T}_\alpha$ par rapport à la matrice d' α -adjacence de Nikiforov pour la classification de graphes.

Abstract – In this paper, we present a spectral similarity measure between two graphs based on a correlation measure between the spectra of their representation matrices $\mathbf{T}_\alpha := \alpha\mathbf{D} + (1 - 2\alpha)\mathbf{A}$, parametrized by $0 \leq \alpha \leq 1$, where $\mathbf{A}$ and $\mathbf{D}$ are respectively the adjacency matrix and the degree matrix. We also show that $\mathbf{T}_\alpha$ is positive semidefinite for $\alpha \geq 1/2$. This work tends to show the relevance of this measure, which, when a SVM is implemented using a Gaussian kernel, allows a powerful classification on well known graph databases of the literature and classification of real signals transformed into a graph thanks to the so-called visibility method. The obtained results in terms of accuracy are similar or even better than those obtained with structural kernels with a much lower computation time and this, by computing only one spectrum for each graph. Moreover, we show the contribution of $\mathbf{T}_\alpha$ compared to the α -adjacency matrix of Nikiforov for graph classification.

1 Introduction

Le graphe constitue une structure bien appropriée à la représentation des relations existantes entre des grands ensembles de données collectées dans des contextes très différents. De nombreux exemples issus de la physique, la génétique, la biologie moléculaire, les réseaux sociaux et bien d'autres encore peuvent être naturellement modélisés sous la forme d'un graphe. Ainsi les personnes, les atomes, les protéines, les zones du cerveau ou les villes sont considérées comme des nœuds et les interactions fonctionnelles ou les connexions structurales sont représentées par des arêtes. L'un des principaux intérêts liés à l'analyse de ces structures est la mesure de la similarité entre deux graphes à des fins de classification. Lors d'un travail de recherche antérieur [1], nous avons développé deux mesures de similarité spectrale entre graphes, les avons insérées dans un noyau de Séparateur à Vaste Marge (SVM) et avons montré leurs intérêts pour la classification de graphes. Les premiers noyaux de graphes ont été présentés par Gärtner *et al.* [2] et n'ont cessé d'être enrichis. Deux approches sont à distinguer. L'une d'elle correspond à l'utilisation de noyaux se basant sur l'analyse d'invariants propres aux graphes comme le nombre de sous-graphes type (*graphlet*) contenus dans la structure de ce graphe [3], le nombre de chemins apparaissant au sein de ce graphe [2] ou la recherche du plus court chemin [4]. La deuxième approche repose sur des matrices de Gram construites à partir de distances et intégrées dans un noyau de type Gaussien. La Covariance Spectrale (CS) et la Similarité Spectrale Conjointe (SSC) que nous avons élaborées sont insérées dans un tel noyau et utilisent conjointement les propriétés énergétiques et structurelles des graphes à travers le spectre de leurs matrices de représentation. Les résultats obtenus, incorporant à la fois les informations spectrales de la matrice d'adjacence et laplacienne, ont confirmé la présence d'informations pertinentes dans ces spectres pour la classifica-

tion de graphes [1] et ont été repris dans des travaux comme la recherche de programmes clones en informatique [5]. Nous élaborons une nouvelle stratégie mettant en avant les informations spectrales permettant de discriminer deux graphes tout en palliant le coût en temps de calcul de la recherche des spectres des deux matrices. Ainsi, nous proposons de traiter le problème de classification de graphes en introduisant une mesure de similarité spectrale définie à partir de la corrélation entre les spectres d'une nouvelle matrice de représentation $\mathbf{T}_\alpha$, combinaison linéaire de la matrice d'adjacence et de la matrice laplacienne pondérée par un paramètre $\alpha \in [0, 1]$. Cette mesure, intégrée dans un noyau SVM, est exploitée à des fins de classification de graphes sur des bases de données connues de la littérature issues de la bio-informatique (MUTAG, PTC-MR, PROTEINS) ainsi que de réseaux sociaux (IMDB-BINARY, IMDB-MULTI). Nous testons également la capacité de cette méthode à détecter la présence ou non d'un drone sous-marin à partir de sa signature sous forme d'une série temporelle. L'analyse des résultats obtenus tend à confirmer l'efficacité de cette nouvelle mesure pour la classification de graphes et de signaux transcrits en graphes.

