

Soutenance de thèse de **Tristan AVERTY** intitulée

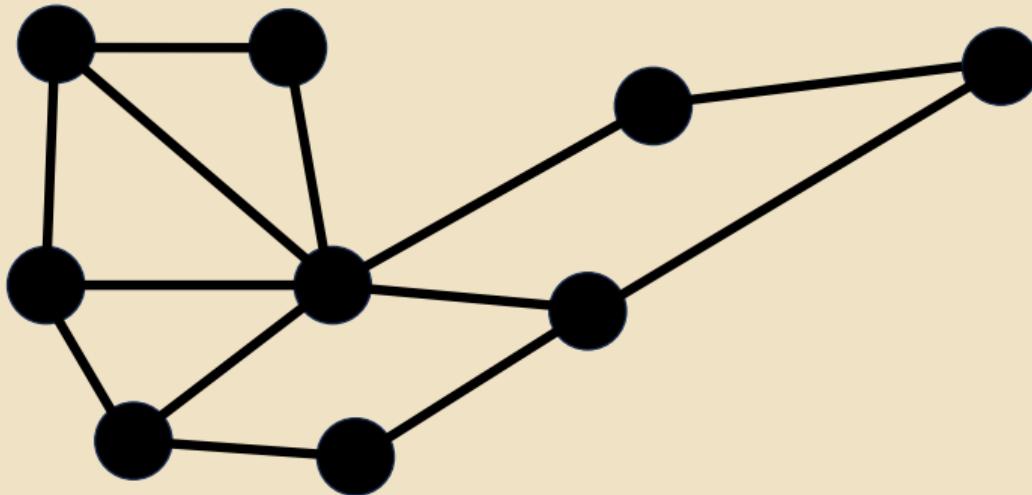
Matrices de représentation généralisées, mesures spectrales et distances statistiques pour l'analyse et la classification de graphes et de signaux

encadrée par

Abdel-Ouahab BOUDRAA
Delphine DARÉ-EMZIVAT

PU ENSAM / École navale, IRENav
MCF ENSAM / École navale, IRENav

à l'Institut de Recherche et d'Études Navales (IRENAV)



Ceci est un graphe

Contexte, enjeux et verrous scientifiques

Données hétérogènes et structurées

- Données **hétérogènes** et **structurées** ...
- ... provenant d'une multitude de **capteurs** (température, pollution, hydrophones, SONAR, RADAR) ...
- ... en quantité **considérable** (en 2020, 22 Go sont créés par habitant et par jour)
- 💡 Représentation idoine par des **graphes**

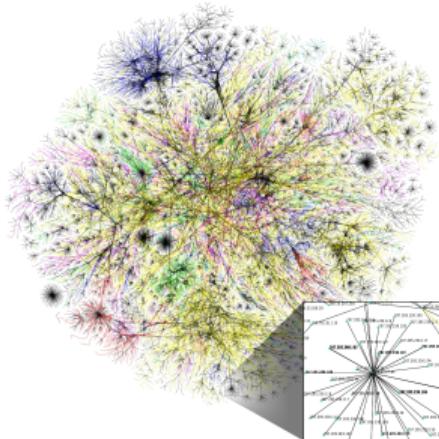


Figure – Carte partielle d'Internet en 2005 (les sommets sont des adresses IP)

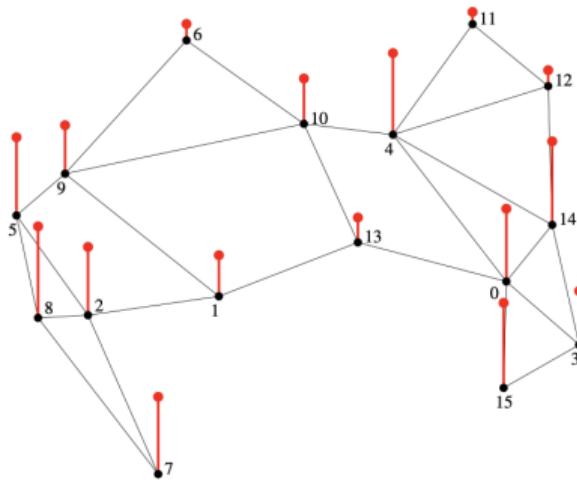


Figure – Mesures de température multicapteurs (les sommets sont les capteurs)

Contexte, enjeux et verrous scientifiques

Réseaux & infrastructures

💡 Représentation naturelle des **réseaux** sous la forme de graphe

- ⚡ Réseaux d'**approvisionnement** (gaz, eau, électricité, ...)
- 🌐 Réseaux **sociaux**
- 📲 Réseaux de **communication** (câbles sous-marins)
- 🚗 Réseaux de **transport** (routier, maritime, aérien, ...)



Figure – Réseau électrique français

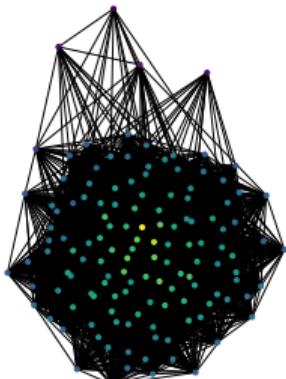


Figure – Relations entre élèves d'un lycée de Quimper



Figure – Carte des câbles sous-marins de communication

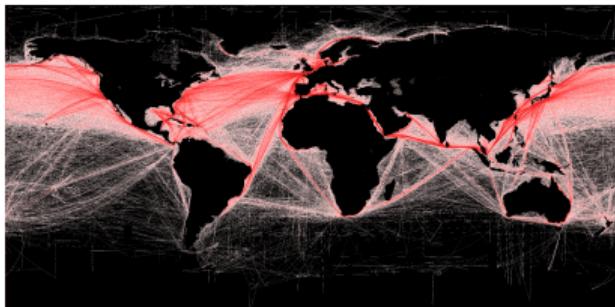


Figure – Réseau de transport maritime

Contexte, enjeux et verrous scientifiques

Vulnérabilité des réseaux

⚠ Sujets à des sabotages ou défaillances

- ✖ États-Unis : Panne de courant en août 2003
- ✖ Inde : Panne de courant en juillet 2012
- ✖ Canal de Suez : obstruction en mars 2021
- ✖ Pipelines NordStream 1 et 2 : sabotage en septembre 2022
- ✖ Câble électrique Estlink 2 : sabotage en décembre 2024

ÉCONOMIE · ÉCONOMIE MONDIALE

Le canal de Suez bloqué à cause d'un cargo échoué en travers

Le porte-conteneurs « Ever-Given » s'était retrouvé en travers du canal reliant la mer Rouge à la Méditerranée, bloquant toute circulation. Le retour à la normale n'était toujours pas acquis jeudi.

Le Monde avec AFP
Publié le 24 mars 2021 à 11h25, modifié le 29 mars 2021 à 08h02 · 0 Lecture 3 min.

💡 Études sur la vulnérabilité des réseaux

- ➡ Impact sur l'« information » qui circule
- ➡ Vulnérabilité des sommets et des arêtes

❓ Comment quantifier la vulnérabilité d'une arête, si possible en utilisant des outils de la théorie de l'information ?

Câbles en Baltique : l'ancre du pétrolier soupçonné de sabotage a été retrouvée

L'ancre soupçonnée d'être impliquée dans la rupture d'un câble sous-marin dans le golfe de Finlande a été récupérée par les autorités finlandaises lundi 6 janvier. Reste à savoir si cette dernière appartient à l'« Eagle S », le pétrolier suspecté d'avoir commis le sabotage.

Le marin
Avec l'AFP.
Publié le 07/01/2025 à 18h15

LIRE PLUS TARD
PARTAGER

Newsletter Le marin

Du lundi au vendredi,
recevez l'actualité
économique maritime
analysée par l'équipe
rédactionnelle du marin.

Votre e-mail OK



L'ancre a été remontée à la surface après 17 h lundi 6 janvier. | POLICE OF FINLAND

Contexte, enjeux et verrous scientifiques

Classification de graphes et de signaux vus comme des graphes

GRAPHES NATURELS

- Cyberdéfense
- Finance
- Médecine
- Traitement du langage naturel
- Bio-informatique
- Naval : Déetecter des routes maritimes suspectes

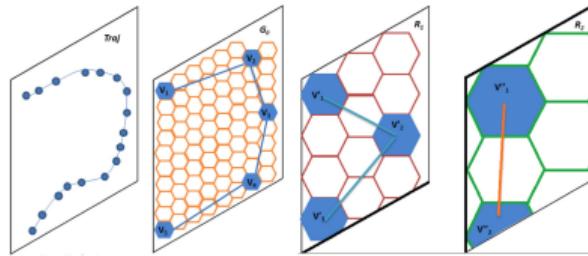


Figure – Graphes à différentes échelles du trafic maritime¹

SIGNALS VUS COMME DES GRAPHES

- Toutes les tâches de classification de signaux
 - 💡 Graphes de visibilité
 - ➡ Sommets : Échantillons / Arêtes : Visibilité

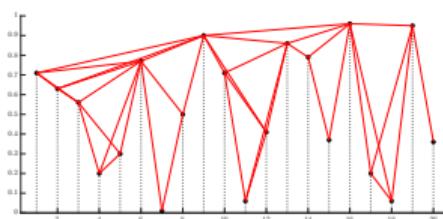
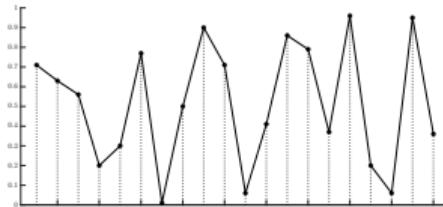


Figure – Construction d'un graphe de visibilité

1. M. Maslek Elayam et al., A hierarchical graph-based model for mobility data representation and analysis, *Data & Knowledge Engineering*, 141:102054, 2022.

Contexte, enjeux et verrous scientifiques

Classification structurelle de graphes

Graphes → Attributs structurels → Classification structurelle

- Comparaison des **marches aléatoires**
- Décompte des **sous-graphes**
- Test d'isomorphisme de **Weisfeiler-Lehman**
- Comparaison des **plus courts chemins**

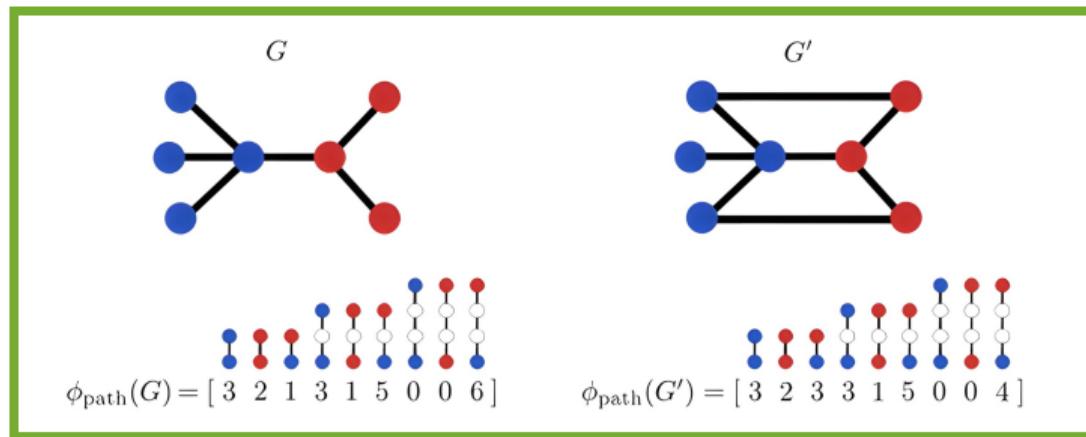


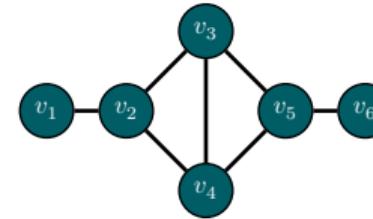
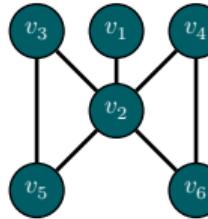
Figure – Principe de la comparaison des plus courts chemins

Contexte, enjeux et verrous scientifiques

Classification spectrale de graphes

Graphes → Matrices de représentation → Valeurs propres → Classification spectrale

- 💡 Des propriétés sur les spectres permettent de retrouver des **éléments structurels**
- **Deux écoles** : celle de la matrice d'adjacence² A et celle de la matrice Laplacienne L ³
- Utilisation d'une **mesure de similarité** entre graphes, jouant le rôle d'une distance
- ⚠ Existence de **graphes cospectraux**



? Comment comparer les graphes de visibilité de manière efficiente ?

? Quelle matrice représente le mieux un graphe ?

? Existe-t-il une mesure de similarité spectrale performante ne requérant qu'un seul spectre ?

2. A. Sandryhaila et J.M.F Moura, Discrete Signal Processing on Graphs : Frequency Analysis, *IEEE Transactions on Signal Processing*, 62(12):3042–3054, 2014

3. D. Shuman *et al.*, The Emerging Field of Signal Processing on Graphs : Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains, *IEEE SPM*, 30(3):83–98, 2013

Plan

- 1 Contexte, enjeux et verrous scientifiques
- 2 Généralités sur les graphes
- 3 Vulnérabilité informationnelle de réseaux
 - Entropie de von Neumann comme mesure de vulnérabilité d'une arête : Algorithme EIVP
 - Optimisations de l'algorithme EIVP basées sur des approximations de l'entropie de von Neumann
- 4 Vers une nouvelle matrice de représentation de graphes ?
 - Définition et propriétés d'un nouveau plan de représentation
 - Nouvelle mesure de similarité spectrale : Application à la classification de graphes
- 5 Graphes de visibilité pour classer et caractériser des séries temporelles
 - Représentation des graphes de visibilité dans un plan informationnel pour une estimation du coefficient de Hurst
 - Classification de signaux grâce à une comparaison des distributions de degrés de leurs graphes de visibilité
- 6 Conclusions & Perspectives

Généralités sur les graphes

Graphe (simple non-orienté) et matrice d'adjacence

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

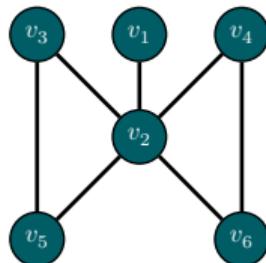
- ↳ d'un ensemble $\mathcal{V} = \{v_i\}_{1 \leq i \leq n}$ de **n sommets**
- ↳ d'un ensemble $\mathcal{E} \subseteq \{\{i, j\} \mid i, j \in \mathcal{V}, i \neq j\}$ de **m arêtes**
- + d'un ensemble de **poids** w_{ij} qui pondèrent (éventuellement) les arêtes $\{i, j\}$

► Deux sommets sont dits **adjacents** s'ils sont reliés par une arête

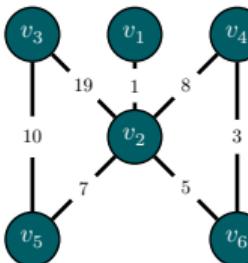
💡 Représentation possible du graphe grâce à une **matrice d'adjacence A** (ou une **matrice de poids W**) de taille $n \times n$.

