

Sur la similarité spectrale des graphes par mesure de corrélation

AER Tristan AVERTY

MCF Delphine DARÉ-EMZIVAT

PU Abdel-Ouahab BOUDRAA

IRENav (Institut de Recherche de l'École Navale)

IRENav & ENSAM (École nationale supérieure d'Arts et Métiers)

IRENav & ENSAM

XXIX^{ème} Colloque Francophone de Traitement du Signal et des Images (Grenoble)



Définition

Un **graphe simple non-orienté** $G = (\mathcal{V}, \mathcal{E})$ est constitué :

- ▶ d'un ensemble \mathcal{V} de n **sommets** ;
- ▶ d'un ensemble $\mathcal{E} \subseteq \{\{v_i, v_j\} \mid (v_i, v_j) \in \mathcal{V}^2, v_i \neq v_j\}$ de m **arêtes**.
- ▶ d'un ensemble de valeurs w_{ij} appelées **poids** qui pondèrent les arêtes (i, j) si le graphe est **pondéré**.

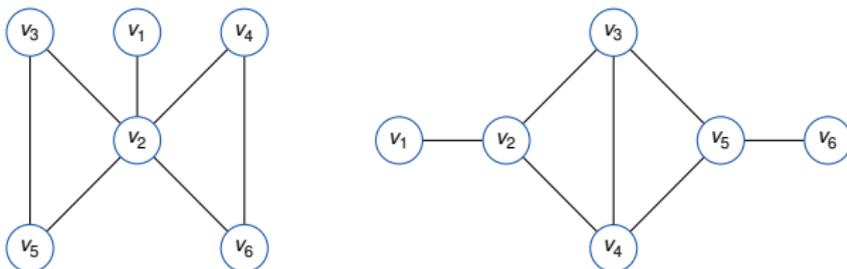


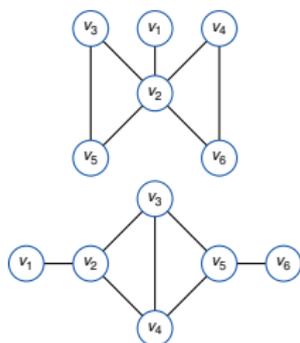
Figure – Deux graphes simples non-orientés et non-pondérés (6 sommets / 7 arêtes)

Deux graphes bien différents, n'est-ce pas ?

Matrices de représentation pour la classification de graphes

Matrices de représentation de la littérature

- Matrice d'adjacence \mathbf{A} où les coefficients \mathbf{A}_{ij} désigne s'il existe une arête (i, j) ou pas ;
- Matrice des degrés \mathbf{D} . Le degré d'un sommet est son nombre d'arêtes incidentes ;
- Matrice laplacienne¹ $\mathbf{L} := \mathbf{D} - \mathbf{A}$;
 - Matrice laplacienne normalisée¹ $\mathbf{L}_N := \mathbf{D}^{-1/2} \mathbf{L} \mathbf{D}^{-1/2} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$;
 - Matrice laplacienne sans-signe² $\mathbf{Q} := \mathbf{D} + \mathbf{A}$;
 - Matrice de densité³ $\rho := \mathbf{L} / \text{Tr}(\mathbf{L})$.



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 5 & -1 & -1 & -1 & -1 \\ 0 & -1 & 2 & 0 & -1 & 0 \\ 0 & -1 & 0 & 2 & 0 & -1 \\ 0 & -1 & -1 & 0 & 2 & 0 \\ 0 & -1 & 0 & -1 & 0 & 2 \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 & 0 \\ 0 & -1 & 3 & -1 & -1 & 0 \\ 0 & -1 & -1 & 3 & -1 & 0 \\ 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

1. F.R. Chung, *Spectral graph theory*, American Mathematical Society, 1997

2. D. Cvetković et al., *Signless Laplacians of finite graphs*. Linear Algebra and its applications, 2007

3. S.L. Braunstein et al., *The Laplacian of a graph as a density matrix : a basic combinatorial approach to separability of mixed states*, Annals of Combinatorics, 2006

Classification de graphes

Exemples d'applications :

- 1 Classification de **graphes naturels**
 - Déterminer si une **molécule** est **cancérogène ou non**.
- 2 Classification de **séries temporelles** (vues comme des graphes)
 - Déterminer si un **signal EEG** provient d'un **patient sain ou épileptique**⁴.
 - Déterminer si un **signal MAD** contient une **signature magnétique ou non**.