2 Définition de la matrice $\mathbf{T}_\alpha$

Un graphe $G = (\mathcal{V}, \mathcal{E})$ est défini à partir d'un ensemble $\mathcal{V}$ de n sommets et d'un ensemble $\mathcal{E}$ de m arêtes. La matrice de représentation usuelle est la matrice d'adjacence $\mathbf{A}$, de taille $n \times n$, où le coefficient $[\mathbf{A}]_{ij}$ est égal à 1 si une arête relie les sommets i et j , et valant 0 sinon. Les graphes considérés ici sont simples (aucun sommet n'est relié à lui-même et aucune arête multiple n'existe) et non orientés (une arête reliant un sommet i et j implique une arête reliant j à i). Pour un tel graphe, la matrice d'adjacence est symétrique et à diagonale nulle. Le degré d'un sommet i , noté $\deg(i)$, correspond au nombre d'arêtes lui étant incidentes. La matrice $\mathbf{D}$ des degrés

du graphe G peut alors être construite en plaçant sur sa diagonale les degrés des sommets. Cette matrice permet d'introduire la matrice laplacienne $\mathbf{L}$ du graphe G comme suit [6] :

$$\mathbf{L} := \mathbf{D} - \mathbf{A} \quad (1)$$

Cette matrice dispose de meilleures propriétés algébriques et spectrales que la matrice d'adjacence. En effet, elle est semi-définie positive, s'apparente à un opérateur de diffusion et le nombre de composantes connectées du graphe est égal à la multiplicité de la valeur propre 0. Une dernière matrice de représentation d'un graphe G utilisée récemment dans [7] est la matrice laplacienne sans signe $\mathbf{Q}$ définie comme la somme de sa matrice des degrés et sa matrice d'adjacence [8] :

$$\mathbf{Q} := \mathbf{D} + \mathbf{A} \quad (2)$$

Cette matrice est aussi semi-définie positive, permet de répondre au problème de la cospectralité entre graphes (deux graphes sont dits $\mathbf{M}$ -cospectraux lorsque les spectres de leurs matrices de représentation $\mathbf{M}$ sont identiques) et la multiplicité de la valeur propre nulle est cette fois égale au nombre de composantes bipartites [9]. Cependant, même si toutes ces matrices ont des propriétés bien spécifiques et sont utiles à bien des égards, il peut être intéressant de proposer une matrice combinant linéairement les matrices d'adjacence et des degrés qui, à l'aide d'un paramètre, permet d'évaluer graduellement l'importance de ces deux matrices, notamment pour le contenu spectral. Ainsi, deux matrices sont proposées : la matrice d'adjacence généralisée $\mathbf{G} = \alpha\mathbf{A} + \beta\mathbf{I} + \gamma\mathbf{J}$ et la matrice d'adjacence universelle $\mathbf{U} = \alpha\mathbf{A} + \beta\mathbf{I} + \gamma\mathbf{J} + \delta\mathbf{D}$ (avec $\alpha, \beta, \gamma, \delta \in \mathbb{R}, \alpha \neq 0, \mathbf{I}$ la matrice identité et $\mathbf{J}$ la matrice formée de un) [10]. Quelques cas particuliers, sous-ensembles de ces généralisations, ont été définis. Par exemple, il existe la matrice laplacienne déformée [11] :

$$\Delta_\alpha := \alpha^2(\mathbf{D} - \mathbf{I}) - \alpha\mathbf{A} + \mathbf{I}, \quad \alpha \in \mathbb{R} \quad (3)$$

Nikiforov présente, dans [12], la matrice d' α -adjacence reposant sur une combinaison convexe linéaire de $\mathbf{D}$ et $\mathbf{A}$:

$$\mathbf{A}_\alpha := \alpha\mathbf{D} + (1 - \alpha)\mathbf{A}, \quad 0 \leq \alpha \leq 1 \quad (4)$$

Cette matrice possède des propriétés intéressantes mais ne fait pas apparaître la matrice $\mathbf{L}$. C'est pourquoi la matrice « α -laplacienne » [13] est introduite :

$$\mathbf{L}_\alpha := \alpha\mathbf{D} + (\alpha - 1)\mathbf{A}, \quad 0 \leq \alpha \leq 1 \quad (5)$$

Cette matrice a l'inconvénient de faire apparaître $-\mathbf{A}$, pour laquelle les N premières valeurs propres (exploitées ultérieurement) ne sont pas les mêmes que celles de $\mathbf{A}$, ce qui est problématique pour l'étude spectrale menée par la suite.