- ↳ Coefficient A_{ij} (resp. W_{ij}) égal à 1 (resp. w_{ij}) s'il existe une arête $\{i, j\}$ et 0 sinon

⚠ Matrices symétriques mais **non semi-définies positives** (elles possèdent des valeurs propres négatives)



(a) Graphe simple non-orienté G



(b) Graphe G avec pondérations

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

(c) Matrice d'adjacence A

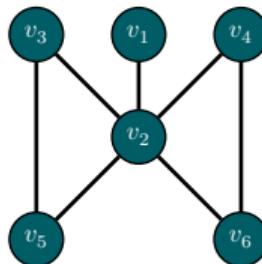
$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 19 & 8 & 7 & 5 \\ 0 & 19 & 0 & 0 & 10 & 0 \\ 0 & 8 & 0 & 0 & 0 & 3 \\ 0 & 7 & 10 & 0 & 0 & 0 \\ 0 & 5 & 0 & 3 & 0 & 0 \end{pmatrix}$$

(d) Matrice de poids W

Généralités sur les graphes

Matrice et distribution des degrés

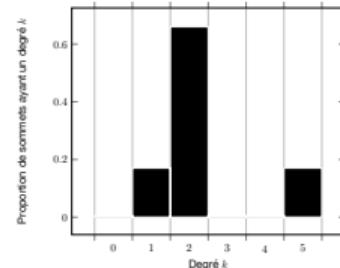
- Le **degré** $\deg(i)$ d'un sommet i est le nombre d'arêtes reliant ce sommet à d'autres : $\deg(i) = \sum_{j \in \mathcal{V}} \mathbf{A}_{ij}$
- 💡 Représentation possible du graphe grâce à une **matrice des degrés D**
 - Coefficient D_{ii} égal à $\deg(i)$
- 💡 Représentation possible du graphe grâce à une **distribution des degrés** $p_k = \frac{|\{i \in \mathcal{V} : \deg(i) = k\}|}{n}$ pour $1 \leq k \leq n$



(a) Graphe G simple non orienté et pondéré

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

(b) Matrice des degrés \mathbf{D}

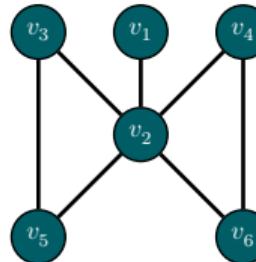


(c) Distribution des degrés

Généralités sur les graphes

Matrice Laplacienne

💡 Représentation possible du graphe grâce à une **matrice Laplacienne** $\mathbf{L} = \mathbf{D} - \mathbf{A}$



(a) Graphe G simple non orienté et pondéré

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

(b) Matrice Laplacienne $\mathbf{L} = \mathbf{D} - \mathbf{A}$

➤ Pourquoi utiliser cette matrice ?

✎ **Semi-définie positivité** : pour tout $\mathbf{x} = (x_i)_{1 \leq i \leq n}$, la matrice \mathbf{L} vérifie $\mathbf{x} \mathbf{L} \mathbf{x}^\top = \frac{1}{2} \sum_{\{i,j\} \in \mathcal{E}} (x_i - x_j)^2 \geq 0$

✎ **Propriétés algébriques et spectrales** : si $0 = \mu_1 \leq \mu_2 \leq \dots \leq \mu_n$ désigne ses valeurs propres, alors

- ➡ $\text{Tr}(\mathbf{L}) = \sum_{\ell=1}^n \mu_\ell = 2m$ et $\text{Tr}(\mathbf{L}^2) = \sum_{\ell=1}^n \mu_\ell^2 = 2m + \sum_{i \in \mathcal{V}} \deg(i)^2$

- ➡ 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**

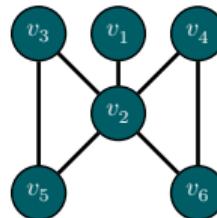
Généralités sur les graphes

Problème de la cospectralité & matrice Laplacienne sans-signe

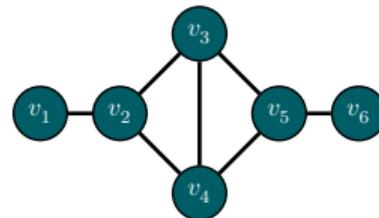
► Soit une matrice de représentation \mathbf{M}

- ↳ Soient $\nu_1^{(1)} \leq \nu_2^{(1)} \leq \dots \leq \nu_n^{(1)}$ les valeurs propres de $\mathbf{M}(G_1)$
- ↳ Soient $\nu_1^{(2)} \leq \nu_2^{(2)} \leq \dots \leq \nu_n^{(2)}$ les valeurs propres de $\mathbf{M}(G_2)$
- G_1 et G_2 sont dits **\mathbf{M} -cospectraux** si $\nu_\ell^{(1)} = \nu_\ell^{(2)}$ pour tout $1 \leq \ell \leq n$

⚠ Véritable problème lors de la création de mesures de similarité spectrale



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



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

💡 Représentation possible du graphe grâce à une **matrice Laplacienne sans-signe**⁴ $\mathbf{Q} = \mathbf{D} + \mathbf{A}$

☞ Matrice **semi-définie positive**

☞ **Moins de graphes Q-cospectraux**

4. D. Cvetković et al., Signless Laplacians of finite graphs, *Linear Algebra and its Applications*, 423(1):155–171, 2007

Généralités sur les graphes

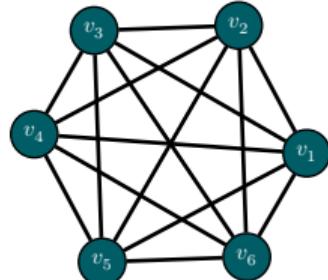
« Graphes-jouets »

► Graphes aux structures particulières

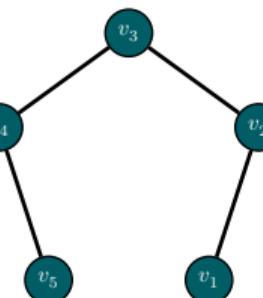
- Graphe complet \mathcal{K}_n , graphe étoile \mathcal{S}_n , graphe chemin \mathcal{P}_n , graphe comète $\mathcal{K}_{n_1} \cup \mathcal{P}_{n_2}, \dots$

► Graphes aléatoires d'Erdős-Rényi

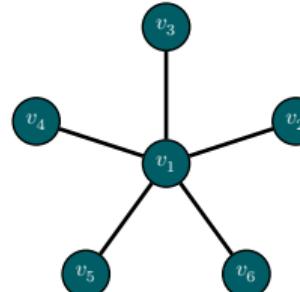
- Graphe $\mathcal{G}_{n,m} = (\mathcal{V}, \mathcal{E})$ d'ordre n
- Les arêtes sont créées de manière aléatoire de sorte à obtenir, au total, m arêtes
- Une variante est $\mathcal{G}_{n,p}$ où la probabilité qu'une arête soit construite est égale à p



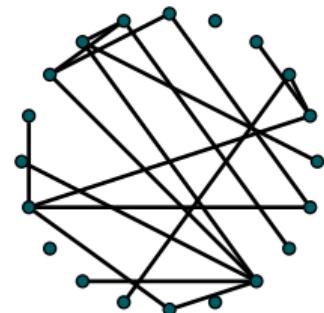
(a) Graphe complet \mathcal{K}_6 (graphe 5-régulier).



(b) Graphe chemin \mathcal{P}_5



(c) Graphe étoile \mathcal{S}_5



(d) Graphe aléatoire d'Erdős-Rényi $\mathcal{G}_{20,0.1}$

Plan

1 Contexte, enjeux et verrous scientifiques

2 Généralités sur les graphes

3 Vulnérabilité informationnelle de réseaux

- Entropie de von Neumann comme mesure de vulnérabilité d'une arête : Algorithme EIVP
- Optimisations de l'algorithme EIVP basées sur des approximations de l'entropie de von Neumann

4 Vers une nouvelle matrice de représentation de graphes ?

- Définition et propriétés d'un nouveau plan de représentation
- Nouvelle mesure de similarité spectrale : Application à la classification de graphes

5 Graphes de visibilité pour classer et caractériser des séries temporelles

- Représentation des graphes de visibilité dans un plan informationnel pour une estimation du coefficient de Hurst
- Classification de signaux grâce à une comparaison des distributions de degrés de leurs graphes de visibilité

6 Conclusions & Perspectives

Vulnérabilité informationnelle de réseaux

Constats relatifs à la vulnérabilité des réseaux

⚠ Pas de définition objective de la vulnérabilité

« La vulnérabilité est l'antonyme naturel de la robustesse⁵ »

- Quelques mesures de vulnérabilité, listées dans l'article de Freitas et al. :
 - ➡ (Globale) La **valeur de Fiedler** μ_2 qui traduit la connectivité du graphe
 - ➡ (Globale) **Indice de Kirchhoff** qui évalue la résistance du graphe vu comme un réseau électrique
 - ➡ (Globale) La valeur propre maximale λ_n de la matrice d'adjacence A , appelée **Shield Value**
 - ➡ (Locale) **Centralité d'intermédiarité** qui dénombre le nombre de plus courts chemins passant par l'arête étudiée
- Possible de passer d'une mesure globale à une mesure locale
- 🚫 Aucune mesure n'est basée sur une **vision « système physique »** d'un graphe

Système physique → Matrice de densité → Information → Entropie (de von Neumann)

5. S. Freitas et al., Graph vulnerability and robustness : A survey. *IEEE Transactions on Knowledge and Data Engineering*, 35(6):5915–5934, 2022

Vulnérabilité informationnelle de réseaux

Matrice de densité et entropie de von Neumann d'un système physique

- Un **système** peut être représenté par une **combinaison convexe d'états purs** $|\Psi_i\rangle$:

$$\rho := \sum_{i=1}^n p_i |\Psi_i\rangle\langle\Psi_i|$$

avec p_i la probabilité que cet état se trouve dans l'état pur $|\Psi_i\rangle$

- La matrice ρ , appelée **matrice de densité**, est

- ☞ hermitienne ($\rho = \rho^H$)
- ☞ semi-définie positive
- ☞ et sa trace $\text{Tr}(\rho)$ est égale à 1

- L'**entropie de von Neumann** d'un système, décrit par une matrice de densité ρ , est donnée par

$$S(\rho) = -\text{Tr}(\rho \ln \rho) = -\sum_{i=1}^n p_i \ln p_i$$

- ☞ L'entropie de von Neumann **généralise l'entropie de Shannon** au cadre des systèmes physiques
- ☞ Elle donne une idée du **contenu informationnel** d'un état quantique

!?

Mais où sont passés les graphes !?

Vulnérabilité informationnelle de réseaux

Matrice de densité et entropie de von Neumann d'un graphe

💡 La matrice de densité⁶ $\rho(G)$ d'un graphe G est égale à

$$\rho(G) = \frac{\mathbf{L}}{\text{Tr}(\mathbf{L})}$$

💡 La matrice Laplacienne \mathbf{L} est

- hermitienne ($\mathbf{L} = \mathbf{L}^\top$)
- semi-définie positive
- mais sa trace $\text{Tr}(\mathbf{L})$ n'est pas égale à 1

💡 Soit $\rho(G)$ la matrice de densité d'un graphe G admettant $0 = \nu_1 \leq \nu_2 \leq \dots \leq \nu_n$ pour valeurs propres.

L'entropie de von Neumann⁷ $S(G)$ du graphe G est égale à

$$S(G) := - \sum_{\ell=1}^n \nu_\ell \ln \nu_\ell$$

📝 Caractérise le contenu informationnel du graphe, compris comme un système physique

📝 Traduit la régularité (entropie nulle pour un graphe vide, maximale pour un graphe complet et intermédiaire pour un graphe étoile)

6. 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*, 10(3):291–317, 2006

7. F. Passerini et S. Severini, Quantifying complexity in networks : the von Neumann entropy, *International Journal of Agent Technologies and Systems*, 1(4):58–67, 2009

Vulnérabilité informationnelle de réseaux

Vulnérabilité informationnelle d'une arête

? Comment quantifier la vulnérabilité d'une arête, si possible en utilisant des outils de la théorie de l'information ?