Méthode :

- Création une **mesure de similarité** entre graphes.
- Mise sous forme d'une **fonction noyau** de cette mesure de similarité ;
- Choix d'un **algorithme de classification** (SVM par exemple) et entraînement à partir de ce noyau.

4. T. Averty et al., *Détection d'épilepsie dans les signaux EEG par graphe de visibilité et un noyau de SVM adapté*, GRETSI, 2022

Deux méthodes principales pour classifier des graphes

1 Classification structurelle : comparaison d'attributs structurels.

- Comparaison des **plus courts chemins**⁵ (SP pour *Shortest Path*);
- Comparaison des **marches aléatoires**⁶ (RW pour *Random Walk*);
- Décompte des **sous-graphes**⁷ (GK pour *Graphlet Count*);
- Test d'isomorphisme de **Weisfeiler-Lehman**⁸ (WL).

Noyaux structurels performants mais souvent très coûteux en calcul.

2 Classification spectrale : comparaison d'attributs spectraux extraits de matrices de représentation.

- **Deux écoles** : celle de la matrice d'adjacence **A** et celle de la matrice laplacienne **L**.
- Volonté de trouver une matrice qui effectue un **balayage** entre ces deux matrices.
- Des propriétés sur les spectres permettent de retrouver des **éléments structurels**.

Recherche d'une mesure de similarité spectrale associée à une nouvelle matrice de représentation permettant une classification spectrale de graphes

5. K.M. Borgwardt et H.P. Kriegel, *Shortest-path kernels on graphs*, 5th IEEE International Conference Data Mining, 2005

6. T. Gärtner et al., *On graph kernels : Hardness results and efficient alternatives*, 16th Annual Conference on Learning Theory and 7th Kernel Workshop, 2003

7. N. Shervashidze et al., *Efficient graphlet kernels for large graph comparison*, Artificial Intelligence and Statistics, 2009

8. N. Shervashidze et al., *Weisfeiler-Lehman graph kernels*, Journal of Machine Learning Research, 2011

Classification spectrale : Théorie spectrale & cospectralité

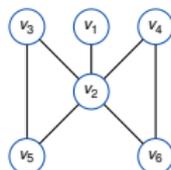
- ▶ Soient les valeurs propres de la matrice \mathbf{A} notées $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$;
- ▶ Soient les valeurs propres de la matrice \mathbf{L} notées $0 = \mu_1 \leq \mu_2 \leq \dots \leq \mu_n$;

Propriétés :

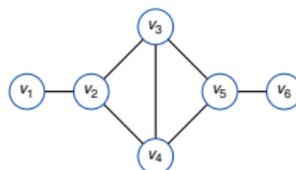
- 📎 Le **rayon spectral** λ_n peut être interprété comme une mesure de **vulnérabilité globale**⁹ (*shield value*);
- 📎 La multiplicité de la valeur propre 0 de \mathbf{L} est le **nombre de composantes connectées** du graphe;
- 📎 La **valeur de Fiedler** μ_2 donne une information sur la **connectivité**;
- 📎 Si m désigne le nombre d'arêtes et T le nombre de triangles dans le graphe, alors

$$\sum_{\ell=1}^n \lambda_{\ell} = 0, \quad \sum_{\ell=1}^n \lambda_{\ell}^2 = 2m, \quad \sum_{\ell=1}^n \lambda_{\ell}^3 = 6T, \quad \sum_{\ell=1}^n \mu_{\ell} = 2m, \quad \sum_{\ell=1}^n \mu_{\ell}^2 = 2m + \sum_{v_i \in \mathcal{V}} \deg(v_i)^2$$

Idee : Calcul d'une distance entre les spectres pour comparer une multitude d'aspects.



