# Metrics of graph Laplacian eigenvectors

*Haotian Li & Naoki Saito*

Department of Mathematics
University of California, Davis

SPIE Conference on Wavelets and Sparsity XVIII
San Diego, CA
August 15, 2019

# Outline

# Acknowledgment


- Support from NSF grants: DMS-1418779 and DMS-1912747
- Support from `Julia` community who helped our computational and graphical issues

# Outline

# Motivations: Problematic Ordering

- Key of building 1D wavelets by the Littlewood-Paley theory is based on *the ordering of Fourier modes*.
- Using *graph Laplacian eigenvectors* as “cosines” or Fourier modes on graphs with eigenvalues as (the square of) their “frequencies” has been popular.
- **Spectral Graph Wavelet Transform (SGWT)** of Hammond et al. derived wavelets on a graph based on the Littlewood-Paley theory that *organized the graph Laplacian eigenvectors corresponding to dyadic partitions of eigenvalues* by viewing the eigenvalues as “frequencies”.
- This view may face difficulty for graphs more complicated than very simple undirected unweighted paths and cycles.

# Motivations: Natural Ordering

- Therefore, we design "metrics" of graph Laplacian eigenvectors to detect the *"behavioral differences"* between them so that we can order the eigenvectors more naturally than using the size of the corresponding eigenvalues.
- Goal: Define proper "metrics" between the eigenvectors such that similar behavior ones are close and distinct behavior ones are far apart.
- The usual $\ell^2$-distance doesn't work since $\left\|\boldsymbol{\phi}_i - \boldsymbol{\phi}_j\right\|_2 = \sqrt{2}\delta_{ij}$.
- Furthermore, these metrics *help us design smooth multiscale basis dictionaries* that are quite important for many applications, e.g., efficiently approximating graph signals and solving differential equations on graphs.

# Outline

# Facts about Graph Laplacian $L(G)$

- Connected undirected graph $G = (V, E, W)$, with $|V| = n$ and $|E| = m$.
- Graph Laplacian is given by $L(G) = D(G) - W(G)$, in which $W(G)$ is the weights matrix of $G$ and $D(G)_{ii} = \sum_j W(G)_{ij}$.
- $L(G)$ is a real symmetric positive semi-definite matrix, so the eigenvalues of $L$ (i.e., $L(G)$) are *nonnegative* and the eigenvectors $\{\boldsymbol{\phi}_l\}_{l=0}^{n-1}$ form *an orthonormal basis*.

$$L\boldsymbol{\phi}_l = \lambda_l \boldsymbol{\phi}_l, \qquad 0 = \lambda_0 < \lambda_1 \le \cdots \le \lambda_{n-1}$$

- $\lambda_0 = 0$ is always an eigenvalue of $L$ and its corresponding eigenvector $\boldsymbol{\phi}_0$ is a constant vector called *the DC component* (vector).
- The eigenvector $\boldsymbol{\phi}_1$ (with the first nonzero eigenvalue) is called *the Fielder vector* which plays an important role in graph partitioning.
- Also, $\{\boldsymbol{\phi}_l\}_{l=0}^{n-1}$ and $\{\lambda_l\}_{l=0}^{n-1}$ commonly viewed as the Fourier modes on graphs and the corresponding "frequencies".

# Outline

# Review: Ramified Optimal Transportation (ROT) Metric

- First, by taking elementwise square, we convert each eigenvector to a probability mass function (pmf) $\boldsymbol{\phi}_i^2$ on the input *undirected* graph $G = (V, E, W)$ with $|V| = n$ and $|E| = m$.
- Define the ROT metric between a pair of the eigenvectors *by the minimal cost to move the probability mass from one pmf $\boldsymbol{\phi}_i^2$ to the other pmf $\boldsymbol{\phi}_j^2$.*
- To do so, we first orient the edges in $E(G)$ in an arbitrary manner to form a directed graph $\tilde{G}$ and compute its incidence matrix $Q := [\boldsymbol{q}_1|\boldsymbol{q}_2|\cdots|\boldsymbol{q}_m] \in \mathbb{R}^{n\times m}$. Here, $\boldsymbol{q}_k$ represents the endpoints of $e_k$ : if $e_k$ connect from vertex $i$ to vertex $j$, then

$$q_k[l] = \begin{cases} -1 & \text{if } l = i; \\ 1 & \text{if } l = j; \\ 0 & \text{otherwise.} \end{cases}$$

# Review: ROT Metric Proposed by Saito (2018)

- For undirected graph G, we form *bidirected* graph $\tilde{\tilde{G}}$ with $\tilde{Q} = [Q| - Q]$.
- Given $\tilde{Q}$, we solve the *balance equation* (underdetermined),

$$\tilde{Q}\boldsymbol{w} = \boldsymbol{\phi}_j^2 - \boldsymbol{\phi}_i^2, \qquad \boldsymbol{w} \in \mathbb{R}_{\geq 0}^{2m}, \tag{1}$$

- Note that any $\boldsymbol{w}$ satisfying Eq. (1) represents a transportation path (or plan) from $\boldsymbol{\phi}_i^2$ to $\boldsymbol{\phi}_j^2$, and there may be multiple solutions.
- Define the cost of a transport path $P \in \mathrm{Path}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2)$ as:

$$\boldsymbol{M}_\alpha(P) := \sum_{e \in E(P)} w(e)^\alpha \,\mathrm{length}(e), \quad \alpha \in [0,1].$$

- Then, define *ROT metric* between $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ as:

$$d_{\mathrm{ROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha) := \min_{P \in \mathrm{Path}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2)} \boldsymbol{M}_\alpha(P).$$

- For $\alpha = 1$, $\min_P \boldsymbol{M}_1(P)$ becomes the optimal transport cost.