► Soit un graphe $G = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ et $\widetilde{G}_{ij,\xi} = (\mathcal{V}, \mathcal{E}, \widetilde{\mathbf{W}})$ le graphe G dont l'arête $\{i, j\}$ est perturbée par $\xi \in [0, 1]$

- ↳ Les entrées w_{ij} et w_{ji} de \mathbf{W} deviennent $\widetilde{w}_{ij} = (1 - \xi)w_{ij}$ et $\widetilde{w}_{ji} = (1 - \xi)w_{ji}$
- ↳ Perturbation pouvant modéliser une **défaillance** ou un **sabotage**

💡 La **vulnérabilité informationnelle** de l'arête $\{i, j\}$ est donnée par la variation relative de l'entropie

$$\eta_{ij,\xi} := 100 \times \frac{S(\widetilde{G}_{ij,\xi}) - S(G)}{S(G)}$$

Vulnérabilité informationnelle de réseaux

Algorithme EIVP

Algorithme de repondération **EIVP** (*Edge Informational Vulnerability to Perturbation*)

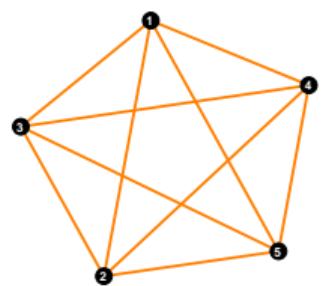
Entrée : Graphe $G = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ et coefficient de perturbation $\xi \in [0, 1]$

- 1 Calcul de l'entropie $S(G)$ du graphe initial
- 2 **for** $\{i, j\} \in \mathcal{E}$ **do**
- 3 $\widetilde{G}_{ij,\xi} \leftarrow (\mathcal{V}, \mathcal{E}, \widetilde{\mathbf{W}})$ avec $[\widetilde{\mathbf{W}}]_{kl} = \begin{cases} w_{kl}(1 - \xi), & \text{si } \{k, l\} = \{i, j\} \\ w_{kl}, & \text{sinon} \end{cases}$
- 4 Calcul de l'entropie $S(\widetilde{G}_{ij,\xi})$ du graphe perturbé
- 5 $[\mathbf{W}_S^{\text{EIVP}}]_{ij} \leftarrow \eta_{ij,\xi}$

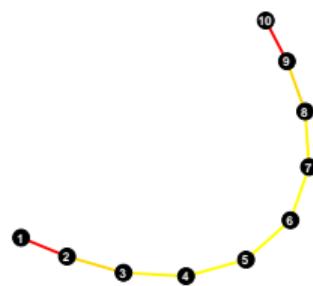
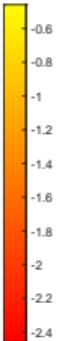
Sortie : Graphe $G_S^{\text{EIVP}} = (\mathcal{V}, \mathcal{E}, \mathbf{W}_S^{\text{EIVP}})$

Vulnérabilité informationnelle de réseaux

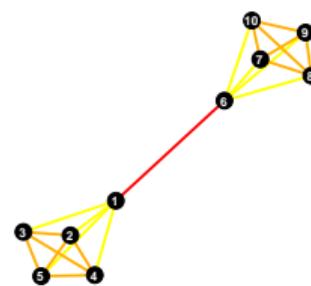
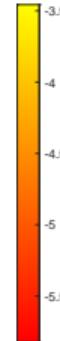
Application de l'algorithme EIVP sur des « graphes-jouets »



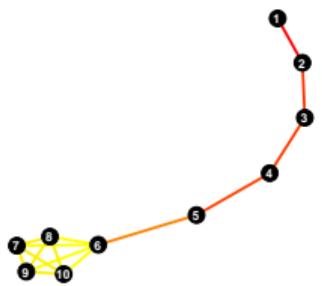
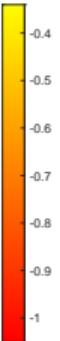
(a) Graphe complet \mathcal{K}_5



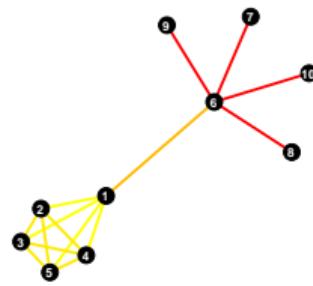
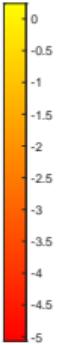
(b) Graphe chemin \mathcal{P}_{10}



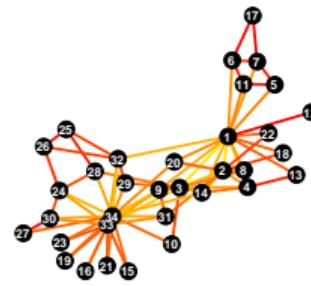
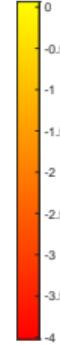
(c) Graphe Barbell



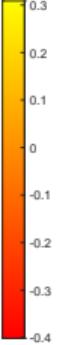
(d) Graphe comète



(e) Graphe Complet+Étoile connecté



(f) Graphe KarateClub (34 sommets / 78 arêtes)

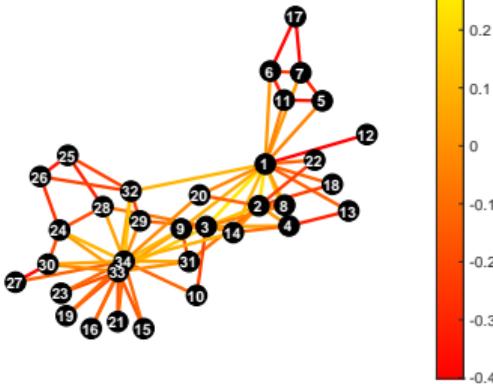


Vulnérabilité informationnelle de réseaux

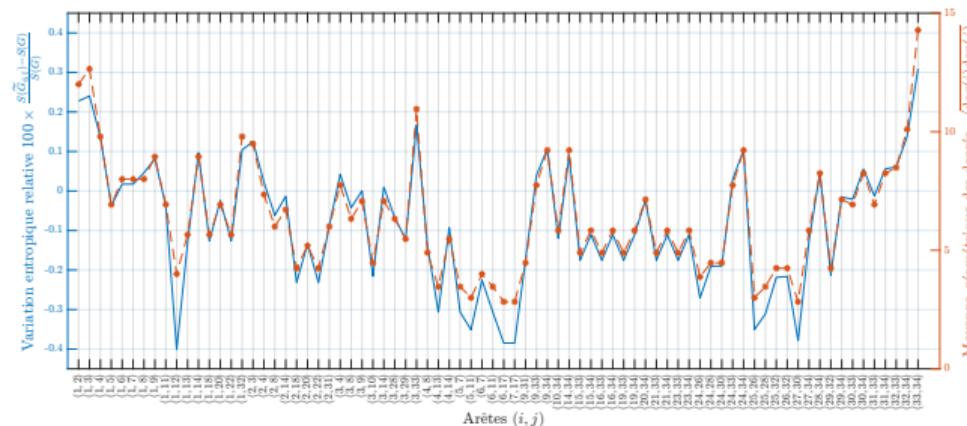
Corrélation entre vulnérabilité informationnelle et moyenne géométrique des degrés

► Résumé des conjectures :

- ➡ Plus forte décroissance de l'entropie ➡ Arêtes dont la suppression déconnecte le graphe
- ➡ Décroissance de l'entropie ➡ Arêtes dont la suppression sépare le graphe en sous-graphes déséquilibrés
- ➡ Croissance de l'entropie ➡ Arêtes situées dans des zones très denses
- ➡ Corrélation entre $\eta_{ij,\xi}$ et la moyenne géométrique $\sqrt{\deg(i) \deg(j)}$ des degrés des sommets connectés ($r = 0.97$)



(a) Vulnérabilités informationnelles du graphe KarateClub



(b) Corrélation entre vulnérabilité informationnelle et moyenne géométrique des degrés

🚫 Difficile de prouver ces conjectures car l'entropie est calculée à partir des valeurs propres

Plan

1 Contexte, enjeux et verrous scientifiques

2 Généralités sur les graphes

3 Vulnérabilité informationnelle de réseaux

- Entropie de von Neumann comme mesure de vulnérabilité d'une arête : Algorithme EIVP
- Optimisations de l'algorithme EIVP basées sur des approximations de l'entropie de von Neumann

4 Vers une nouvelle matrice de représentation de graphes ?

- Définition et propriétés d'un nouveau plan de représentation
- Nouvelle mesure de similarité spectrale : Application à la classification de graphes

5 Graphes de visibilité pour classer et caractériser des séries temporelles

- Représentation des graphes de visibilité dans un plan informationnel pour une estimation du coefficient de Hurst
- Classification de signaux grâce à une comparaison des distributions de degrés de leurs graphes de visibilité

6 Conclusions & Perspectives

Vulnérabilité informationnelle de réseaux

Complexité de l'algorithme EIVP

⚠ Problème de l'algorithme EIVP : Complexité en $O(mn^3)$

Algorithme de repondération **EIVP** (*Edge Informational Vulnerability to Perturbation*)

Entrée : Graphe $G = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ et coefficient de perturbation $\xi \in [0, 1]$

1 Calcul de l'entropie $S(G)$ du graphe initial → **Complexité en $O(n^3)$**

2 **for** $\{i, j\} \in \mathcal{E}$ **do**

3 $\widetilde{G}_{ij,\xi} \leftarrow (\mathcal{V}, \mathcal{E}, \widetilde{\mathbf{W}})$ avec $[\widetilde{\mathbf{W}}]_{kl} = \begin{cases} w_{kl}(1 - \xi), & \text{si } \{k, l\} = \{i, j\} \\ w_{kl}, & \text{sinon} \end{cases}$

4 Calcul de l'entropie $S(\widetilde{G}_{ij,\xi})$ du graphe perturbé → **Complexité en $O(n^3)$**

5 $[\mathbf{W}_S^{\text{EIVP}}]_{ij} \leftarrow \eta_{ij,\xi} := 100 \times \frac{S(\widetilde{G}_{ij,\xi}) - S(G)}{S(G)}$

Sortie : Graphe $G_S^{\text{EIVP}} = (\mathcal{V}, \mathcal{E}, \mathbf{W}_S^{\text{EIVP}})$

💡 Utilisation d'approximations de l'entropie de von Neumann. Mais lesquelles ?

$$1 \quad S(G) = - \sum_{\ell=1}^n \boxed{\nu_\ell \ln \nu_\ell}$$

$$2 \quad S(G) = - \sum_{\ell=1}^n \boxed{\nu_\ell} \ln \boxed{\nu_\ell}$$

Vulnérabilité informationnelle de réseaux

Première approximation de l'entropie de von Neumann

Première approximation (Développement limité) de $S(G) = -\sum_{\ell=1}^n \nu_\ell \ln \nu_\ell$

Soit le **développement limité** à l'ordre 2 de la fonction $\nu_\ell \ln \nu_\ell$ autour d'un point $0 < \nu_* \leq 1$:

$$\nu_\ell \ln \nu_\ell \underset{\nu_*}{\sim} -\frac{\nu_*}{2} + \nu_\ell \ln \nu_* + \frac{1}{2\nu_*} \nu_\ell^2.$$

Soit un graphe G d'ordre n . Ce développement limité permet d'obtenir l'approximation de l'entropie suivante :

$$S_{\text{DL},\nu_*}(G) = -\sum_{\ell=1}^n \left(-\frac{\nu_*}{2} + \nu_\ell \ln \nu_* + \frac{1}{2\nu_*} \nu_\ell^2 \right) = \frac{n\nu_*}{2} - \ln \nu_* \sum_{\ell=1}^n \nu_\ell - \frac{1}{2\nu_*} \sum_{\ell=1}^n \nu_\ell^2 = \frac{n\nu_*}{2} - \ln \nu_* - \frac{1}{2\nu_*} \text{Tr}(\rho^2).$$

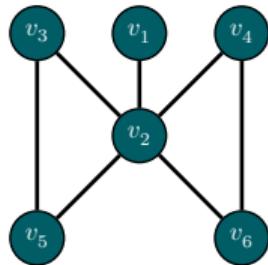
- ! Dans les faits, $\nu_* = 1/n$ car, la somme des valeurs propres valant 1, leur moyenne est égale à $1/n$
- ! Algorithme **DL-EIVP** en remplaçant l'entropie de von Neumann $S(G)$ par l'approximation $S_{\text{DL},1/n}(G)$
- ! Complexité censée passer de $O(mn^3)$ à $O(mn^2)$ mais atteint $O(m(n+m))$ (voire $O(mn)$ si non-pondéré) car

$$\text{Tr}(\rho^2) = \frac{1}{\text{Tr}(\mathbf{L})^2} \left(\sum_{i \in \mathcal{V}} \deg(i)^2 + 2 \sum_{\{i,j\} \in \mathcal{E}} w_{ij}^2 \right) = \frac{1}{(2m)^2} \left(\sum_{i \in \mathcal{V}} \deg(i)^2 + 2m \right)$$

Vulnérabilité informationnelle de réseaux

Impact d'une perturbation sur les coefficients des matrices de représentation

$$2 \quad S(G) := - \sum_{\ell=1}^n |\nu_\ell| \ln |\nu_\ell|$$



(a) Graphe G

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

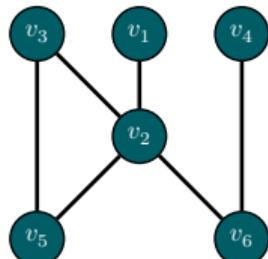
(b) Matrice d'adjacence \mathbf{A}

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

(c) Matrice des degrés \mathbf{D}

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

(d) Matrice Laplacienne \mathbf{L}



(a) Graphe perturbé \widetilde{G}

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & \textcolor{red}{0} & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & \textcolor{red}{0} & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

(b) Matrice d'adjacence $\widetilde{\mathbf{A}}$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \textcolor{red}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \textcolor{red}{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

(c) Matrice des degrés $\widetilde{\mathbf{D}}$

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & \textcolor{red}{4} & -1 & 0 & -1 & -1 \\ 0 & -1 & 2 & 0 & -1 & 0 \\ 0 & \textcolor{red}{0} & 0 & 1 & 0 & -1 \\ 0 & -1 & -1 & 0 & 2 & 0 \\ 0 & -1 & 0 & -1 & 0 & 2 \end{pmatrix}$$

(d) Matrice Laplacienne $\widetilde{\mathbf{L}}$

Vulnérabilité informationnelle de réseaux

Deuxième approximation de l'entropie de von Neumann

Lemme sur les valeurs propres d'une matrice perturbée⁸

Soit une matrice symétrique $\mathbf{M} = [m_{ij}]_{1 \leq i,j \leq n}$ d'*eigenpair* $(\mu_\ell, \mathbf{u}_\ell)_{1 \leq \ell \leq n}$. Si les entrées m_{kl} sont perturbées ($\tilde{m}_{kl} = m_{kl} + \Delta m_{kl}$), la valeur propre μ_ℓ devient

$$\tilde{\mu}_\ell = \mu_\ell + \sum_{k,l=1}^n [\mathbf{u}_\ell]_k [\mathbf{u}_\ell]_l \Delta m_{kl}.$$

Deuxième approximation (Théorie de perturbation matricielle)

La matrice Laplacienne \mathbf{L} , d'*eigenpair* $(\mu_\ell, \mathbf{u}_\ell)_{1 \leq \ell \leq n}$, voit 4 de ses coefficients modifiés :

$$[\tilde{\mathbf{L}}]_{ij} = [\mathbf{L}]_{ij} + \xi w_{ij}, \quad [\tilde{\mathbf{L}}]_{ji} = [\mathbf{L}]_{ji} + \xi w_{ji}, \quad [\tilde{\mathbf{L}}]_{ii} = [\mathbf{L}]_{ii} - \xi w_{ij}, \quad [\tilde{\mathbf{L}}]_{jj} = [\mathbf{L}]_{jj} - \xi w_{ji}.$$