$$\begin{aligned} \text{spec}(\mathbf{A}) &= \{-1.9, -1, -1, 0.2, 1, 2.7\} \\ \text{spec}(\mathbf{L}) &= \{0, 1, 1, 3, 3, 6\} \end{aligned}$$



$$\begin{aligned} \text{spec}(\mathbf{A}) &= \{-1.9, -1, -1, 0.2, 1, 2.7\} \\ \text{spec}(\mathbf{L}) &= \{0, 0.5, 1.2, 3.4, 4, 4.7\} \end{aligned}$$

Classification spectrale : Similarité spectrale conjointe

Similarité spectrale conjointe¹⁰

Soient deux graphes G_1 et G_2 d'ordre n_1 et n_2 . La **similarité spectrale conjointe** (SSC) est définie par :

$$SSC_{\beta}(G_1, G_2) = \beta \sum_{\ell=1}^{\min(n_1, n_2)} (\mu_{1\ell} - \mu_{2\ell})^2 + (1 - \beta) \sum_{\ell=1}^{\min(n_1, n_2)} (\lambda_{1\ell} - \lambda_{2\ell})^2$$

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).

Forces :

- Cette mesure tient compte du problème de cospectralité entre graphes.
- $\beta = 1$: comparaison des spectres laplacien / $\beta = 0$: comparaison des spectres d'adjacence.

Limitations :

- Il est nécessaire de calculer **deux spectres**.
- Selon les ordres n_1 et n_2 , cette mesure omet un certain nombre de valeurs propres.
 - Méthode type *Dynamic Time Warping*
 - Complétion par des nœuds (*node-padding*)

Idee : Définir une mesure de similarité spectrale à partir d'un unique calcul de spectre

10. H.A. Bay-Ahmed et al., *A joint spectral similarity measure for graphs classification*, Pattern Recognition Letters, 2019

Matrices d'adjacence généralisées

Unique calcul de spectre \longrightarrow Unique matrice \longrightarrow Matrice qui « balaie » les représentations classiques

Matrices d'adjacence généralisées

- ▶ Matrice d'adjacence généralisée¹¹ $\mathbf{G} := \alpha\mathbf{A} + \beta\mathbf{I} + \gamma\mathbf{J}$
✗ **Problème** : 3 paramètres.
- ▶ Matrice d'adjacence universelle¹¹ $\mathbf{U} := \alpha\mathbf{A} + \beta\mathbf{I} + \gamma\mathbf{J} + \delta\mathbf{D}$ (avec $\alpha, \beta, \gamma, \delta \in \mathbb{R}, \alpha \neq 0$)
✗ **Problème** : 4 paramètres.
- ▶ Matrice d' α -adjacence¹² $\mathbf{A}_\alpha := \alpha\mathbf{D} + (1 - \alpha)\mathbf{A}$ (avec $0 \leq \alpha \leq 1$)
✗ **Problème** : Ne fait pas apparaître \mathbf{L} mais \mathbf{Q} .
- ▶ Matrice α -laplacienne¹³ $\mathbf{L}_\alpha := \alpha\mathbf{D} + (\alpha - 1)\mathbf{A}$ (avec $0 \leq \alpha \leq 1$).
✗ **Problème** : Ne fait pas apparaître \mathbf{A} mais $-\mathbf{A}$.

11. W.H. Haemers et G.R Omid, *Universal adjacency matrices with two eigenvalues*, Linear Algebra Appl., 2011

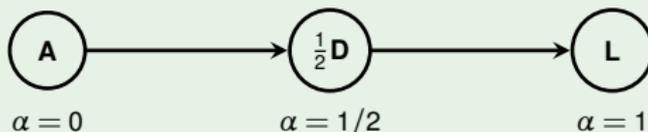
12. V. Nikiforov, *Merging the \mathbf{A} - and \mathbf{Q} -spectral theories*, Applicable Analysis and Discrete Mathematics, 2017.

13. S. Wang et al., *Bounds for the largest and the smallest \mathbf{A}_α eigenvalues of a graph in terms of vertex degrees*, Linear Algebra and its Applications, 2020.

