

Quantification de l'étendue d'un signal sur graphe

Antoine MAZARGUIL^{1,2}, Laurent OUDRE³, Nicolas VAYATIS^{1,2}

¹CMLA (ENS Paris Saclay, CNRS) 61 Avenue du Président Wilson, 94230 Cachan, France

²COGNAC-G (Université Paris Descartes, CNRS, SSA) 45 Rue des Saints-Pères, 75006 Paris, France

³L2TI (Université Paris 13) 99 avenue Jean Baptiste Clément 93430 Villetaneuse, France

antoine.mazarguil@cmla.ens-cachan.fr, laurent.oudre@univ-paris13.fr
nicolas.vayatis@cmla.ens-cachan.fr

Résumé – Cet article présente une nouvelle métrique pour la quantification de l'étendue d'un signal sur graphe. Par analogie avec les réseaux de capteurs physiques, nous décrivons une expression mathématique de cette notion qui se base sur la représentation des noeuds du graphe dans un espace euclidien (graph embedding). La métrique proposée évalue de manière efficace si le phénomène décrit par le signal sur graphe est localisé ou non. Différentes propriétés et applications sont présentées, ainsi que des résultats sur des données de synthèse et réelles.

Abstract – This article presents a new metric for quantifying the spread of a graph signal. By analogy with physical sensor networks, we describe a mathematical formulation for this notion, that is based on the concept of graph embedding. The proposed metric efficiently measures if the phenomenon captured by a graph signal is localized or not. Several properties and use cases for this metric are presented, along with practical results on synthetic and real data. In particular, this metric shows promising results for the characterization of graph signals.

1 Introduction

Au cours des dernières années, l'utilisation croissante des réseaux de capteurs a permis de produire une quantité importante de données structurées [1]. Le Traitement de Signal sur Graphe (GSP) est apparu comme un paradigme permettant d'intégrer ces structures et dépendances aux outils usuels de traitement du signal (échantillonnage [2], filtrage [3], interpolation [4]). Dans le cadre de cette théorie, un signal sur graphe est un ensemble de valeurs associées aux noeuds d'un graphe.

Afin de décrire les propriétés de tels signaux, plusieurs métriques ont été introduites par la communauté du GSP. Par analogie avec la transformée de Fourier usuelle, une transformée de Fourier sur graphe (GFT) a été définie, apportant une analyse spectrale des signaux et introduisant un concept de fréquences spatiales [1]. De façon analogue à l'analyse classique, un signal composé de basses fréquences correspond à un signal dont les valeurs varient peu et de façon régulière entre noeuds voisins sur le graphe, tandis que les signaux composés de hautes fréquences présentent de fortes variations entre voisins. Comme dans l'analyse de signaux temporels, la notion de largeur de bande d'un signal sur graphe est une propriété intéressante. Elle permet notamment, pour les signaux à largeur de bande finie, de s'attaquer à différentes problématiques, comme la compression, la reconstruction [4] et le débruitage [3] de signaux sur graphes.

Dans cet article, nous proposons une métrique permettant d'évaluer l'étendue d'un signal sur graphe. Par analogie avec des réseaux de capteurs physiques, nous introduisons une quan-

tification mathématique de la localisation spatiale du signal sur graphe. Cette propriété s'avère utile dans plusieurs contextes comme le filtrage, l'échantillonnage et la reconstruction.

2 Métrique proposée

Les signaux sur graphes sont particulièrement utiles pour mesurer des phénomènes structurés sur des réseaux (diffusion, propagation, anomalies). Dans un grand nombre de situations, il est pertinent de caractériser l'étendue d'un tel phénomène (ou de manière analogue sa localité). En effet, cette notion est cruciale par exemple pour déterminer si des anomalies se présentent sur une portion localisée du graphe ou si celles-ci corrompent tous les noeuds. L'étendue du phénomène peut être estimée directement à partir des intensités du signal sur graphe mesuré. Nous proposons dans cette partie une définition intuitive de l'étendue d'un signal sur graphe ainsi qu'une approche mathématique permettant sa quantification.