En utilisant le lemme précédent, les valeurs propres $(\tilde{\mu}_\ell)_{1 \leq \ell \leq n}$ de $\tilde{\mathbf{L}}$ peuvent s'écrire

$$\tilde{\mu}_\ell = \mu_\ell - \xi w_{ij} ([\mathbf{u}_\ell]_i - [\mathbf{u}_\ell]_j)^2.$$

Ainsi, comme $\tilde{\rho} = \tilde{\mathbf{L}} / \text{Tr}(\tilde{\mathbf{L}})$, une approximation de l'entropie de von Neumann peut être

$$S(\tilde{G}_{ij,\xi}) \approx \sum_{\ell=1}^n \tilde{\nu}_\ell \ln \tilde{\nu}_\ell, \quad \tilde{\nu}_\ell = \frac{\tilde{\mu}_\ell}{\text{Tr}(\mathbf{L}) - 2\xi w_{ij}}.$$

8. G.W. Stewart et J.-G. Sun, Matrix Perturbation Theory, Academic Press, 1990

Vulnérabilité informationnelle de réseaux

Algorithme fast-EIVP

Algorithme de repondération **fast-EIVP**

Entrée : Graphe $G = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ et coefficient de perturbation $\xi \in [0, 1]$

```

1 Calcul des eigenpairs de la matrice Laplacienne ( $\mu_\ell, \mathbf{u}_\ell$ )
2 Calcul de l'entropie  $S(G)$  du graphe initial
3 for  $\{i, j\} \in \mathcal{E}$  do
4    $S(\widetilde{G}_{ij,\xi}) \leftarrow 0$ 
5   for  $1 \leq \ell \leq n$  do
6      $\tilde{\nu}_\ell \leftarrow \frac{\mu_\ell - \xi w_{ij} ([\mathbf{u}_\ell]_i - [\mathbf{u}_\ell]_j)^2}{\text{Tr}(\mathbf{L}) - 2\xi w_{ij}}$ 
7      $S(\widetilde{G}_{ij,\xi}) \leftarrow S(\widetilde{G}_{ij,\xi}) - \tilde{\nu}_\ell \ln \tilde{\nu}_\ell$ 
8   [math]\mathbf{W}_S^{\text{f-EIVP}}]_{ij} \leftarrow \eta_{ij,\xi}

```

Sortie : Graphe $G_S^{\text{f-EIVP}} = (\mathcal{V}, \mathcal{E}, \mathbf{W}_S^{\text{f-EIVP}})$



Complexité totale en $O(n^3 + mn)$

Vulnérabilité informationnelle de réseaux

Temps de calcul et erreurs sur des graphes aléatoires d'Erdős-Rényi

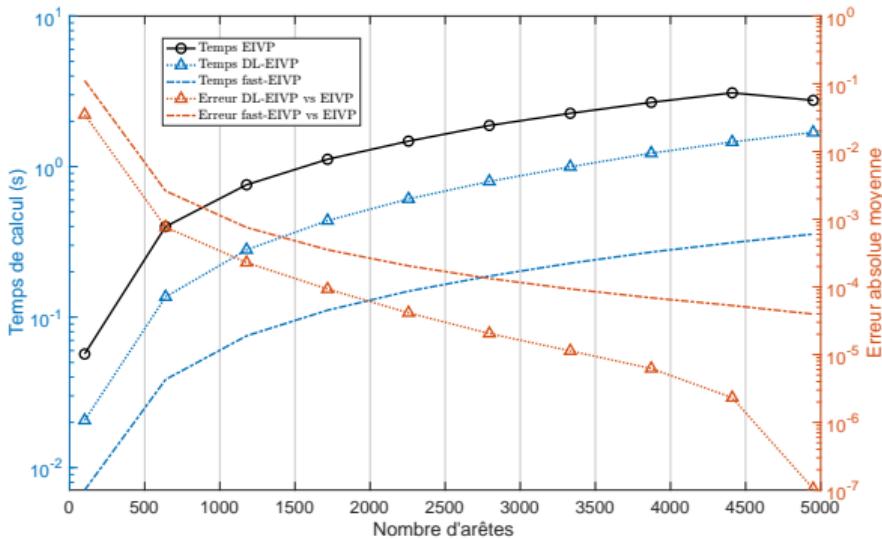


Figure – Temps de calcul (courbes bleues) et erreurs (courbes orange) des algorithmes EIVP sur des graphes $G_{100,m}$ pour $100 \leq m \leq 4950$.

► Algorithmique DL-EIVP

- Erreur commise « DL-EIVP vs EIVP »
- Temps de calcul : complexité en $O(m(n + m))$

► Algorithmique fast-EIVP

- Erreur commise « fast-EIVP vs EIVP »
- Temps de calcul : complexité en $O(n^3 + mn)$

Vulnérabilité informationnelle de réseaux

Temps de calcul et erreurs sur des graphes aléatoires d'Erdős-Rényi

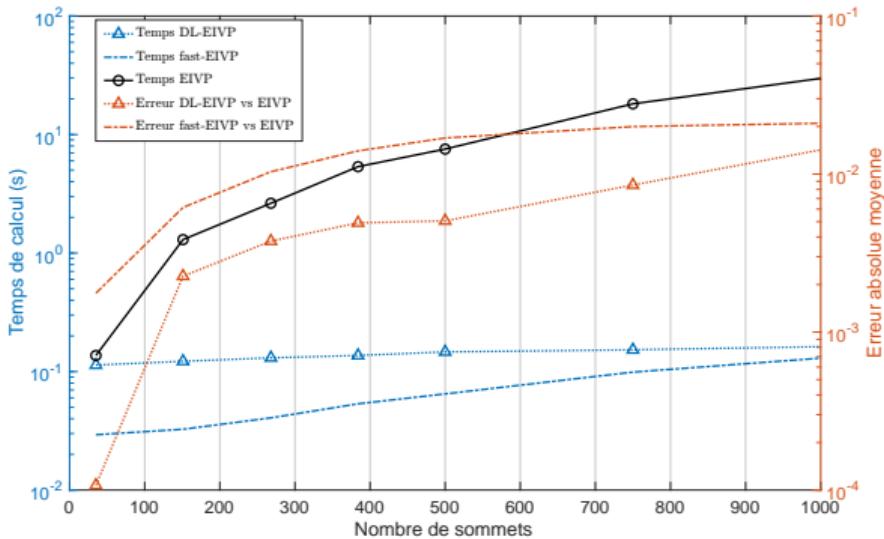


Figure – Temps de calcul (courbes bleues) et erreurs (courbes orange) des algorithmes EIVP sur des graphes $G_{n,500}$ pour $1 \leq n \leq 1000$.

► Algorithmique DL-EIVP

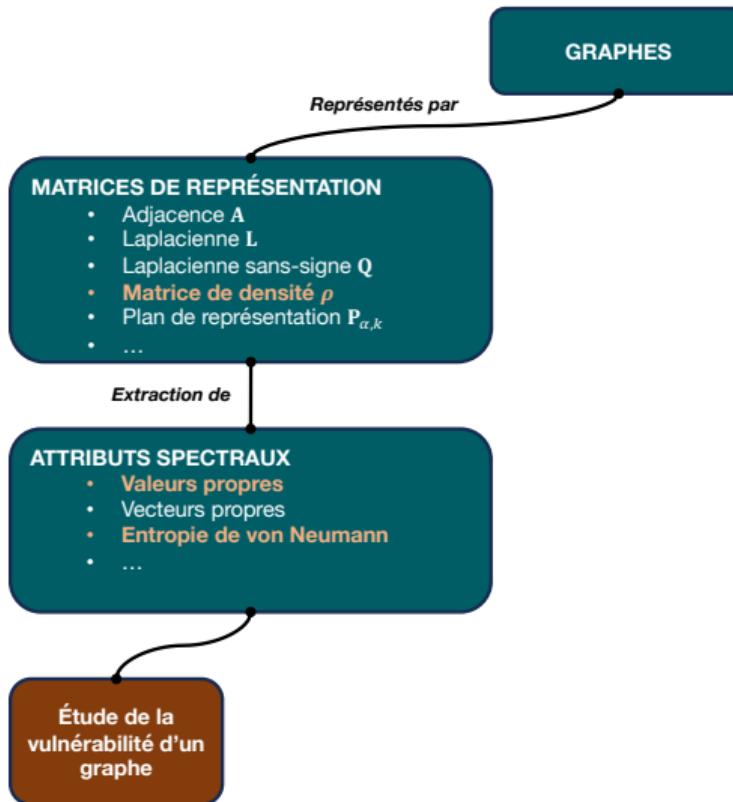
- ✓ Erreur commise « DL-EIVP vs EIVP »
- ✗ Temps de calcul : complexité en $O(m(n + m))$

► Algorithmique fast-EIVP

- ✗ Erreur commise « fast-EIVP vs EIVP »
- ✓ Temps de calcul : complexité en $O(n^3 + mn)$

Vulnérabilité informationnelle de réseaux

Conclusion de la partie



Plan

1 Contexte, enjeux et verrous scientifiques

2 Généralités sur les graphes

3 Vulnérabilité informationnelle de réseaux

- Entropie de von Neumann comme mesure de vulnérabilité d'une arête : Algorithme EIVP
- Optimisations de l'algorithme EIVP basées sur des approximations de l'entropie de von Neumann

4 Vers une nouvelle matrice de représentation de graphes ?

- Définition et propriétés d'un nouveau plan de représentation
- Nouvelle mesure de similarité spectrale : Application à la classification de graphes

5 Graphes de visibilité pour classer et caractériser des séries temporelles

- Représentation des graphes de visibilité dans un plan informationnel pour une estimation du coefficient de Hurst
- Classification de signaux grâce à une comparaison des distributions de degrés de leurs graphes de visibilité

6 Conclusions & Perspectives

Vers une nouvelle matrice de représentation de graphes ?

Une solution au problème de cospectralité

⚠ Un problème pour la classification spectrale de graphe : la **cospectralité**

Similarité Spectrale Conjointe⁹

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

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

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).

- ✓ Cette mesure de similarité spectrale apporte une réponse au problème de cospectralité
- ✓ $\beta = 1$: comparaison des spectres des matrices Laplaciennes / $\beta = 0$: comparaison des spectres d'adjacence
- ✗ Il est nécessaire de calculer **deux spectres**
- ✗ Selon les ordres n_1 et n_2 , cette mesure omet un certain nombre de valeurs propres

? Existe-t-il une mesure de similarité spectrale performante ne requérant qu'un seul spectre ?

9. H.A. Bay-Ahmed et al., A joint spectral similarity measure for graphs classification, *Pattern Recognition Letters*, 120:1–7, 2019

Vers une nouvelle matrice de représentation de graphes ?

Matrices de représentation généralisées

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

Matrices d'adjacence généralisées

💡 Matrice d'adjacence généralisée¹⁰ $\mathbf{G} := a\mathbf{A} + b\mathbf{I} + c\mathbf{J}$ (avec $a, b, c \in \mathbb{R}, a \neq 0$)

🚫 Contrainte : 3 paramètres

💡 Matrice d'adjacence universelle¹¹ $\mathbf{U} := a\mathbf{A} + b\mathbf{I} + c\mathbf{J} + d\mathbf{D}$ (avec $a, b, c, d \in \mathbb{R}, a \neq 0$)

🚫 Contrainte : 4 paramètres

💡 Matrice d' α -adjacence¹² $\mathbf{A}_\alpha := \alpha\mathbf{D} + (1 - \alpha)\mathbf{A}$ (avec $0 \leq \alpha \leq 1$)

🚫 Contrainte : Ne fait pas apparaître \mathbf{L} mais \mathbf{Q}

💡 Matrice α -Laplaciennne¹³ $\mathbf{L}_\alpha := \alpha\mathbf{D} + (\alpha - 1)\mathbf{A}$ (avec $0 \leq \alpha \leq 1$).

🚫 Contrainte : Ne fait pas apparaître \mathbf{A} mais $-\mathbf{A}$

10. E.R. Van Dam et al., Cospectral graphs and the generalized adjacency matrix, *Linear Algebra and its Applications*, 423(1):33-41, 2007

11. W.H. Haemers et G.R Omidi, Universal adjacency matrices with two eigenvalues, *Linear Algebra and its Applications*, 435(10):2520-2529, 2011

12. V. Nikiforov, Merging the \mathbf{A} - and \mathbf{Q} -spectral theories, *Applicable Analysis and Discrete Mathematics*, 11(1):81-107, 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*, 590:210-223, 2020

Vers une nouvelle matrice de représentation de graphes ?

Nouveau plan de représentation d'un graphe

Plan de représentation $\mathbf{P}_{\alpha,k}$ ¹⁴

$$\mathbf{P}_{\alpha,k} := \alpha \mathbf{D} + (2k - 1)(\alpha - 1) \mathbf{A}, \quad \alpha, k \in [0, 1] \quad (1)$$

Semi-définie positivité

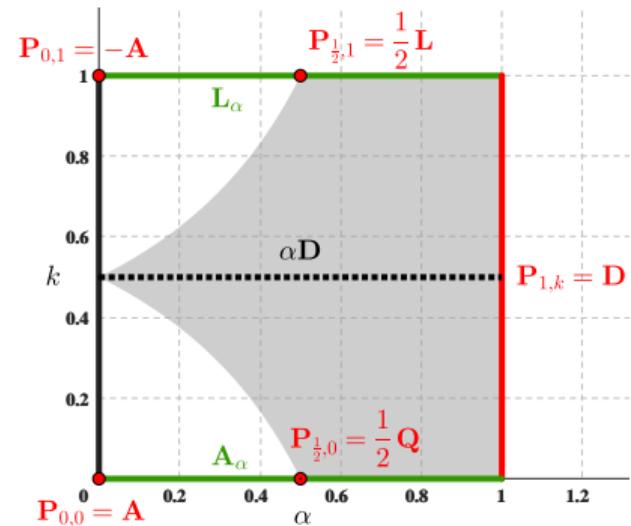
Soit G un graphe et $\mathbf{P}_{\alpha,k}$ sa matrice de représentation. Alors

$\mathbf{P}_{\alpha,k}$ est semi-définie positive si $\frac{2\alpha - 1}{2(\alpha - 1)} \leq k \leq \frac{1}{2(1 - \alpha)}$.