Nouvelle matrice de représentation

Matrice de représentation \mathbf{T}_α ¹⁴

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



Semi-définie positivité :

- ➔ Si $\alpha \geq 1/2$, alors \mathbf{T}_α est semi-définie positive.
- ➔ Condition suffisante mais non nécessaire : il existe des $\alpha \leq 1/2$ tels que $\mathbf{T}_\alpha \geq 0$.

Quelques propriétés algébriques :

- ➔ Si σ_ℓ désigne les valeurs propres de \mathbf{T}_α et m le nombre d'arêtes du graphe, alors

$$\text{Tr}(\mathbf{T}_\alpha) = \sum_{\ell=1}^n \sigma_\ell = 2\alpha m, \quad \text{Tr}(\mathbf{T}_\alpha^2) = \sum_{\ell=1}^n \sigma_\ell^2 = \alpha^2 \sum_{v_i \in \mathcal{V}} \text{deg}(v_i)^2 + 2m(1 - 2\alpha)^2$$

- ➔ Soit $\alpha, \beta \in [0, 1]$. Alors

$$\mathbf{T}_{\alpha+\beta} = \mathbf{T}_\alpha + \mathbf{T}_\beta - \mathbf{A}, \quad \mathbf{T}_{\alpha-\beta} = \mathbf{T}_\alpha - \mathbf{T}_\beta + \mathbf{A}$$

Similarité spectrale par mesure de corrélation ¹⁵

Soient deux graphes G_1 et G_2 d'ordre n . Si $\tilde{\mathbf{v}}_\alpha(G_1)$ (resp. $\tilde{\mathbf{v}}_\alpha(G_2)$) représente le **spectre standardisé** de la matrice $\mathbf{T}_\alpha(G_1)$ (resp. $\mathbf{T}_\alpha(G_2)$), alors une mesure de similarité entre les graphes G_1 et G_2 basée sur une distance construite à partir de la corrélation ¹⁶ est définie comme suit :

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

- **Problème** : une corrélation ne peut être calculée qu'entre deux vecteurs de même taille.
 - **Solution temporaire** : *node-padding* ;
 - **Conjecture** : « Il est moins grave de comparer de l'information artificielle que d'oublier de l'existant » ;
- Mise en place de la classification spectrale à l'aide de \mathbf{T}_α et du **noyau de SVM** défini par les coefficients

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

15. T. Averty et al., *Sur la similarité spectrale des graphes par mesure de corrélation*, GRETSI, 2023

16. S. Van Dongen et A. J. Enright, *Metric distances derived from cosine similarity and Pearson and Spearman correlations*, arXiv :1208.3145, 2012

Classification spectrale de graphes

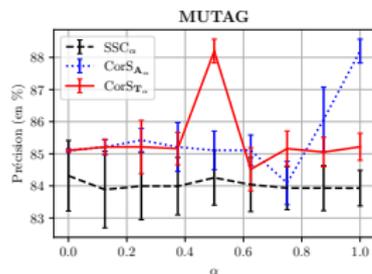
Noyau \ Données	MUTAG	PTC-MR	PROTEINS	IMDB-BINARY	IMDB-MULTI
Shortest Path	85.79 ± 2.51	58.24 ± 2.44	75.07 ± 0.54	71.26 ± 1.04	–
Random Walk	83.72 ± 1.50	57.85 ± 1.30	74.22 ± 0.42	64.54 ± 1.22	34.54 ± 0.76
Graphlet Count	81.58 ± 2.11	57.26 ± 1.41	71.67 ± 0.55	65.87 ± 0.98	43.89 ± 0.38
Weisfeiler-Lehman	82.88 ± 0.57	–	73.52 ± 0.43	71.88 ± 0.77	49.50 ± 0.49
SCNN	90.08 ± 2.01	63.47 ± 2.65	76.51 ± 1.37	73.24 ± 0.96	46.82 ± 1.83
SSC_{β}	84.32 ± 1.10	57.83 ± 2.08	69.99 ± 0.39	67.46 ± 1.01	48.19 ± 0.19
$CorS_{A,\alpha}$	88.20 ± 0.37	59.22 ± 0.59	74.48 ± 0.17	71.24 ± 0.58	49.85 ± 0.5
$CorS_{T,\alpha}$	88.20 ± 0.37	59.22 ± 0.61	74.48 ± 0.17	72.25 ± 0.55	49.69 ± 0.35

Table – 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.