2.1 Précédentes définitions

Une quantification de l'étendue d'un signal sur un graphe a été proposée dans les études sur les principes d'incertitudes sur graphe [5].

$$\Delta_G(s) = \frac{1}{\|s\|_2^2} \min_{u \in V(G)} \sum_{v \in V(G)} d_G(u, v) s_v^2$$

Si cette métrique quantifie avec efficacité la notion d'étendue (elle vérifie par exemple, pour des distances strictement

positives, les propriétés 1 à 4 présentées en partie 3.1), son utilisation dans le cadre du traitement de signal sur graphe est rendu difficile pour trois raisons :

- Elle nécessite le calcul de la matrice géodésique
- Son expression non matricielle ne permet pas son utilisation dans des problèmes de minimisation
- Sa définition est basée sur les distances, là où l'approche du traitement de signal sur graphe préfère les affinités. La génération de distances pour l'utilisation de cette métrique d'étendue constitue une problématique à part entière [6]

2.2 Analogie avec un réseau de capteurs physiques

Lorsque l'on considère un réseau de capteur physique fixe, pour lequel nous disposons des coordonnées de chaque noeud dans le plan ou dans l'espace, la notion d'étendue du signal est naturelle et intuitive. Celle-ci peut être quantifiée par la variance de la position des noeuds actifs, possiblement pondérée par les activations du signal à ces noeuds. Cet article propose d'étendre cette notion pour n'importe quel graphe, en particulier lorsque les affinités entre les noeuds ne sont pas nécessairement inférées par des positions dans l'espace. Pour cela, nous proposons d'effectuer un plongement du graphe dans un espace \mathbb{R}^d , qui sera assimilable à un espace physique dans lequel cette quantité pourra aisément être évaluée.

2.3 Prérequis et notations

Un graphe pondéré \mathcal{G} est un triplet (V, E, W) composé d'un ensemble fini de noeuds $V = \{v_1, \dots, v_N\}$, d'un ensemble d'arêtes $E \subset V \times V$ et d'une matrice de poids positifs $W \in \mathbb{R}_+^{N \times N}$ dont les éléments $W_{i,j}$ quantifient l'affinité entre les noeuds v_i et v_j . Les poids $W_{i,j}$ sont égaux à 0 pour $(v_i, v_j) \notin E$, et le graphe est dit non dirigé quand $W = W^T$. Enfin, le graphe est dit connexe si il existe un chemin d'éléments de E entre toute paire de noeuds. Le Laplacien du graphe \mathcal{G} est défini par $L^{\mathcal{G}} = D - W$, avec D la matrice diagonale des degrés du graphe, i.e $D = \text{Diag}(W\mathbf{1})$. Un signal réel sur le graphe \mathcal{G} est une fonction $V \rightarrow \mathbb{R}$. Il se représente par un vecteur $s \in \mathbb{R}^N$, pour lequel s_i est le signal positionné au noeud v_i . Etant donné un signal sur graphe $s \in \mathbb{R}^N$, on peut définir la distribution de probabilité P_s telle que $\forall i, P_s(i) \geq 0$ et $\sum_{i=1}^N P_s(i) = 1$. Cette distribution peut s'interpréter comme la contribution du noeud i donnée par l'intensité du signal s . Enfin, une d -représentation du graphe est une fonction $f : v_i \rightarrow X_i^{\mathcal{G}} \in \mathbb{R}^d$ qui maintient une notion de proximité entre les espaces de départ et d'arrivée. Une d -représentation du graphe \mathcal{G} peut être assimilée à une matrice $X^{\mathcal{G}} \in \mathbb{R}^{d \times N} = (X_1^{\mathcal{G}}, \dots, X_N^{\mathcal{G}})$, de telle sorte que l'on a $f(v_i) = X_i^{\mathcal{G}} = X^{\mathcal{G}} e_i$.

2.4 Étendue d'un signal sur graphe

Soit \mathcal{G} un graphe connexe et $X^{\mathcal{G}}$ une d -représentation de \mathcal{G} telle que :

- (1) La représentation est centrée : $X^{\mathcal{G}}\mathbf{1} = 0_D$
- (2) La représentation est injective : $\forall i \neq j, X_i^{\mathcal{G}} \neq X_j^{\mathcal{G}}$
- (3) La représentation est normalisée : $\sum \|X_i^{\mathcal{G}}\|^2 = N$

Considérons un signal s sur le graphe \mathcal{G} et la distribution de probabilité P_s associée. On peut définir grâce à P_s les statistiques usuelles induites par s , telle que l'espérance :

$$\mathbb{E}_s[f(X)] = \sum_{i=1}^N f(X_i^{\mathcal{G}}) P_s(i). \quad (1)$$

La première étape dans le calcul de la variance de la position des noeuds consiste à exprimer l'espérance de la position des noeuds dans l'espace de représentation :

$$\mathbb{E}_s[X] = \sum_{i=1}^N P_s(i) X_i^{\mathcal{G}} = \sum_{i=1}^N P_s(i) X^{\mathcal{G}} e_i = X^{\mathcal{G}} P_s \quad (2)$$

L'espérance de la position des noeuds constitue d'ores et déjà une statistique intéressante caractérisant le signal. Celle-ci nous donne une indication sur le positionnement du signal dans l'espace, permettant ainsi de définir la position moyenne d'une anomalie.