Superposition des spectres

Soit G un graphe. Si $\nu_\ell^{(\alpha,k)}$ (resp. $\nu_\ell^{(\alpha',k)}$) désigne la ℓ^{e} valeur propre de sa matrice $\mathbf{P}_{\alpha,k}$ (resp. $\mathbf{P}_{\alpha',k}$), alors, pour $\alpha \in [0, 1]$ et $\alpha' \in [\alpha, 1]$,

$$\nu_\ell^{(\alpha,k)} \leq \nu_\ell^{(\alpha',k)}, \quad \forall k \in [0, 1].$$



14. T. Averty et al., A New Family of Graph Representation Matrices : Application to Graph and Signal Classification, *IEEE Signal Processing Letters*, 31:2935–2939, 2024

Plan

1 Contexte, enjeux et verrous scientifiques

2 Généralités sur les graphes

3 Vulnérabilité informationnelle de réseaux

- Entropie de von Neumann comme mesure de vulnérabilité d'une arête : Algorithme EIVP
- Optimisations de l'algorithme EIVP basées sur des approximations de l'entropie de von Neumann

4 Vers une nouvelle matrice de représentation de graphes ?

- Définition et propriétés d'un nouveau plan de représentation
- Nouvelle mesure de similarité spectrale : Application à la classification de graphes

5 Graphes de visibilité pour classer et caractériser des séries temporelles

- Représentation des graphes de visibilité dans un plan informationnel pour une estimation du coefficient de Hurst
- Classification de signaux grâce à une comparaison des distributions de degrés de leurs graphes de visibilité

6 Conclusions & Perspectives

Vers une nouvelle matrice de représentation de graphes ?

Nouvelle mesure de similarité spectrale

Similarité spectrale par mesure de corrélation¹⁵

Soient deux graphes G_1 et G_2 d'ordre n . Si $\tilde{\nu}_{\alpha,k}(G_1)$ (resp. $\tilde{\nu}_{\alpha,k}(G_2)$) représente le **spectre standardisé** de la matrice $\mathbf{P}_{\alpha,k}(G_1)$ (resp. $\mathbf{P}_{\alpha,k}(G_2)$), alors une mesure de similarité entre les graphes G_1 et G_2 basée sur une distance construite à partir de la corrélation¹⁶ est définie comme suit :

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

ⓧ **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 $\mathbf{P}_{\alpha,k}$ et du **noyau de SVM** défini par les coefficients

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

15. T. Averty et al., Sur la similarité spectrale des graphes par mesure de corrélation, *GRETSI*, pages 1–4, 2023

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

Vers une nouvelle matrice de représentation de graphes ?

Bases de données de la littérature

- **MUTAG**¹⁷ : composés chimiques étiquetés en fonction de leur effet mutagène sur une bactérie
- **PTC_MR**¹⁸ : composés chimiques étiquetés en fonction de leur cancérogénicité sur des rats mâles
- **PROTEINS**¹⁹ : protéines classifiées selon qu'elles soient des enzymes ou non
- **IMDB-BINARY**²⁰ : réseaux ego de 1000 acteurs/actrices qui ont joué des rôles dans des films de 2 genres différents
- **IMDB-MULTI**²⁰ : réseaux ego de 1500 acteurs/actrices qui ont joué des rôles dans des films de 3 genres différents

	Nombre de graphes N	Nombre moyen de sommets \bar{n}	Nombre moyen d'arêtes \bar{m}
MUTAG	188	17.93	19.79
PTC_MR	344	14.29	14.69
PROTEINS	1113	39.06	72.82
IMDB-BINARY	1000	19.77	96.53
IMDB-MULTI	1500	13.00	65.94

Table – Statistiques descriptives des bases de données

17. A. K. Debnath et al., Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity, *Journal of Medicinal Chemistry*, 34(2):786–797, 1991

18. C. Helma et al., The predictive toxicology challenge 2000–2001, *Bioinformatics*, 17:107–108, 2001

19. K. M. Borgwardt et al., Protein function prediction via graph kernels, *Bioinformatics* 21:47–56, 2005

20. P. Yanardag et S. V. N. Vishwanathan, Deep Graph Kernel, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1365–1374, 2015

Vers une nouvelle matrice de représentation de graphes ?

Classification spectrale de graphes

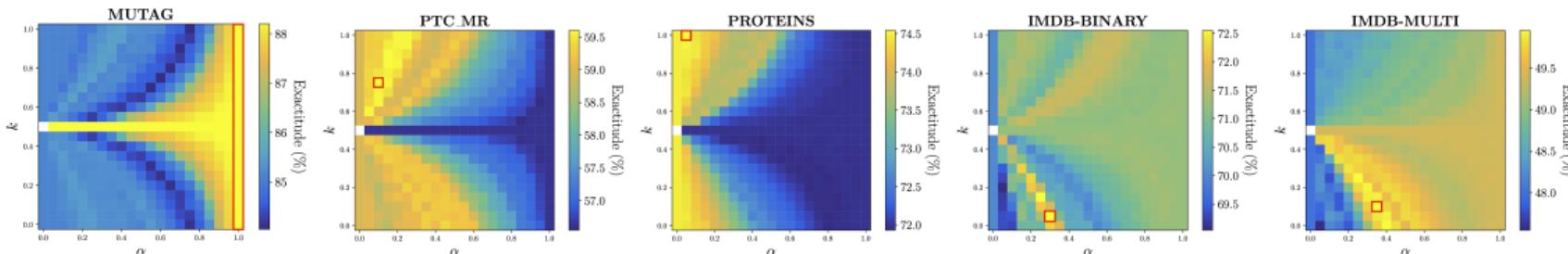


Figure – Plan $\mathbf{P}_{\alpha,k}$ dont la couleur représente l'exactitude moyenne de classification pour des couples de paramètres (α, k) , le pixel encadré en rouge correspondant au meilleur couple

	Meilleures matrices
MUTAG	$88.20 \pm 0.37\%$ pour $\mathbf{P}_{1,k}$
PTC_MR	$59.60 \pm 0.66\%$ pour $\mathbf{P}_{0.1,0.75}$
PROTEINS	$74.54 \pm 0.12\%$ pour $\mathbf{P}_{0.05,1}$
IMDB-BINARY	$72.55 \pm 0.47\%$ pour $\mathbf{P}_{0.3,0.05}$
IMDB-MULTI	$49.95 \pm 0.51\%$ pour $\mathbf{P}_{0.35,0.1}$

Table – Exactitudes obtenues pour la meilleure matrice $\mathbf{P}_{\alpha,k}$



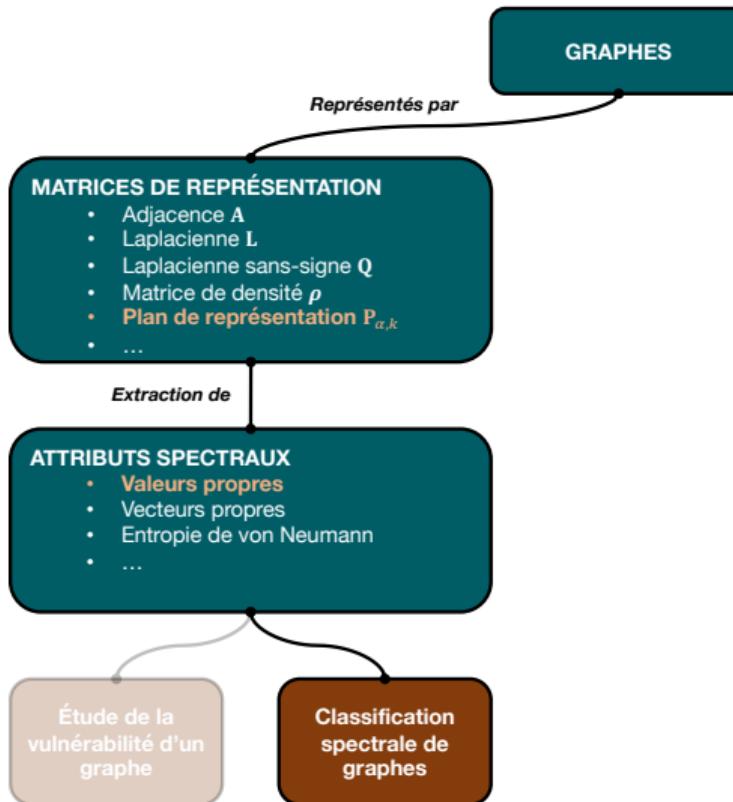
La meilleure matrice $\mathbf{P}_{\alpha,k}$ est différente pour chaque base de données



La meilleure matrice n'est pas une des matrices conventionnelles mais bien une combinaison de ces dernières

Vers une nouvelle matrice de représentation de graphes ?

Conclusion de la partie



Plan

1 Contexte, enjeux et verrous scientifiques

2 Généralités sur les graphes

3 Vulnérabilité informationnelle de réseaux

- Entropie de von Neumann comme mesure de vulnérabilité d'une arête : Algorithme EIVP
- Optimisations de l'algorithme EIVP basées sur des approximations de l'entropie de von Neumann

4 Vers une nouvelle matrice de représentation de graphes ?

- Définition et propriétés d'un nouveau plan de représentation
- Nouvelle mesure de similarité spectrale : Application à la classification de graphes

5 Graphes de visibilité pour classer et caractériser des séries temporelles

- Représentation des graphes de visibilité dans un plan informationnel pour une estimation du coefficient de Hurst
- Classification de signaux grâce à une comparaison des distributions de degrés de leurs graphes de visibilité

6 Conclusions & Perspectives

Graphes de visibilité pour classer et caractériser des séries temporelles

Graphes de visibilité naturelle et horizontale

► **Le graphe de visibilité naturelle**²¹ d'une série temporelle $x = (x_i)_{1 \leq i \leq n}$ est un graphe d'ordre n où $\{i, j\}$ existe si et seulement si l'échantillon x_k vérifie

$$x_k < x_j + (x_i - x_j) \frac{j - k}{j - i}, \quad \forall k \in \llbracket i, j \rrbracket$$

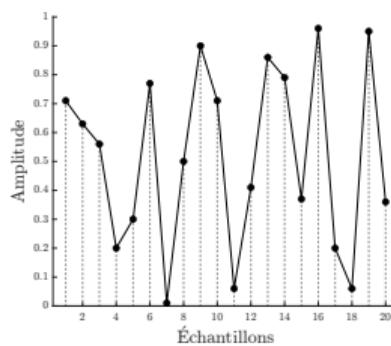


Figure – Construction du graphe de visibilité naturelle de x ($n = 20 / m = 40$)

► **Le graphe de visibilité horizontale**²² d'une série temporelle $x = (x_i)_{1 \leq i \leq n}$ est un graphe d'ordre n où $\{i, j\}$ existe si et seulement si l'échantillon x_k vérifie

$$x_k < \min(x_i, x_j), \quad \forall k \in \llbracket i, j \rrbracket$$

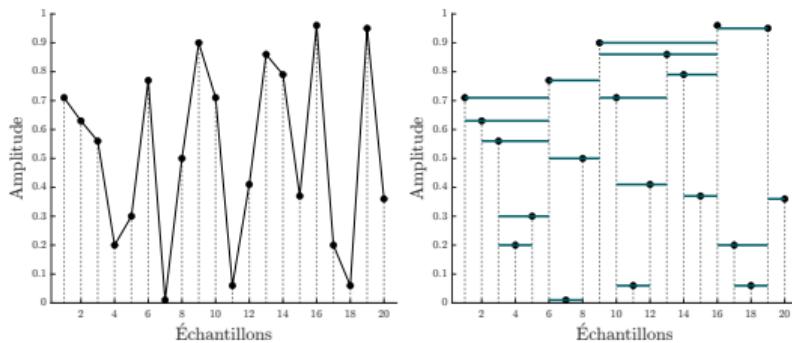


Figure – Construction du graphe de visibilité horizontale de x ($n = 20 / m = 33$)

21. L. Lacasa et al., From time series to complex networks : The visibility graph, *Proceedings of the National Academy of Sciences*, 105(13):4972–4975, 2008

22. B. Luque et al., Horizontal visibility graphs : Exact results for random time series, *Physical Review E*, 80(4):046103, 2009

Graphes de visibilité pour classer et caractériser des séries temporelles

Pondérations possibles d'un graphe de visibilité

► Il existe bien d'autres variantes du graphe de visibilité

- *Penetrable Visibility Graph*
- **Graphe de visibilité signée**
- ...

► Ajout possible de **pondérations**

Nom	Formule
Distance euclidienne	$w_{ij} = \sqrt{(j - i)^2 + (x_j - x_i)^2}$
Distance euclidienne au carré	$w_{ij} = (j - i)^2 + (x_j - x_i)^2$
Écart vertical	$w_{ij} = x_j - x_i$
Écart vertical absolu	$w_{ij} = x_j - x_i $
Écart horizontal	$w_{ij} = j - i$
Pente	$w_{ij} = \frac{x_j - x_i}{j - i}$
Pente absolue	$w_{ij} = \left \frac{x_j - x_i}{j - i} \right $
Angle de vue	$w_{ij} = \arctan \left(\frac{x_j - x_i}{j - i} \right)$
Angle de vue absolu	$w_{ij} = \arctan \left(\left \frac{x_j - x_i}{j - i} \right \right)$

Graphes de visibilité pour classer et caractériser des séries temporelles

Processus stochastiques étudiés

- **Mouvement Brownien fractionnaire (fBm)** $B_H(t)$ paramétré par le **coefficient de Hurst** H : unique processus Gaussien centré, nul en zéro, continu et dont la covariance est donnée par

$$\mathbb{E}[B_H(t)B_H(s)] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$$

- 💡 Quand $H \in (0, 1/2)$, la série est **anti-persistante** (une hausse est souvent suivie d'une baisse, et vice versa)
- 💡 Quand $H \in (1/2, 1)$, la série est **persistante** (une tendance haussière (ou baissière) a de fortes chances de continuer)
- 💡 Quand $H = 1/2$, le **mouvement Brownien standard** est obtenu

➤ Intérêt de ces processus stochastiques (et du H) ?