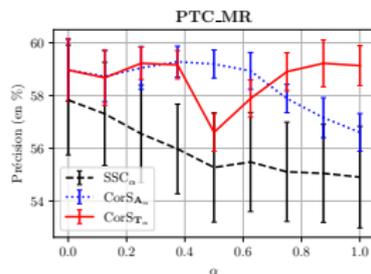
Les résultats précédents sont affichés pour le meilleur α . Existe-t-il un α optimal ?

Classification spectrale de graphes

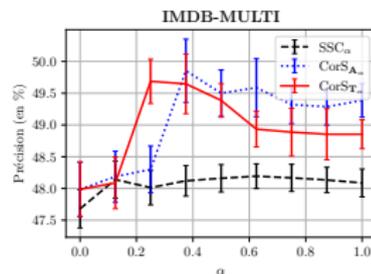
$$\mathbf{T}_\alpha = \alpha \mathbf{D} + (1 - 2\alpha) \mathbf{A}, \quad 0 \leq \alpha \leq 1$$



$\alpha_{opt} \approx 0.5$



$\alpha_{opt} \approx 0.3$



$\alpha_{opt} \approx 0.3$

- ▶ Courbe **rouge** : Noyau gaussien avec la mesure $CorS_{T_\alpha}$
- ▶ Courbe **bleue** : Noyau gaussien avec la mesure $CorS_{A_\alpha}$
- ▶ Courbe **noire** : Noyau gaussien avec la mesure SSC_α

Classification spectrale de séries temporelles

- **Signaux MAD** (*Magnetic Anomaly Detector*)
 - Détection d'objets sous-marins d'intérêt
- **Base de données.** Pour chaque profondeur :
 - 1 000 signaux sans signature
 - 1 000 signaux avec signature
- Signaux \longrightarrow **graphes de visibilité naturelle**¹⁷ (GVN) ;

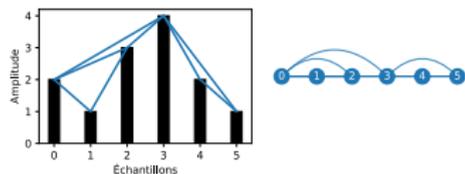
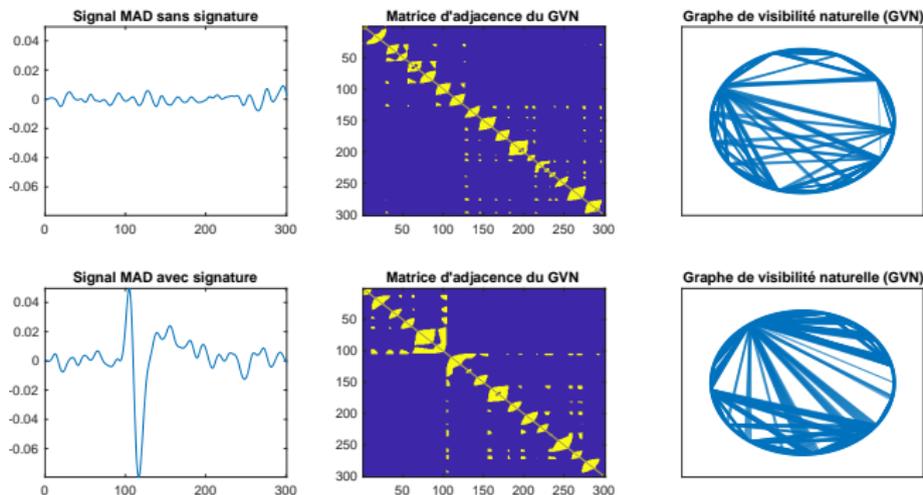


Figure – Construction de graphes de visibilité naturelle



17. L. Lacasa et al., *From time series to complex networks : The visibility graph*, Proceedings of the National Academy of Sciences, 2008