Notre définition de l'étendue est alors donnée par la variance des coordonnées des noeuds dans l'espace de représentation. Une définition équivalente est l'espérance du carré de la norme des positions recentrées :

$$S_{X^{\mathcal{G}}}(s) = \mathbb{E}_s[\|X - \mathbb{E}_s[X]\|^2]. \quad (3)$$

Cette quantité peut être exprimée à partir de la matrice de Gram de la représentation, donnant ainsi la définition ci-après.

Définition (étendue d'un signal sur graphe). Etant donné un graphe \mathcal{G} , une représentation (embedding) $X^{\mathcal{G}}$ et une distribution de probabilité P_s associé à un signal s , l'étendue du signal s est définie comme suit :

$$S_{X^{\mathcal{G}}}(s) = P_s^T (w - C P_s), \quad (4)$$

avec

- $C = (X^{\mathcal{G}})^T (X^{\mathcal{G}})$ La matrice de Gram de la représentation $X^{\mathcal{G}}$.
- $w = (\|X_1^{\mathcal{G}}\|^2 \dots \|X_N^{\mathcal{G}}\|^2)^T$ le vecteur colonne des normes au carré de la représentation (embedding).

2.5 Considérations pratiques

Dans cet article la représentation (embedding) choisie est celle des espaces propres du Laplacien (Laplacian Eigenmaps) avec les $d = 10$ espaces propres associés aux premières valeurs propres non nulles. La distribution de probabilité P_s sera obtenue grâce à une normalisation L_k , avec $k = 1$ ou $k = 2$ selon les utilisations.

3 Propriétés et applications

Dans cette section, nous présentons quelques propriétés vérifiées par la métrique proposée ainsi que quelques applications possibles.

3.1 Propriétés

Prop 1 (Positivité).

$$\forall s \in \mathcal{R}^N, S_X(s) \geq 0 \quad (5)$$

Cette propriété est triviale étant donnée la définition de l'étendue en tant qu'espérance d'une quantité positive.

Prop 2 (Normalisation). Si le signal de distribution a été généré par une normalisation L_k , on a

$$\forall s \in \mathcal{R}^N, \forall \lambda \in \mathcal{R}^*, S_X(\lambda s) = S_X(s) \quad (6)$$

Cette propriété d'invariance montre que la métrique proposée décrit la distribution du signal sur son support, indépendamment de l'ordre de grandeur du signal.

Prop 3 (Etendue d'un Dirac).

$$\forall i \in [1, N], S_X(e_i) = 0 \quad (7)$$

Prop 4 (Ensemble des signaux d'étendue nulle). Dans le cas d'une représentation (embedding) injective, on a :

$$S_X(s) = 0 \implies \exists i, s = e_i \quad (8)$$

Si le support d'un signal est réduit à un seul noeud, la variance obtenue est nulle. Si par contre on considère un signal dont au moins deux points différents présentent des valeurs non nulles, la variance est strictement positive.

Prop 5 (Etendue d'un signal uniforme). Pour une représentation (embedding) centrée,

$$S_X(\mathbf{1}) = \frac{1}{N} \text{trace}((X^G)^T(X^G)) = \frac{1}{N} \|X^G\|^2 \quad (9)$$

Les propriétés (3), (4) et (5) nous permettent de définir une échelle d'étendue. Dans le cas d'une représentation normalisée de telle sorte que $\|X^G\|^2 = N$, la métrique définie varie de 0 pour un signal très localisé à 1 pour un signal très étendu.

Prop 6 (Distance de paire de noeuds). Soient v_1 et v_2 deux noeuds tirés au hasard selon la distribution P_s . On a alors :

$$\mathbb{E}_s[|X_{v_1} - X_{v_2}|^2] = 2 S_X(s) \quad (10)$$

Cette propriété présente un lien plus direct entre la proximité des noeuds sur le graphe et la métrique d'étendue. Dans le cas où la représentation préserve la proximité, une faible étendue est interprétable par une faible distance sur le graphe entre les fortes intensités du signal.