- ➡ Modélisation d'une très large classe de phénomènes physiques
- ➡ Utile pour l'analyse des **marchés financiers** (recherche de tendance ou de cycles)
- ➡ Utile pour les **séries hydrologiques** (flux de rivières, pluviométrie)

➤ Les accroissements de $B_H(t)$ sont **stationnaires** et le processus $G_H(t)$ constitué de ces accroissements est appelé **Bruit Gaussien fractionnaire (fGn)** :

$$G_H(t) = B_H(t) - B_H(t-1)$$

Graphes de visibilité pour classer et caractériser des séries temporelles

Intérêt de la distribution de degrés des graphes de visibilité des processus stochastiques

? Mais quel est le lien avec les graphes de visibilité ?

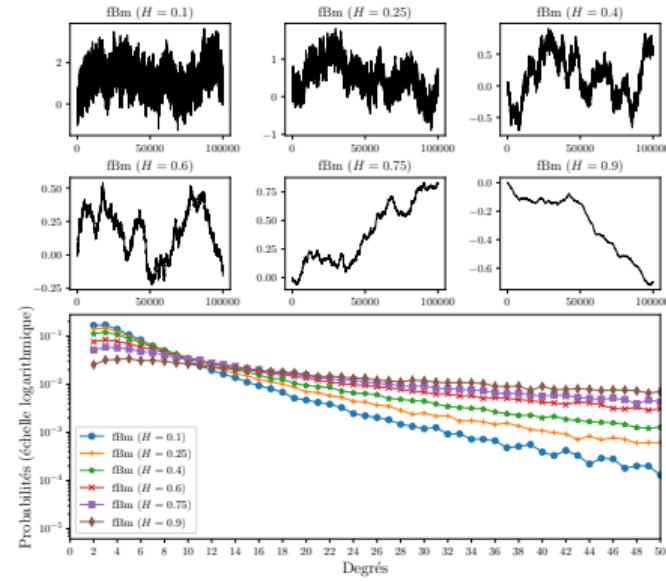


Figure – Distributions de degrés des graphes de visibilité associées à différents fBm

► Les **distributions de degrés** des graphes de visibilité sont riches d'information quant aux processus stochastiques sous-jacents

💡 Représentation de ces distributions dans le **plan informationnel de Fisher-Shannon**²³ (versions discrètes)

► Entropie de Shannon (quantité globale) :

$$S(\mathbf{p}) = - \sum_{i=1}^n p_i \log p_i$$

► Information de Fisher²⁴ (quantité locale) :

$$F(\mathbf{p}) = \frac{1}{2} \sum_{i=1}^{n-1} (\sqrt{p_{i+1}} - \sqrt{p_i})^2$$

23. C. Vignat et J.-F. Bercher, Analysis of signals in the Fisher-Shannon information plane, *Physics Letters A*, 312(1-2):27–33, 2003

24. P. Sánchez-Moreno et al., Discrete densities and Fisher information, *Proceedings of the 14th International Conference on Difference Equations and Applications*, pages 291–298, 2009.

Graphes de visibilité pour classer et caractériser des séries temporelles

Projection des distributions des degrés dans le plan Fisher-Shannon

► Méthode proposée par Gonçalves et al.²⁵

- Graphes de visibilité **horizontale**
- Distributions des degrés **entières**

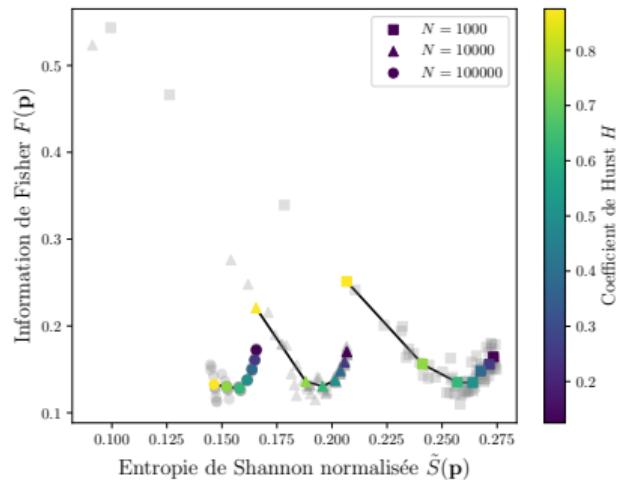


Figure – Méthode de Gonçalves pour représenter les distributions des degrés dans le plan FS

► Notre méthode

- Graphes de visibilité **naturelle**
- Distributions des degrés **tronquées**

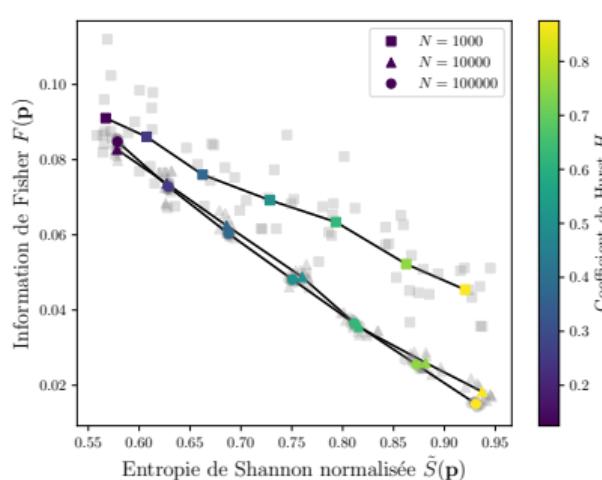


Figure – Notre méthode pour représenter les distributions des degrés dans le plan FS

💡 Si un squelette précis est construit, il est possible d'estimer le coefficient de Hurst d'un signal réel 💡

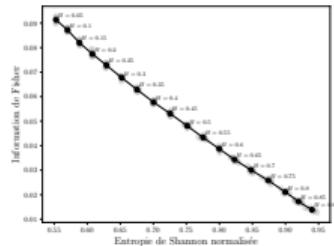
25. B. A. Gonçalves et al., Time series characterization via horizontal visibility graph and Information Theory, *Physica A : Statistical Mechanics and its Applications*, 464:93–102, 2016

Graphes de visibilité pour classer et caractériser des séries temporelles

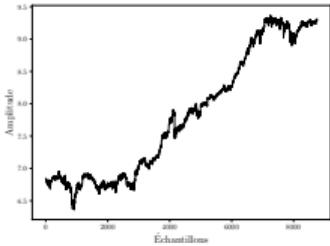
Une nouvelle méthode d'estimation du coefficient de Hurst

MÉTHODE VG/FS

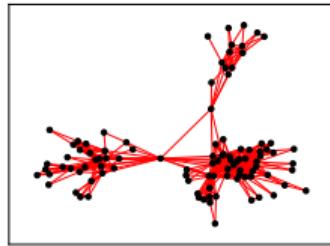
- ① Importer (ou construire) le squelette de référence dans le plan informationnel



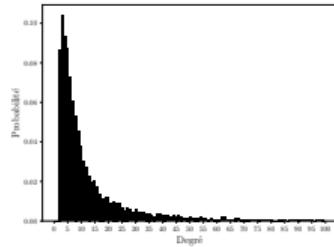
- ② Lire la série temporelle x à étudier



- ③ Construire le graphe de visibilité naturelle G associé à x

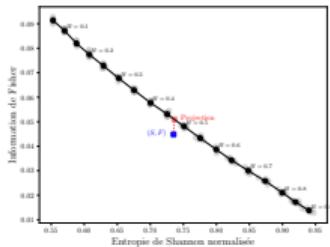


- ④ Extraire la distribution des degrés tronquée p du graphe G



- ⑤ Calculer le point (S, F) où F est l'information de Fisher et S est l'entropie de Shannon normalisée de la distribution de degrés tronquée p

- ⑥ Obtenir la projection orthogonale $(S_{\text{proj}}, F_{\text{proj}})$ de (S, F) sur le squelette



Graphes de visibilité pour classer et caractériser des séries temporelles

Résultats de notre méthode

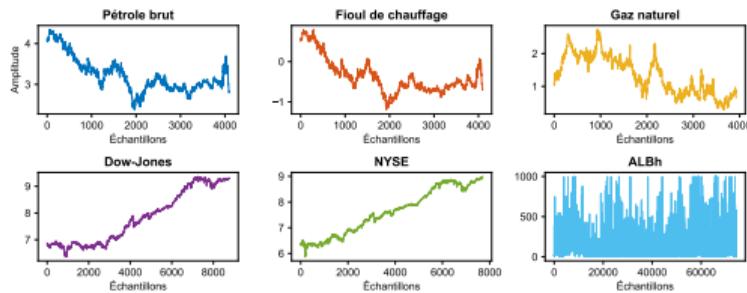


Figure – Signaux issus du monde de la finance (5 fBm et 1 fGn)

Méthodes	VG/FS	PSD ²⁶	WAV ²⁷	R/S ²⁸	DFA ²⁹
Pétrole brut (fBm)	0.472	0.423	0.452	0.498	0.466
Fioul de chauffage (fBm)	0.498	0.441	0.443	0.423	0.458
Gaz naturel (fBm)	0.512	0.433	0.465	0.654	0.481
Dow-Jones (fBm)	0.470	0.491	0.480	0.554	0.481
NYSE (fBm)	0.508	0.491	0.487	0.545	0.505
ALBh (fGn)	0.899	0.923	0.829	0.838	0.855

Table – Résultats de différents estimateurs du coefficient de Hurst

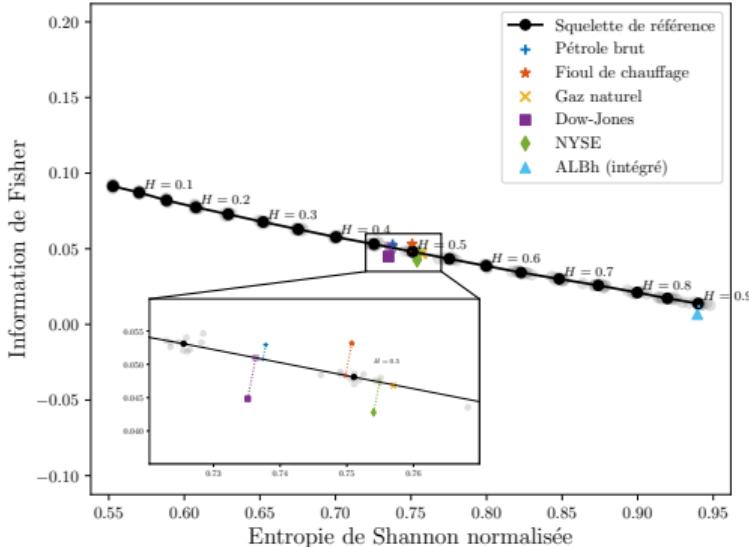


Figure – Méthode VG/FS appliquée aux 6 signaux

26. V. Marmelat et al., Relative roughness : an index for testing the suitability of the monofractal model, *Frontiers in Physiology*, 3:208, 2012

27. I. Simonsen et al., Determination of the Hurst exponent by use of wavelet transforms, *Physical Review E*, 58(3):2779, 1998

28. C.-F. Li, Rescaled-range and power spectrum analyses on well-logging data, *Geophysical Journal International*, 153:201–212, 2003

29. C.-K. Peng et al., Mosaic organization of DNA nucleotides, *Physical Review E*, 49(2):1685, 1994

Plan

1 Contexte, enjeux et verrous scientifiques

2 Généralités sur les graphes

3 Vulnérabilité informationnelle de réseaux

- Entropie de von Neumann comme mesure de vulnérabilité d'une arête : Algorithme EIVP
- Optimisations de l'algorithme EIVP basées sur des approximations de l'entropie de von Neumann

4 Vers une nouvelle matrice de représentation de graphes ?

- Définition et propriétés d'un nouveau plan de représentation
- Nouvelle mesure de similarité spectrale : Application à la classification de graphes

5 Graphes de visibilité pour classer et caractériser des séries temporelles

- Représentation des graphes de visibilité dans un plan informationnel pour une estimation du coefficient de Hurst
- Classification de signaux grâce à une comparaison des distributions de degrés de leurs graphes de visibilité

6 Conclusions & Perspectives

Graphes de visibilité pour classer et caractériser des séries temporelles

Tâche 1 – Détection d'épilepsie dans des signaux EEG

► Base de données **libre d'accès** de l'université de Bonn

► Signaux EEG **monocapteur**

► **5 sous-ensembles :**

- ↳ **Ensemble A** : Personnes saines avec les yeux ouverts
- ↳ **Ensemble B** : Personnes saines avec les yeux fermés
- ↳ **Ensemble C** : Personnes malades sans crise hippocampique
- ↳ **Ensemble D** : Personnes malades sans crise d'épilepsie
- ↳ **Ensemble E** : Personnes malades durant une crise d'épilepsie

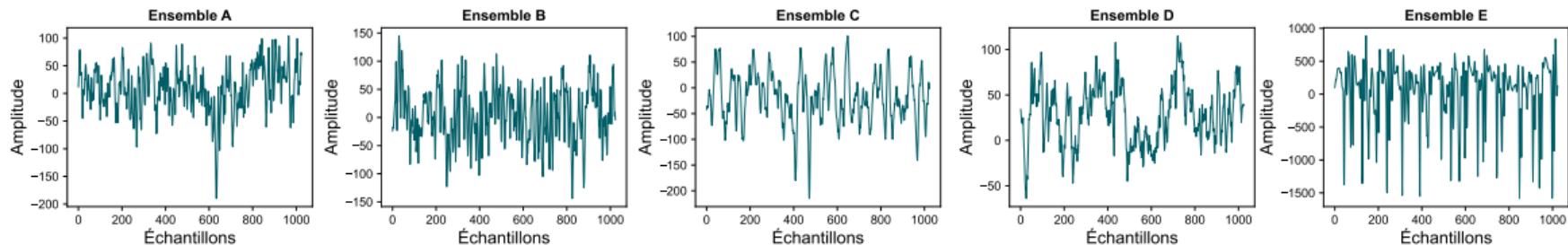


Figure – Exemples de signaux EEG de la base de données de l'université de Bonn

Graphes de visibilité pour classer et caractériser des séries temporelles

Travaux de la littérature

► Pour la tâche 1, les travaux de la littérature extraient et comparent un **nombre limité d'attributs**

→ **Travaux de Zhu et al.**³⁰ :

- GVH pondéré ($w_{ij} = |(x_i - x_j)(i - j) + 1|$)
- 2 attributs structurels

→ **Travaux de Supriya et al.**³¹ :

- GVN pondéré (angle de vue)
- 2 attributs structurels

→ **Travaux de Rajadurai et al.**³² :

- GVN pondéré (angle de vue)
- 3 attributs structurels

💡 Nous pourrions **extraire plus d'attributs** (et ça fonctionne)

? Est-il possible d'obtenir de bons résultats en ne comparant que les distributions des degrés ?