Largement inspirés de ces dernières, nous introduisons une nouvelle matrice de représentation d'un graphe, notée $\mathbf{T}_\alpha$, pour répondre aux attentes évoquées précédemment :

$$\mathbf{T}_\alpha := \alpha\mathbf{D} + (1 - 2\alpha)\mathbf{A}, \quad 0 \leq \alpha \leq 1 \quad (6)$$

Cette matrice vérifie $\mathbf{T}_0 = \mathbf{A}$, $\mathbf{T}_1 = \mathbf{L}$ et $2\mathbf{T}_{1/2} = \mathbf{D}$. La positivité de cette matrice fait l'objet du lemme suivant.

Lemme. La matrice $\mathbf{T}_\alpha$ est semi-définie positive si $\alpha \in [\frac{1}{2}, 1]$.

Preuve. Soit $\mathbf{x} \in \mathbb{R}^n$ où n est l'ordre du graphe étudié. On a

$$\begin{aligned} \langle \mathbf{T}_\alpha \mathbf{x}, \mathbf{x} \rangle &= \langle [\alpha\mathbf{D} + (1 - 2\alpha)\mathbf{A}] \mathbf{x}, \mathbf{x} \rangle \\ &= \langle [\alpha\mathbf{D} + (1 - 2\alpha)(\mathbf{D} - \mathbf{L})] \mathbf{x}, \mathbf{x} \rangle \\ &= \langle [(1 + \alpha)\mathbf{D} + (2\alpha - 1)\mathbf{L}] \mathbf{x}, \mathbf{x} \rangle \geq 0 \end{aligned}$$

si $\alpha \in [\frac{1}{2}, 1]$ car $\langle \mathbf{D}\mathbf{x}, \mathbf{x} \rangle \geq 0$ et $\langle \mathbf{L}\mathbf{x}, \mathbf{x} \rangle \geq 0$. $\square$

Attention la réciproque n'est pas vraie : selon les graphes, il semble exister un $\alpha_0 < \frac{1}{2}$ tel que $\mathbf{T}_\alpha$ est semi-définie positive pour $\alpha \geq \alpha_0$. Par ailleurs, nous proposons une expression de l'énergie d'un graphe semblable à celles définies par Gutan [14] et basée sur les valeurs propres $(\nu_\ell)_{1 \leq \ell \leq n}$ de $\mathbf{T}_\alpha$:

$$E(\mathbf{T}_\alpha) = \sum_{\ell=1}^n \left| \nu_\ell - \frac{2\alpha m}{n} \right| = \sum_{\ell=1}^n |\nu_\ell - \alpha \bar{d}| \quad (7)$$

avec $\bar{d} = 2m/n$ le degré moyen du graphe. Pour $\alpha = 0$, nous retrouvons l'énergie d'un graphe et pour $\alpha = 1$, c'est l'énergie laplacienne. Pour $\alpha = 1/2$, nous obtenons la moitié de la somme des écarts au degré moyen comme dans [15]. Une conjecture concernant l'énergie de cette matrice est établie.

Conjecture. L'énergie $E(\mathbf{T}_\alpha)$ est décroissante pour $0 \leq \alpha \leq \alpha_{\min} \leq 1/2$ puis croissante pour $\alpha_{\min} \leq \alpha \leq 1$. Elle atteint son maximum en $\alpha_{\max} = 1$.