3.2 Applications possibles

L'expression de l'étendue comme métrique de quantification d'un signal peut s'avérer fort utile pour différentes problématiques de traitement de signal sur graphe.

- **Echantillonnage** : Lors d'un échantillonnage sur graphe, une analyse de l'étendue des noeuds sélectionnés peut être effectuée afin de quantifier leur répartition sur le graphe. Notre notion peut alors être utilisée comme contrainte sur ces noeuds, de façon à favoriser les échantillonnage à forte étendue qui sont probablement plus robustes.
- **Ondelettes sur graphe** : La construction d'ondelettes sur graphe requiert une quantification de l'étendue d'un atome tant sur le graphe que dans le domaine spectral. Notre approche mathématique pour l'évaluation de l'étendue peut être une piste intéressante pour chercher un compromis entre localisation spatiale et fréquentielle de telles familles.
- **Débruitage** : De façon analogue à la reconstruction de signal régulier en GSP, il est possible de s'intéresser à la reconstruction d'un signal localisé. Pour l'étude d'un phénomène local superposé à un bruit global, une reconstruction est possible en ajoutant la métrique d'étendue à la fonction de coût à optimiser, ne sélectionnant ainsi qu'un support réduit pour le signal débruité.
- **Décomposition de signal** : Des événements locaux simultanés peuvent être mesurés par des réseaux de capteurs, et isoler ces composantes spatialement distinctes est une tâche complexe. La décomposition d'un tel signal peut être obtenue en minimisant l'étendue de ces composantes, sous la contrainte que leur somme reconstruise le signal initial.

4 Expérimentations et résultats

Dans cette section, la notion d'étendue est utilisée pour caractériser deux types de signaux classiquement utilisés : un processus de diffusion et des ondelettes sur graphe, mais également sur des données réelles d'activité neuronale.

4.1 Processus de diffusion sur un nuage de point

La métrique d'étendue est utilisée pour suivre l'évolution un processus de diffusion sur un graphe généré à partir d'un nuage de point représentant un lapin (Stanford Bunny). Le processus est initialisé par un Dirac situé sur un noeud (signal localisé), et est diffusé à chaque étape selon un opérateur de diffusion sur ses voisins.

L'étendue du signal en fonction de l'étape de diffusion est présentée sur la figure 1 pour trois initialisations différentes (patte, dos et oreille du lapin).. Les valeurs d'étendue commencent à 0 et tendent vers 1, ce qui concorde avec les propriétés de l'étendue et de la diffusion. Le profil d'étendue change en fonction du noeud initial, et laisse supposer que le choix d'un noeud optimal pour une diffusion large ne sera pas le même selon le terme de l'objectif recherché. Dans ce cas particulier, les oreilles du lapin sont composées d'un grand nombre de noeuds qui sont difficiles d'accès pour une diffusion de source extérieure. Ainsi, commencer une diffusion dans une telle compo-

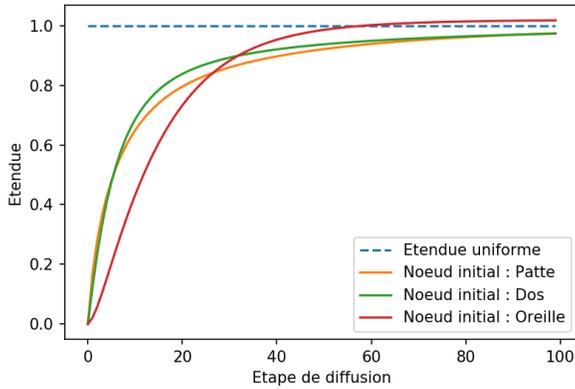


FIGURE 1 – Evolution de l'étendue d'un signal au cours d'une diffusion sur un nuage de point

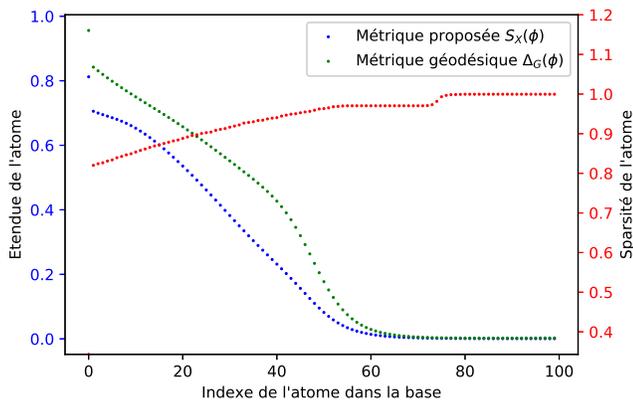


FIGURE 2 – Etendue et sparsité d'une base d'ondelettes sur graphe en fonction de l'échelle de l'atome

sante n'est pas efficace sur le court terme, mais permet une plus grande diffusion sur le long terme. Cet exemple permet de penser à l'étendue comme métrique permettant l'évaluation de qualité d'un processus de diffusion.