Classification spectrale de séries temporelles

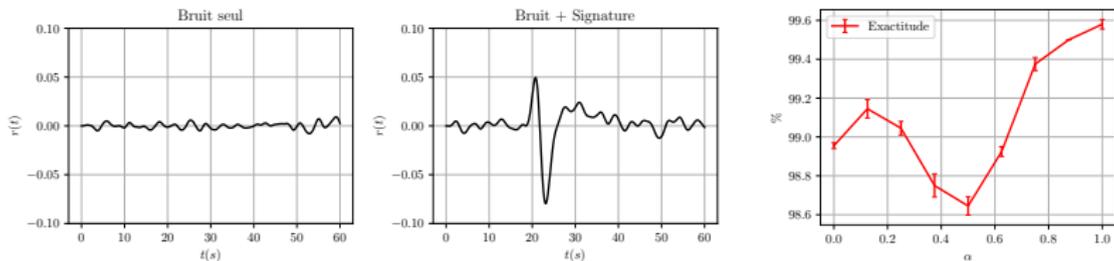


Figure – À gauche : signaux MAD sans et avec signature d'objets sous-marins (profondeur : 150 m). À droite, l'évolution de l'exactitude en fonction de α pour la détection.

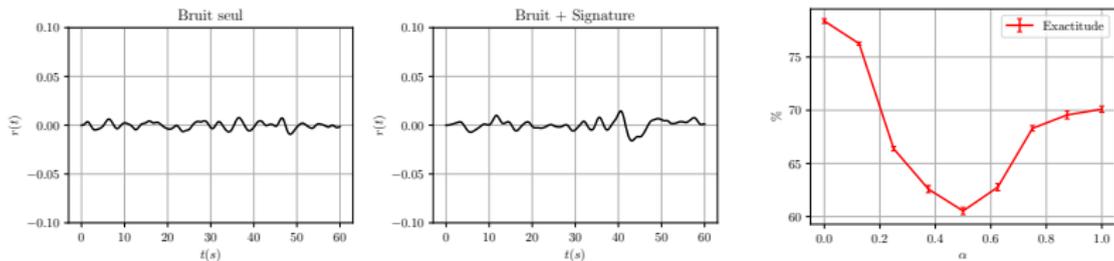


Figure – À gauche : signaux MAD sans et avec signature d'objets sous-marins (profondeur : 250 m). À droite, l'évolution de l'exactitude en fonction de α pour la détection.

Conclusion & Perspectives

👉 **Nouvelle matrice de représentation** définie comme une combinaison linéaire de **A** et **D**.

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

👉 Outil permettant de comprendre l'**évolution graduelle du spectre** entre les matrices **A**, **D** et **L**.

👉 Nouvelle **mesure de similarité spectrale** basée sur une corrélation entre spectres.

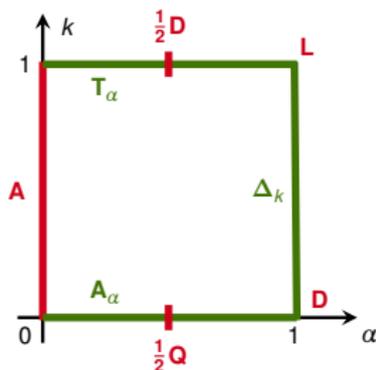
👉 **Résultats satisfaisants** sur une large gamme de graphes.

💡 Trouver un critère déterminant pour le choix d'un **paramètre α optimal**.

💡 Considérer des distances de **transport optimal** (Wassertein) pour éviter le recours au *node-padding*.

💡 Ajouter et formaliser des **propriétés algébriques et spectrales** de \mathbf{T}_α .

💡 Penser un **plan de représentation** grâce à la matrice $\mathbf{T}_{\alpha,k} = \alpha \mathbf{D} + (1 - (k + 1)\alpha) \mathbf{A}$ où $\alpha, k \in [0, 1]$



Sur la similarité spectrale des graphes par mesure de corrélation

AER Tristan AVERTY

MCF Delphine DARÉ-EMZIVAT
PU Abdel-Ouahab BOUDRAA

IRENav (Institut de Recherche de l'École Navale)

IRENav & ENSAM (École nationale supérieure d'Arts et Métiers)
IRENav & ENSAM

XXIX^{ème} Colloque Francophone de Traitement du Signal et des Images (Grenoble)