3 Similarité spectrale $\text{CorS}_{\mathbf{T}_\alpha}$

Dans un travail précédent qui avait pour objectif de proposer une nouvelle méthode de classification spectrale de graphes [1], une mesure de similarité appelée SSC_β a été introduite pour remplacer la distance euclidienne du traditionnel noyau Gaussien d'un SVM. Cette mesure de similarité entre deux graphes $G_1 = (\mathcal{V}_1, \mathcal{E}_1)$ et $G_2 = (\mathcal{V}_2, \mathcal{E}_2)$ est définie comme une combinaison convexe linéaire de paramètre $0 \leq \beta \leq 1$ d'une « distance euclidienne » calculée entre les spectres des matrices d'adjacence et entre ceux des matrices laplacienne :

$$\text{SSC}_\beta(G_1, G_2) = \beta \text{SSL}(G_1, G_2) + (1 - \beta) \text{SSA}(G_1, G_2) \quad (8)$$

avec

$$\text{SSA}(G_1, G_2) = \sum_{\ell=1}^N (\lambda_{1\ell} - \lambda_{2\ell})^2 \quad (9)$$

$$\text{SSL}(G_1, G_2) = \sum_{\ell=1}^N (\mu_{1\ell} - \mu_{2\ell})^2 \quad (10)$$

où $N = \min(n_1, n_2)$ avec n_1 (resp. n_2) le nombre de sommets du graphe G_1 (resp. G_2), où $(\lambda_{1\ell})_{1 \leq \ell \leq n_1}$ (resp. $(\lambda_{2\ell})_{1 \leq \ell \leq n_2}$) représente le spectre de la matrice d'adjacence de G_1 (resp. G_2) et où $(\mu_{1\ell})_{1 \leq \ell \leq n_1}$ (resp. $(\mu_{2\ell})_{1 \leq \ell \leq n_2}$) représente le spectre de la matrice laplacienne de G_1 (resp. G_2). Le terme distance ne peut être employé ici car la propriété de séparation n'est pas vérifiée. En effet, $\text{SSA}(G_1, G_2) = 0$ n'implique pas $G_1 = G_2$ car il est connu qu'il existe des graphes dits cospectraux qui, bien qu'ayant des structures différentes, possèdent les mêmes spectres. Quelques points d'attention peuvent être mis en avant. Le premier est que la SSC_β est calculée entre les deux spectres (d'adjacence et laplacien) et non pas entre des spectres issus de matrices variant entre $\mathbf{A}$ et $\mathbf{L}$. De plus, cette mesure nécessite le calcul de deux spectres qui, pour de grands graphes, est coûteux en temps de calcul. Le deuxième est que la SSC_β utilisée pour comparer les spectres calcule une différence locale. Cette dernière, par définition, peut omettre beaucoup de valeurs propres (en fonction du nombre de sommets des graphes comparés). Aussi, l'objectif est de trouver une mesure qui ne requiert qu'un spectre et qui puisse comparer les spectres de manière globale. Pour cela, une mesure de similarité spectrale basée sur la corrélation entre les spectres des matrices $\mathbf{T}_\alpha$ des graphes étudiés ($\text{CorS}_{\mathbf{T}_\alpha}$) est introduite.

	Nombre de graphes	Classes	$\bar{n}$	$\bar{m}$
MUTAG	188	2	17.93	19.79
PTC-MR	344	2	14.29	14.69
PROTEINS	1113	2	39.06	72.82
IMDB-B	1000	2	19.77	96.53
IMDB-M	1500	3	13	65.94

TABLEAU 1 : Statistiques des bases de données utilisées ($\bar{n}$ le nombre moyen de sommets et $\bar{m}$ le nombre moyen d'arêtes).

La corrélation ne pouvant être calculée qu'entre des vecteurs de même taille, nous opérons une simple complétion par des nœuds (*node-padding*) qui, bien que n'ayant pas de sens physique, permet à tous les graphes d'être du même ordre. Soient deux graphes G_1 et G_2 d'ordre n . Si $\tilde{\nu}_1^{(\alpha)}$ (resp. $\tilde{\nu}_2^{(\alpha)}$) représente le spectre standardisé (de moyenne nulle et d'écart-type égal à 1) de la matrice $\mathbf{T}_\alpha(G_1)$ (resp. $\mathbf{T}_\alpha(G_2)$), alors nous définissons une mesure de similarité entre les graphes G_1 et G_2 basée sur une distance construite à partir de la corrélation de Pearson comme suit [16] :

$$\text{CorS}_{\mathbf{T}_\alpha}(G_1, G_2) := \sqrt{1 - \left(\frac{1}{n} \langle \tilde{\nu}_1^{(\alpha)}, \tilde{\nu}_2^{(\alpha)} \rangle \right)^2} \quad (11)$$

où $\langle \cdot, \cdot \rangle$ désigne le produit scalaire.