30. G. Zhu et al., Epileptic seizure detection in EEGs signals using a fast weighted horizontal visibility algorithm, *Computer Methods and Programs in Biomedicine*, 115(2):64–75, 2014

31. S. Supriya et al., Weighted visibility graph with complex network features in the detection of epilepsy, *IEEE Access*, 4:6554–6566, 2016

32. T. Rajadurai et al., Epileptic Seizure Prediction Using Weighted Visibility Graph, *International Conference on Soft Computing Systems (ICSCS)*, pages 453–461, 2018

Graphes de visibilité pour classer et caractériser des séries temporelles

Distances statistiques pour comparer les distribution de degrés

💡 Utilisation de distances statistiques 💡

- Soient deux distributions discrètes $\mathbf{p} = (p_i)_{1 \leq i \leq n}$ et $\mathbf{q} = (q_i)_{1 \leq i \leq n}$

💡 Distance en variation totale

$$d_{\text{VT}}(\mathbf{p}, \mathbf{q}) = 2 \sup_{1 \leq i \leq n} |p_i - q_i|$$

💡 Distance de Bhattacharyya

$$d_{\text{B}}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n \sqrt{p_i q_i}$$

💡 Distance de Hellinger

$$d_{\text{H}}(\mathbf{p}, \mathbf{q}) = \sqrt{\frac{1}{2} \sum_{i=1}^n (\sqrt{p_i} - \sqrt{q_i})^2}$$

💡 Distance de Jensen-Shannon

$$d_{\text{JS}}(\mathbf{p}, \mathbf{q}) = S(\mathbf{m}) - \frac{S(\mathbf{p}) + S(\mathbf{q})}{2}, \quad \mathbf{m} = \frac{\mathbf{p} + \mathbf{q}}{2}$$

- Soit \mathbf{p}_i (resp. \mathbf{p}_j) la **distribution des degrés du graphe de visibilité horizontale** construit à partir du i^{e} (resp. j^{e}) signal
- Entraînement d'un SVM à partir de la **matrice de Gram**

$$[\mathbf{K}]_{ij} = \exp(-\gamma d(\mathbf{p}_i, \mathbf{p}_j)), \quad 1 \leq i, j \leq N$$

Graphes de visibilité pour classer et caractériser des séries temporelles

Résultats de classification pour la tâche 1 dans le cas (D vs. E)

Distances statistiques	Exactitude	AUC
Jensen-Shannon	96.62 %	0.97
Variation totale	94.88 %	0.95
Bhattacharyya	96.62 %	0.97
Hellinger	97.00 %	0.97

Table – Comparaison des résultats de la méthode avec différentes distances statistiques

Méthodes	Exactitude	AUC
Zhu <i>et al.</i>	96.12 %	0.96
Supriya <i>et al.</i>	68.62 %	0.69
Rajadurai <i>et al.</i>	82.62 %	0.83
Méthode proposée	97.00 %	0.97

Table – Comparaison des résultats avec des méthodes de la littérature utilisant des GV

- Meilleure distance statistique : **Distance d'Hellinger**
- Utiliser **moins d'attributs mieux comparés** permet d'obtenir de meilleurs résultats

Graphes de visibilité pour classer et caractériser des séries temporelles

Tâche 2 – Détection d'anomalies magnétiques

- ▶ Signaux simulés par **DGA-TN**
- ▶ Signatures simulées à **5 différents SNR** pour traduire **différentes profondeurs d'immersion** de l'objet d'intérêt
- ▶ **2 sous-ensembles :**
 - **Ensemble B** : Signaux avec uniquement du bruit
 - **Ensemble S+B** : Signaux avec présence d'une signature

💡 Implémentation des méthodes utilisées pour la tâche 1

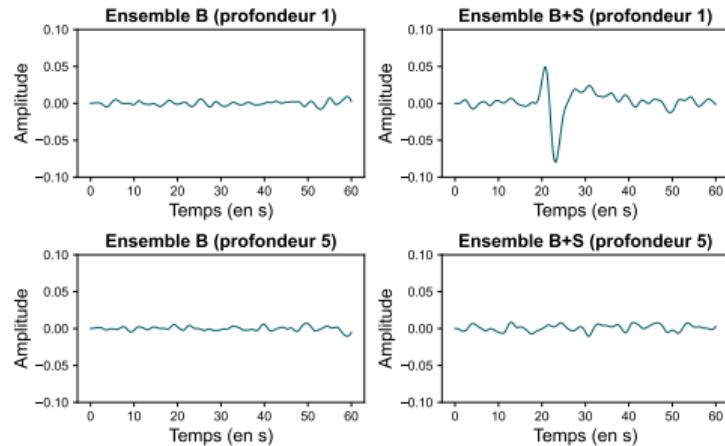


Figure – Exemples de signaux MAD de la base de données de DGA-TN

Graphes de visibilité pour classer et caractériser des séries temporelles

Résultats de classification pour la tâche 2

Distances statistiques	Exactitude	AUC
Jensen-Shannon	73.00 %	0.73
Variation totale	70.30 %	0.70
Bhattacharyya	73.85 %	0.74
Hellinger	73.10 %	0.73

Table – Comparaison des résultats de la méthode avec différentes distances statistiques

Méthodes	Exactitude	AUC
Zhu <i>et al.</i>	74.05 %	0.74
Supriya <i>et al.</i>	58.00 %	0.58
Rajadurai <i>et al.</i>	71.95 %	0.72
Méthode proposée	73.85 %	0.74

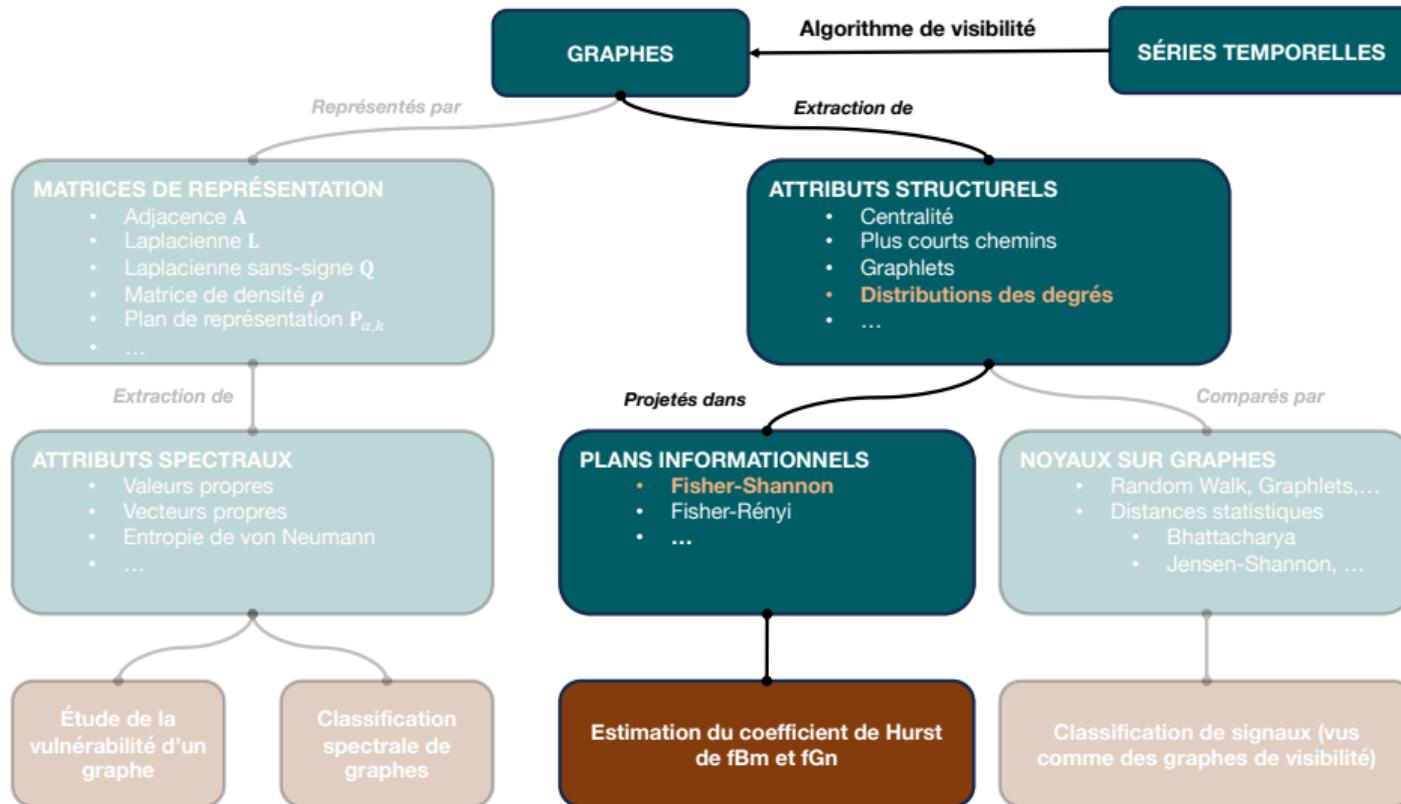
Table – Comparaison des résultats avec des méthodes de la littérature utilisant des GV

Meilleure distance statistique : **Distance de Bhattacharyya**

Utiliser **moins d'attributs mieux comparés** a ses limites pour certaines tâches

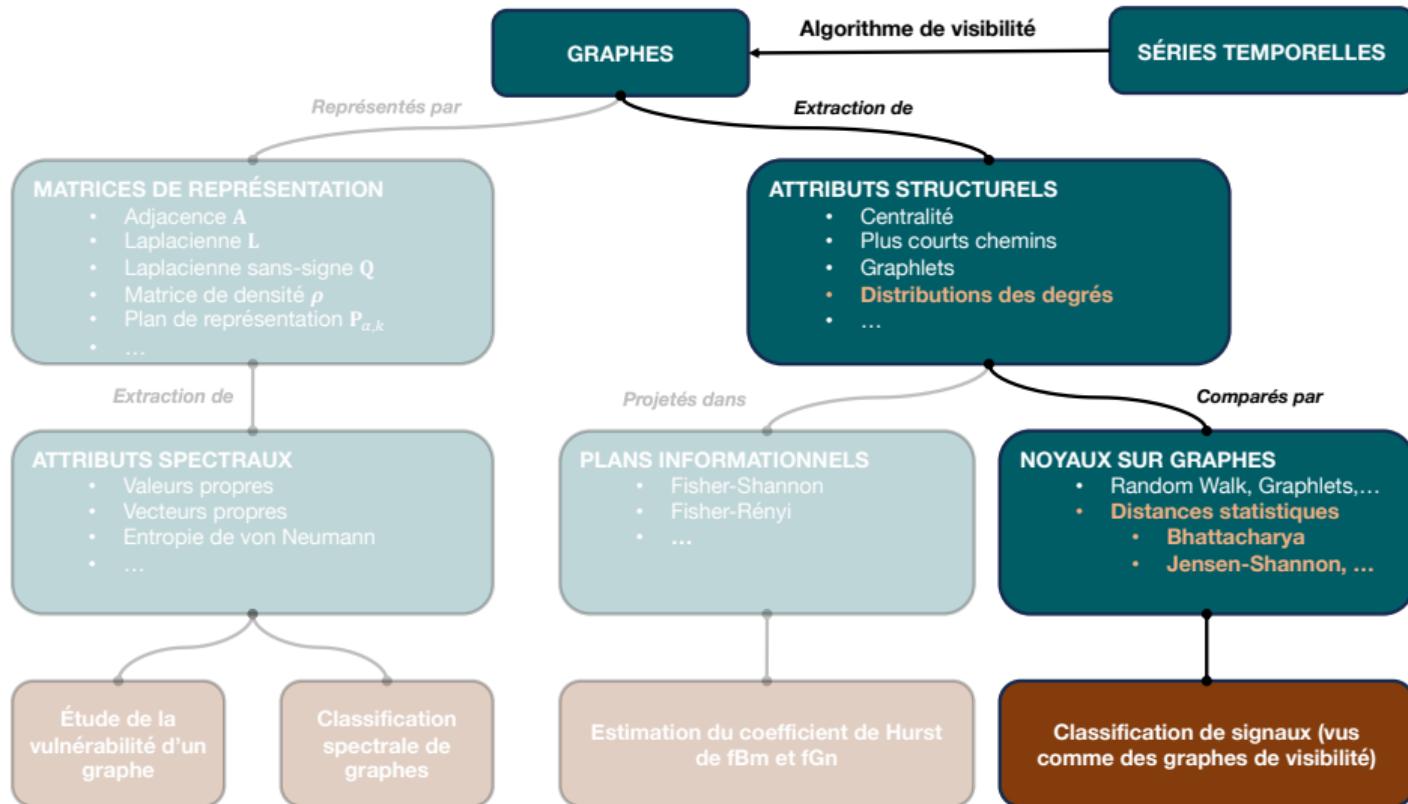
Graphes de visibilité pour classer et caractériser des séries temporelles

Conclusion de la partie



Graphes de visibilité pour classer et caractériser des séries temporelles

Conclusion de la partie



Conclusions & Perspectives

Conclusion – Contributions principales

🔍 Vulnérabilité informationnelle

- 📝 Interprétation de la **variation relative de l'entropie de von Neumann** comme **mesure de vulnérabilité** d'une arête
- 📝 Implémentation d'un **algorithme EIVP** qui **repondère les arêtes** grâce à leurs vulnérabilités informationnelles
- 📝 Proposition d'une **approximation de l'entropie de von Neumann** pour **optimiser l'algorithme EIVP** (approximation qui peut également s'appliquer à d'autres applications faisant intervenir cette entropie)

🔍 Matrices de représentation généralisées

- 📝 Introduction d'une **nouvelle matrice de représentation $P_{\alpha,k}$**
- 📝 Preuve de **deux propriétés algébriques** concernant $P_{\alpha,k}$
- 📝 Utilisation d'une **nouvelle mesure de similarité** basée sur une corrélation spectrale pour une **classification de graphes**

🔍 Graphes de visibilité

- 📝 Comparaison des **distributions de degrés** des **graphes de visibilité horizontale** à l'aide de **distances statistiques** permettant une **classification de signaux**
- 📝 Développement d'une méthode d'**estimation du coefficient de Hurst** de processus stochastiques basée sur la **projection des distributions de degrés tronquées** des **graphes de visibilité naturelle** dans le **plan informationnel de Fisher-Shannon**

Conclusions & Perspectives

Perspectives

Vulnérabilité informationnelle

- 🔍 Perturber un groupe d'arêtes
- 🔍 Développer un algorithme **NIVP** (*Node Informational Vulnerability to Perturbation*)
- 🔍 Utiliser d'autres métriques de la théorie de l'information (divergence de Kullback-Leibler, distances statistiques, etc.)