4.2 Étendue d'une base d'ondelette

L'étude locale des signaux sur graphes est souvent réalisée grâce à des bases d'ondelettes. La quantification de l'étendue d'une base d'ondelettes sur graphes est une problématique importante, généralement traitée en calculant la parcimonie des atomes générés. Nous proposons d'évaluer la localité de chaque signal d'ondelette grâce à l'étendue définie précédemment.

Notre approche est illustrée sur le graphe "Stanford Bunny". La base d'ondelettes utilisée est la base d'ondelettes spectrale [7] centrée sur le noeud 0, situé sur le dos du lapin. La figure 2 présente l'étendue des différents atomes de la famille, l'indice de l'atome étant analogue à l'inverse de l'échelle sur laquelle l'atome opère. La variation simultanée de la métrique

d'étendue et de la parcimonie apparaît clairement, permettant de comprendre que les atomes se rapprochent d'un Dirac avec la réduction de l'échelle. On observe que l'étendue évolue de manière plus régulière que la parcimonie, permettant ainsi une quantification plus claire de l'étendue. On peut noter également que ces deux notions ne sont pas équivalentes : on peut trouver des signaux locaux et non parcimonieux, comme des signaux globaux et parcimonieux.

4.3 Signaux d'activité neuronale

Un champ d'application important du GSP est l'analyse de signaux d'activité neuronale. Afin d'évaluer la capacité de notre métrique à caractériser un signal, nous avons utilisé la base de donnée GIPSA-lab Neuroimaging [8]. Celle-ci est composée de signaux d'activation mesurés sur 100 patients sains ayant effectué 2 tests successifs acquis par IRM fonctionnelle. Le graphe associé est composé de 89 noeuds correspondant à des régions d'intérêt du cerveau. Pour chaque acquisition, le signal considéré est le différentiel temporel de l'activation et nous avons calculé la variance de la métrique d'étendue de ce signal. La question qui nous intéresse ici est de savoir si les étendues associées aux deux tests d'un même sujet étaient plus similaires que celles associées à deux tests de sujets différents. On trouve que l'écart moyen de variance d'étendue entre deux acquisitions d'un même sujet est de l'ordre de 0.13, contre 0.29 pour l'écart moyen sur toutes les acquisitions. Le coefficient de corrélation interclasse (ICC) associé est de 0.773, ce qui démontre que cette métrique constitue bien une caractéristique associée au patient, démontrant ainsi la pertinence de l'utilisation de la métrique d'étendue pour l'étude de signaux sur graphe.

Références

- [1] A. Ortega, P. Frossard, J. Kovačević, J. Moura, and P. Vandergheynst, "Graph signal processing : Overview, challenges, and applications," *Proceedings of the Institute of Electrical and Electronics Engineers (IEEE)*, pp. 808–828, 2018.
- [2] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *The Applied and Computational Harmonic Analysis*, vol. 44, no. 2, pp. 446–475, 2018.
- [3] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovacevic, "Signal denoising on graphs via graph filtering," in *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2014, pp. 872–876.
- [4] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Transaction on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.
- [5] A. Agaskar and Y. M. Lu, "A spectral graph uncertainty principle," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4338–4356, 2013.
- [6] B. Pasdeloup, R. Alami, V. Gripon, and M. Rabbat, "Toward an uncertainty principle for weighted graphs," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 1496–1500.
- [7] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "Signal processing on graphs : Extending high-dimensional data analysis to networks and other irregular data domains," *Computing Research Repository (CoRR)*, vol. abs/1211.0053, 2012.
- [8] M. Termenon, A. Jaillard, C. Delon-Martin, and S. Achard, "Reliability of graph analysis of resting state fmri using test-retest dataset from the human connectome project," *Neuroimage*, vol. 142, pp. 172–187, 2016.