À l'image du travail précédent [1], cette mesure est intégrée dans un SVM en remplacement de la distance euclidienne dans le traditionnel noyau Gaussien. Ainsi, la matrice de Gram $\mathbf{K}$ est définie par ses coefficients

$$[\mathbf{K}]_{ij} = \exp(-\gamma \text{CorS}_{\mathbf{T}_\alpha}(G_i, G_j)) \quad (12)$$

Le paramètre de normalisation γ du noyau comme le paramètre de régularisation C de la SVM sont fixés à 1, et une validation croisée à 10 couches est effectuée dans le but de pouvoir comparer nos résultats avec ceux de la littérature. Les graphes utilisés sont issus de bases de données bio-informatiques (MUTAG, PTC-MR, PROTEINS) et de réseaux sociaux (IMDB-BINARY, IMDB-MULTI) classiquement utilisées dont les statistiques sont rappelées dans le Tableau 1.

Les résultats de classification sont présentés en Tableau 2 : les astérisques signifient que les résultats sont directement retranscrits à partir de l'article de Ma *et al.* [18] qui utilisent la même méthode de validation croisée que ce travail. Pour chaque base de données, les résultats des deux meilleurs noyaux sont affichés en gras. Deux points sont marquants. Le

premier est que seuls les graphes de PROTEINS sont ceux pour lesquels le noyau basé sur la mesure $\text{CorS}_{\mathbf{T}_\alpha}$ n'est pas parmi les deux meilleurs. Le deuxième est que la méthode SCNN basée sur des réseaux de neurones est la meilleure dans 4 cas sur 5. Il faut alors préciser que la méthode introduite dans ce travail n'a pas vocation à dépasser les résultats qui peuvent être obtenus avec ces nouvelles méthodes d'apprentissage profond mais se veut plus interprétable et moins coûteuse en paramètres d'entrée et en temps d'entraînement ce qui semble intuitif bien qu'une analyse de cet aspect soit absente dans [18]. Par ailleurs, la méthode développée doit être plus performante que les noyaux structurels, ce qui est vérifié. En effet, elle améliore la précision de 2.5 % pour MUTAG par rapport au plus performant d'entre eux, de 1 % pour PTC-MR, de 1 % pour IMDB-BINARY. Pour IMDB-MULTI, le noyau $\text{CorS}_{\mathbf{T}_\alpha}$ obtient des résultats similaires au noyau Weisfeiler-Lehman pour un temps de calcul qui devrait être beaucoup plus faible. Bien qu'une étude comparative des temps de calcul ne soit pas présentée dans ce travail, les temps nécessaires aux calculs de tous les noyaux pour la base de données PROTEINS sont listés dans la documentation de la librairie Grakel utilisée pour ce travail : il faut plus de 51 minutes pour le noyau RW, 7 minutes pour le noyau SP, 32 secondes pour le noyau WL et le noyau GK a quant à lui saturé la mémoire. À titre de comparaison, avec un ordinateur moins performant, le noyau SSC_β nécessite 28 secondes et le noyau $\text{CorS}_{\mathbf{T}_\alpha}$ requiert 14 secondes.