Matrices de représentation généralisées

- 🔍 Établir d'autres propriétés algébriques à la matrice $\mathbf{P}_{\alpha,k}$
- 🔍 Rechercher le couple (α, k) optimal
- 🔍 Faire le lien avec les travaux de Lutzeyer et Vazirgiannis (GNN)

Graphes de visibilité

- 🔍 Créer un graphe de visibilité prenant en compte l'enveloppe inférieure
- 🔍 Réfléchir à une troncature non arbitraire des distributions de degrés
- 🔍 Utiliser d'autres entropies globales (Havrda-Charvát, Burg, ...)
- 🔍 Développer une méthode d'estimation du coefficient de Hurst adaptée au fGn

Liste des publications

Actes de conférences nationales à comité de lecture

-  **T. Averty**, D. Daré-Emzivat et A.-O. Boudraa. Détection d'épilepsie dans les signaux EEG par graphe de visibilité et un noyau de SVM adapté. *GRETSI*, pages 1–4, 2022
-  **T. Averty**, D. Daré-Emzivat, A.-O. Boudraa et Y. Préaux. Approximation de l'entropie de von Neumann de graphes pour une analyse de vulnérabilité. *GRETSI*, pages 1–4, 2022
-  **T. Averty**, D. Daré-Emzivat et A.-O. Boudraa. Sur la similarité spectrale des graphes par mesure de corrélation. *GRETSI*, pages 1–4, 2023

Articles de journaux internationaux

-  **T. Averty**, A.-O. Boudraa et D. Daré-Emzivat. A New Family of Graph Representation Matrices : Application to Graph and Signal Classification. *IEEE Signal Processing Letters*, 31:2935–2939, 2024
-  **T. Averty**, A.-O. Boudraa et D. Daré-Emzivat. Hurst exponent estimation using natural visibility graph embedding in Fisher–Shannon plane. *Signal Processing*, 230:109884, 2025
-  **T. Averty**, H.-A. Bay-Ahmed, D. Daré-Emzivat, A.-O. Boudraa et C. Richard. Identifying vulnerable links in large networks using von Neumann graph entropy. *IEEE Transactions on Signal and Information Processing over Networks*, prochainement soumis

Merci à tous pour votre attention,

et merci encore à mes encadrants, collègues, services de soutien, secrétariats et directions de l'École navale.



Liste des 26 attributs pour la classification de graphes

Attributs	Valeurs pour G_1	Valeurs pour G_2
Nombre de sommets n	6	6
Nombre d'arêtes m	7	8
Degré minimal δ	1	2
Degré moyen \bar{d}	2.333	2.667
Degré maximal Δ	5	3
Force moyenne \bar{s}	17.667	53.667
Déviation des degrés $\text{dev}(G)$	5.333	2.667
Variance des degrés $\text{var}(G)$	1.556	0.22
Indice de Zagreb $Z_1(G)$	42	44
Indice de Zagreb $Z_2(G)$	53	60
Indice hyper-Zagreb $Z_h(G)$	264	244
Indice de Randić $R(G)$	2.712	2.966
Indice harmonique $H(G)$	1.238	1.467
<i>Inverse Sum Indeg index ISI(G)</i>	8.548	10.8
Rayon spectral λ_n	2.709	2.741
Écart spectral $\lambda_n - \lambda_{n-1}$	1.709	2.03
Énergie $E_A(G)$	7.806	8.139
Indice d'Estrada EE(G)	19.836	20.494
Connectivité naturelle $\bar{\lambda}(G)$	1.196	1.228
Valeur de Fiedler μ_2	1	1.586
Rayon spectral Laplacien μ_n	6	5
Indice de Kirchhoff $K(G)$	17	11.343
Énergie $E_L(G)$	10	8.828
<i>Laplacian Energy Like LEL(G)</i>	7.914	8.743
Énergie $E_Q(G)$	9.411	8.27
Énergie $E_L(G)$	3.306	3.064

Nombre de graphes cospectraux

Nombre de sommets n	Nombre de graphes N	Matrices de représentation		
		A	L	Q
1	1	0 (0)	0 (0)	0 (0)
2	2	0 (0)	0 (0)	0 (0)
3	4	0 (0)	0 (0)	0 (0)
4	11	0 (0)	0 (0)	2 (0.182)
5	34	2 (0.059)	0 (0)	4 (0.118)
6	156	10 (0.064)	4 (0.026)	16 (0.103)
7	1 044	110 (0.105)	130 (0.125)	102 (0.098)
8	12 346	1 722 (0.139)	1 767 (0.143)	1 201 (0.097)
9	274 668	51 039 (0.186)	42 595 (0.155)	19 001 (0.069)
10	12 005 168	2 560 516 (0.213)	1 412 438 (0.118)	636 607 (0.053)
11	1 018 997 864	215 264 372 (0.211)	91 274 836 (0.090)	38 966 935 (0.038)

Table – Nombre et proportion de graphes cospectraux selon différentes représentations (avec un ordre $n \in \llbracket 1, 12 \rrbracket$)³³

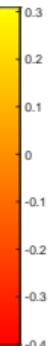
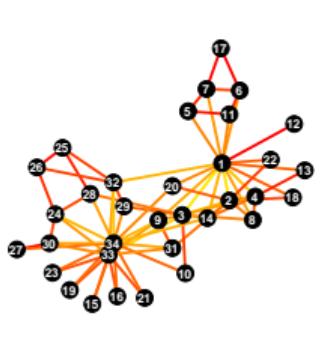
Nombre de sommets n	Nombre de graphes connectés N	Combinaisons de matrices de représentation		
		A \wedge L	L \wedge Q	Q \wedge A
6	112	0	0	0
7	853	0	16	0
8	11 117	0	232	0
9	261 080	82	4139	8
10	11 716 571	13 864	107 835	10 716

Table – Nombre de graphes (avec $n \in \llbracket 6, 10 \rrbracket$) cospectraux en combinant les spectres de différentes matrices³⁴

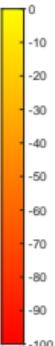
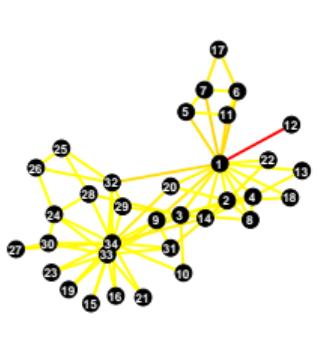
33. W.H. Haemers et E. Spence, Enumeration of cospectral graphs, *European Journal of Combinatorics*, 25(2):199–211, 2004

34. R.C. Wilson et P. Zhu, A study of graph spectra for comparing graphs and trees, *Pattern Recognition*, 41(9):2833–2841, 2008

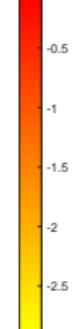
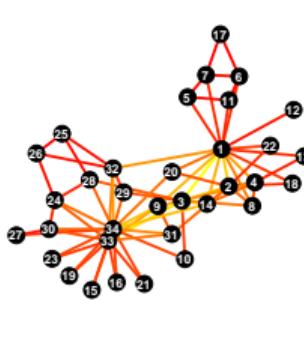
Corrélation de la vulnérabilité informationnelle avec d'autres mesures



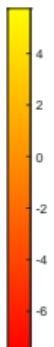
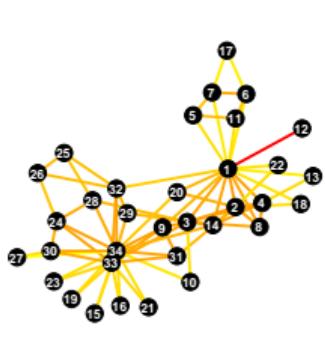
(a) Variation relative de l'entropie



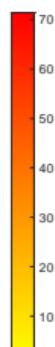
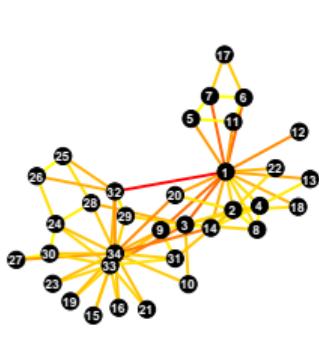
(b) Évolution de la valeur de Fiedler : $r = 0.17$



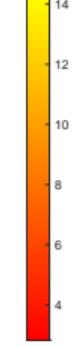
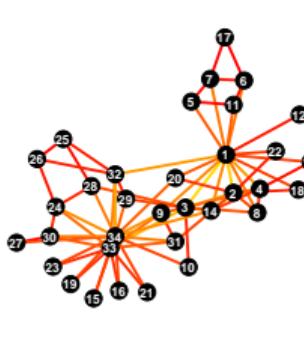
(c) Évolution de la *Shield Value* : $r = -0.9$



(d) Évolution de l'indice de Kirchhoff : $r = -0.28$



(e) Centralité d'intermédiairité des arêtes : $r = 0.38$



(f) Moyenne géométrique des degrés : $r = 0.97$

Résultats pour la tâche 1 dans le cas (D vs. E)

Tests	Distances statistiques	Exactitude	AUC
A vs. E	Jensen-Shannon	99.50 %	1.00
	Variation totale	99.00 %	0.99
	Bhattacharyya	99.25 %	0.99
	Hellinger	99.50 %	1.00
B vs. E	Jensen-Shannon	96.88 %	0.97
	Variation totale	92.12 %	0.92
	Bhattacharyya	96.62 %	0.97
	Hellinger	96.88 %	0.97
C vs. E	Jensen-Shannon	99.75 %	1.00
	Variation totale	98.50 %	0.99
	Bhattacharyya	99.88 %	1.00
	Hellinger	99.75 %	1.00
D vs. E	Jensen-Shannon	96.62 %	0.97
	Variation totale	94.88 %	0.95
	Bhattacharyya	96.62 %	0.97
	Hellinger	97.00 %	0.97

Table – Comparaison des résultats de la méthode avec différentes distances statistiques

Tests	Méthodes	Exactitude	AUC
A vs. E	Zhu et al.	99.00 %	0.99
	Supriya et al.	84.88 %	0.85
	Rajadurai et al.	99.12 %	0.99
	Méthode proposée	99.50 %	1.00
B vs. E	Zhu et al.	93.25 %	0.93
	Supriya et al.	81.75 %	0.82
	Rajadurai et al.	93.25 %	0.93
	Méthode proposée	96.88 %	0.97
C vs. E	Zhu et al.	98.50 %	0.99
	Supriya et al.	75.12 %	0.75
	Rajadurai et al.	92.00 %	0.92
	Méthode proposée	99.88 %	1.00
D vs. E	Zhu et al.	96.12 %	0.96
	Supriya et al.	68.62 %	0.69
	Rajadurai et al.	82.62 %	0.83
	Méthode proposée	97.00 %	0.97

Table – Comparaison des résultats avec des méthodes de la littérature utilisant des GV

Comparaison de la mesure de similarité CorS par rapport à d'autres noyaux

- Le noyau **SSC**³⁵ de Similarité Spectrale Conjointe
- Le noyau **SP**³⁶ comparant les plus courts chemins dans les graphes étudiés
- Le noyau **RW**³⁷ comparant les marches aléatoires dans les graphes étudiés
- Le noyau **GK**³⁸ dénombrant les sous-graphes types dans les graphes étudiés
- Le noyau **WL**³⁹ basé sur le test d'isomorphisme de Weisfeiler-Lehman

	CorS	SSC	SP	RW	GK	WL
MUTAG	$88.2 \pm 0.4\%$	$84.3 \pm 1.1\%$	$83.6 \pm 1.2\%$	$88.3 \pm 0.5\%$	$87.0 \pm 0.6\%$	$85.6 \pm 0.9\%$
PTC_MR	$59.6 \pm 0.7\%$	$57.8 \pm 2.1\%$	$58.2 \pm 1.9\%$	$54.2 \pm 1.8\%$	$55.6 \pm 0.3\%$	$62.3 \pm 1.0\%$
PROTEINS	$74.5 \pm 0.1\%$	$70.0 \pm 0.4\%$	$71.4 \pm 0.4\%$	-	$71.2 \pm 0.4\%$	$74.4 \pm 0.5\%$
IMDB-BINARY	$72.6 \pm 0.5\%$	$67.5 \pm 0.9\%$	$69.8 \pm 1.3\%$	$47.9 \pm 1.4\%$	-	$72.7 \pm 1.2\%$
IMDB-MULTI	$50.0 \pm 0.5\%$	$48.3 \pm 0.4\%$	$49.3 \pm 0.6\%$	$29.3 \pm 0.7\%$	-	$50.8 \pm 0.5\%$

Table – Résultats de classification (exactitudes en % moyennées sur 10 itérations \pm écarts-types) de différents noyaux sur des bases de données de la littérature.

35. H.A. Bay-Ahmed et al., A joint spectral similarity measure for graphs classification, *Pattern Recognition Letters*, 120:1–7, 2019

36. K.M. Borgwardt et H.P. Kriegel, Shortest-path kernels on graphs, In *5th IEEE International Conference Data Mining*, pages 1–8, 2005

37. T. Gärtner et al., On graph kernels : Hardness results and efficient alternatives, In *16th Annual Conference on Learning Theory and 7th Kernel Workshop*, pages 129–143, 2003

38. N. Pržulj, Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):177–183, 2007

39. N. Shervashidze et al., Weisfeiler-Lehman graph kernels, *Journal of Machine Learning Research*, 12(9):2539–2561, 2011