Les résultats obtenus avec le noyau SSC_α , le noyau CorS développé à la fois à partir de la matrice $\mathbf{A}_\alpha$ introduite par Nikiforov ($\text{CorS}_{\mathbf{A}_\alpha}$) et la matrice $\mathbf{T}_\alpha$ proposé précédemment ($\text{CorS}_{\mathbf{T}_\alpha}$) sont présentés en Figure 1. Il apparaît que le noyau $\text{CorS}_{\mathbf{T}_\alpha}$ est plus performant que la SSC_α pour tout α et toutes les bases de données. Les taux de classification sont comparables pour les matrices $\mathbf{A}_\alpha$ et $\mathbf{T}_\alpha$, cette dernière permettant toutefois une évolution graduelle entre la matrice $\mathbf{A}$ et la matrice $\mathbf{L}$. Cela nous conforte dans l'idée qu'une mesure qui compare de manière globale les spectres semble plus judicieuse qu'une mesure de différence plus locale. Par notre approche paramétrique, il est tout à fait possible de mettre en évidence l'influence de la matrice de représentation dans la classification des différents graphes. Ainsi, pour la base de données MUTAG, la matrice des degrés ($\alpha = 1/2$) a un meilleur taux de classification tandis que pour PTC-MR, l'inverse semble

Noyau	MUTAG	PTC-MR	PROTEINS	IMDB-BINARY	IMDB-MULTI
SP [4]	85.79 ± 2.51 (*)	58.24 ± 2.44 (*)	75.07 ± 0.54 (*)	71.26 ± 1.04 (*)	– (*)
RW [2]	83.72 ± 1.50 (*)	57.85 ± 1.30 (*)	74.22 ± 0.42 (*)	64.54 ± 1.22 (*)	34.54 ± 0.76 (*)
GK [3]	81.58 ± 2.11 (*)	57.26 ± 1.41 (*)	71.67 ± 0.55 (*)	65.87 ± 0.98 (*)	43.89 ± 0.38 (*)
WL [17]	82.88 ± 0.57 (*)	– (*)	73.52 ± 0.43 (*)	71.88 ± 0.77 (*)	49.50 ± 0.49 (*)
SCNN [18]	90.08 ± 2.01 (*)	63.47 ± 2.65 (*)	76.51 ± 1.37 (*)	73.24 ± 0.96 (*)	46.82 ± 1.83 (*)
SSC_β [1]	84.32 ± 1.10	57.83 ± 2.08	69.99 ± 0.39	67.46 ± 1.01	48.19 ± 0.19
$\text{CorS}_{\mathbf{A}_\alpha}$	88.20 ± 0.37	59.22 ± 0.59	74.48 ± 0.17	71.24 ± 0.58	49.85 ± 0.5
$\text{CorS}_{\mathbf{T}_\alpha}$	88.20 ± 0.37	59.22 ± 0.61	74.48 ± 0.17	72.25 ± 0.55	49.69 ± 0.35

TABLEAU 2 : Résultats de classification (précisions en % moyennées sur 10 itérations ± écarts-types) de différents noyaux (Shortest Path (SP), Random Walk (RW), Graphlet (GK), Weisfeiler-Lehman (WL), Spatial Convolutional Neural Network (SCNN)) sur des bases de données de la littérature. Les résultats suivis d'une astérisque (*) sont directement issus de l'article de Ma *et al.* [18].

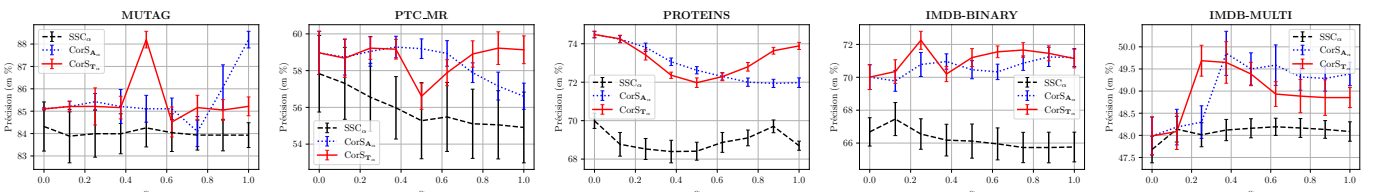


FIGURE 1 : Résultats de classification (précision en %) en fonction de α entre les noyaux SSC_α et CorS_α sur les bases de données de la littérature.

se produire. En effet, la courbe pleine (en rouge) décroît brusquement lorsque α est égal à $1/2$. Pour IMDB-BINARY et IMDB-MULTI, les meilleures précisions sont obtenues pour $\alpha = 0.25$. Sur ces premiers résultats et compte tenu des bases de données exploitées, aucune matrice de représentation n'est à privilégier. Une étude approfondie doit être effectuée pour intuitiver, *a priori*, la valeur α optimale.

4 Mise en œuvre sur des signaux

Nous illustrons également notre mesure spectrale à des fins de classification de séries temporelles issus de drones sous-marins. Les signaux utilisés, dont une illustration est fournie en Figure 2, sont échantillonnés à 5 Hz et ont une durée de 1 minute. Cette classification est binaire : signal bruité ou signal bruité avec signature. L'objectif est de savoir s'il y a une signature dans le signal ou non. Pour cela, les signaux sont transformés en graphe de visibilité naturelle [19] puis les corrélations spectrales sont calculées à partir des matrices $\mathbf{T}_\alpha$ de ces graphes. Les courbes décrivant l'évolution de l'exactitude, de la sensibilité et de la spécificité en fonction de α sont affichées en Figure 3. Ces résultats sont issus de la moyenne de 10 itérations en effectuant une nouvelle fois une validation croisée à 10 couches. Les résultats de classification sont très encourageants avec des mesures d'exactitude dépassant 99.5 %. Pour une telle application, il est impératif de minimiser les faux négatifs, c'est-à-dire les signaux qui seraient classés comme composés uniquement de bruit alors qu'il y a bien une signature. Un intérêt particulier doit alors être porté sur la sensibilité. Cette grandeur possède des valeurs très satisfaisantes lorsque $\alpha = 1$ qui correspond au maximum de l'exactitude atteint lorsque $\mathbf{T}_\alpha = \mathbf{L}$. La mesure paramétrique développée permet de se rendre compte que ce sont toutefois les spectres des matrices d'adjacence des graphes de visibilité qui sont les plus prompts à contenir l'information nécessaire à la classification.

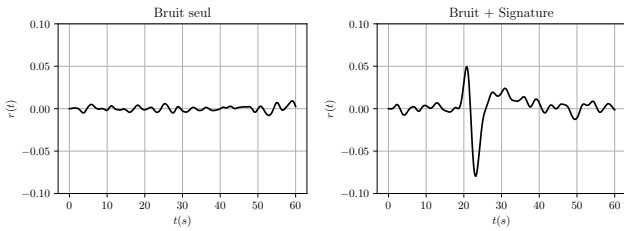


FIGURE 2 : Exemple de séries temporelles étudiées. À gauche : bruit seul / À droite : bruit avec une signature de drones sous-marins.

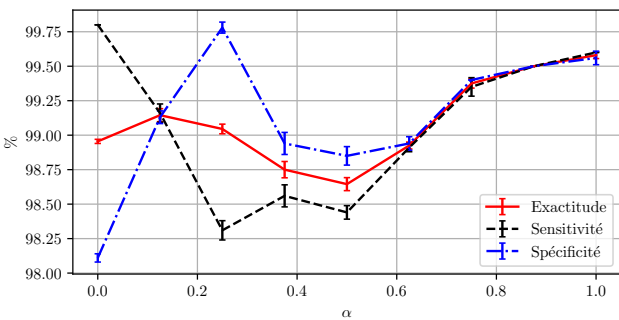


FIGURE 3 : Évolution de l'exactitude, de la sensibilité et de la spécificité en fonction de α pour la détection de drones sous-marins dans des séries temporelles.

5 Conclusion

Dans cet article, nous avons proposé une nouvelle méthode de classification spectrale de graphes reposant sur une mesure de similarité établie à partir de la corrélation entre les spectres des matrices $\mathbf{T}_\alpha := \alpha \mathbf{D} + (1 - 2\alpha) \mathbf{A}$ (avec $0 \leq \alpha \leq 1$) représentatives des graphes étudiés. Cette mesure est associée à un noyau de SVM et des tests ont été menés à la fois sur des bases de données de graphes connues de la littérature mais aussi sur des séries temporelles contenant des signaux réels associés à des drones sous-marins. Les résultats obtenus montrent la pertinence et l'efficacité de la méthode avec un faible taux d'erreur de classification et des temps de calculs réduits. Des propriétés de la matrice $\mathbf{T}_\alpha$ doivent encore être établies dans le but de poursuivre la construction du cadre formel autour de cette matrice, à l'instar du travail de Nikiforov avec $\mathbf{A}_\alpha$. Parmi celles-ci, nous pouvons citer la recherche de la valeur α optimale ou l'extraction d'attributs spectraux pour une classification de graphes et/ou de signaux sur graphes.

Références

- [1] H. A. BAY-AHMED *et al.*, « Classification des signaux sur graphes par mesures spectrales algébriques », *GRETSI*, p. 1–4, 2017.
- [2] T. GÄRTNER *et al.*, « On graph kernels : Hardness results and efficient alternatives », in *Learning Theory and Kernel Machines : 16th Annual Conference on Learning Theory and 7th Kernel Workshop*, p. 129–143, Springer, 2003.
- [3] N. SHERVASHIDZE *et al.*, « Efficient graphlet kernels for large graph comparison », in *Artificial intelligence and statistics*, p. 488–495, PMLR, 2009.
- [4] K. M. BORGWARDT et H.-P. KRIEGLER, « Shortest-path kernels on graphs », in *Fifth IEEE Int. Conf. Data Mining*, p. 1–8, 2005.
- [5] T. BENOIT *et al.*, « Scalable program clone search through spectral analysis », *arXiv preprint arXiv :2210.13063*, 2022.
- [6] F. CHUNG, *Spectral graph theory*, vol. 92. AMS, 1997.
- [7] J. B. LIU *et al.*, « On the generalized adjacency, Laplacian and signless Laplacian spectra of the weighted edge corona networks », *Physica A : Statistical Mechanics and Its Applications*, vol. 540, p. 1–11, 2020.
- [8] E. R. VAN DAM et W. H. HAEMERS, « Which graphs are determined by their spectrum? », *Linear Algebra Appl.*, vol. 373, p. 241–272, 2003.
- [9] D. CVETKOVIĆ *et al.*, « Signless laplacians of finite graphs », *Linear Algebra Appl.*, vol. 423, no. 1, p. 155–171, 2007.
- [10] W. H. HAEMERS et G. R. OMIDI, « Universal adjacency matrices with two eigenvalues », *Linear Algebra Appl.*, vol. 435, no. 10, p. 2520–2529, 2011.
- [11] L. DALL'AMICO *et al.*, « Classification spectrale par la laplacienne déformée dans des graphes réalistes », *GRETSI*, p. 1–4, 2019.
- [12] V. NIKIFOROV, « Merging the $\mathbf{A}$ - and $\mathbf{Q}$ -spectral theories », *Applicable Analysis and Discrete Mathematics*, vol. 11, no. 1, p. 81–107, 2017.
- [13] S. WANG *et al.*, « Bounds for the largest and the smallest $\mathbf{A}_\alpha$ eigenvalues of a graph in terms of vertex degrees », *Linear Algebra and its Applications*, vol. 590, p. 210–223, 2020.
- [14] I. GUTMAN et B. FURTULA, « Survey of graph energies », *Mathematics Interdisciplinary Research*, vol. 2, no. 2, p. 85–129, 2017.
- [15] V. NIKIFOROV, « Eigenvalues and degree deviation in graphs », *Linear Algebra and its Applications*, vol. 414, no. 1, p. 347–360, 2006.
- [16] S. VAN DONGEN et A. J. ENRIGHT, « Metric distances derived from cosine similarity and Pearson and Spearman correlations », *arXiv preprint arXiv :1208.3145*, 2012.
- [17] N. SHERVASHIDZE *et al.*, « Weisfeiler-Lehman graph kernels », *Journal of Machine Learning Research*, vol. 12, no. 9, p. 2539–2561, 2011.
- [18] T. MA *et al.*, « Graph classification based on structural features of significant nodes and spatial convolutional neural networks », *Neurocomputing*, vol. 423, p. 639–650, 2021.
- [19] L. LACASA *et al.*, « From time series to complex networks : The visibility graph », *PNAS*, vol. 105, no. 13, p. 4972–4975, 2008.