# Review: Hadamard (HAD) Product Affinity Measure Proposed by Cloninger and Steinerberger (2018)

- On a compact Riemannian manifold $(\mathcal{M}, g)$, the *HAD affinity measure* between eigenfunctions is defined as:

$$a_{\mathrm{HAD}}(\phi_i, \phi_j)^2 := \|\phi_i \phi_j\|_2^{-2} \int_{\mathcal{M}} \left( \int_{\mathcal{M}} p(t,x,y)(\phi_i(y) - \phi_i(x))(\phi_j(y) - \phi_j(x)) \mathrm{d}y \right)^2 \mathrm{d}x$$
$$= \frac{\|e^{t\Delta}(\phi_i \phi_j)\|_{L^2}^2}{\|\phi_i \phi_j\|_{L^2}^2}$$

  where $(\lambda_i, \phi_i)_i$ is an eigenpair of the Laplace-Beltrami operator $\Delta$ on $\mathcal{M}$, $p(t,x,y)$ is the classical heat kernel, and the value of $t$ should satisfy $e^{-t\lambda_i} + e^{-t\lambda_j} = 1$.

- It can be interpreted as *a global average of local correlation* between these two eigenfunctions.

# Outline

# Outline

# Time-Stepping Diffusion (TSD) Metric

- The purpose of TSD metric is to design an optimal transport-like metric that *depends on time*. In other words, at each given time, we have a cost scheme (or metric).
- Given a time $T$, we consider the heat diffusion process on the graph. We want to measure *the cost of "flatten" the initial graph signal via diffusion process up to the time $T$*.
- We expect the graph signal will be flatten out by this process and *the final cost ,as $T \to \infty$, behave similar with the $d_{\mathrm{ROT}}(\alpha = 1)$.*
- Notations: Denote the factorization of graph Laplacian matrix as $L = \Phi \Lambda \Phi^T$, in which $\Phi = [\boldsymbol{\phi}_0 | \boldsymbol{\phi}_1 | \cdots | \boldsymbol{\phi}_{n-1}]$ and $\Lambda = \mathrm{diag}([\lambda_0, \lambda_1, \cdots, \lambda_{n-1}])$; the incidence matrix as $Q \in \mathbb{R}^{n \times m}$, which is treated as the graph gradient, i.e., $Q^{\mathsf{T}} = \nabla_G$.

# TSD Cost Functional $K$:

**(Heat diffusion)** Given initial $\boldsymbol{f}_0$, the governing ODE system which describes the graph signal $\boldsymbol{u}(t)$'s ($\in \mathbb{R}^n$) evolution is following:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}(t) + L \cdot \boldsymbol{u}(t) = \mathbf{0} \quad t \geq 0, \qquad \boldsymbol{u}(0) = \boldsymbol{f}_0 \in \mathbb{R}^n$$

Since $\{\boldsymbol{\phi}_0, \cdots, \boldsymbol{\phi}_{n-1}\}$ forms an ONB of $\mathbb{R}^n$, we can get the general solution:

$$\boldsymbol{u}(t) = \sum_{k=0}^{n-1} \langle \boldsymbol{f}_0, \boldsymbol{\phi}_k \rangle e^{-\lambda_k t} \boldsymbol{\phi}_k$$

At a certain time T, we define the following *TSD cost functional*:

$$K(\boldsymbol{f}_0, T) := \int_0^T \|\nabla_G \boldsymbol{u}(t)\|_1 \mathrm{d}t \qquad \nabla_G \text{ is the graph gradient.}$$

which can be interpreted as *accumulated total variation* of $\boldsymbol{u}(t)$.

# Convergence of TSD Cost and TSD Metirc

- We can show that $\lim_{T\to\infty} K(\boldsymbol{f}_0, T) < \infty$ for any $\boldsymbol{f}_0 \in \mathbb{R}^n$.
- After setting the input signal $\boldsymbol{f}_0 = \boldsymbol{\phi}_i - \boldsymbol{\phi}_j$, we define the *TSD metric between the eigenvectors at time T* by

$$d_{\mathrm{TSD}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j; T) := K(\boldsymbol{f}_0, T)$$

- Furthermore, we can show that for any $T > 0$ (including $T = \infty$), $K(\cdot, T)$ is a norm on $\mathscr{L}_0^2(V) := \{\boldsymbol{f} \in \mathscr{L}^2(V) | \sum_{x\in V} \boldsymbol{f}(x) = 0\}$.
- Cost Conjecture: As $T \to \infty$, we expect

$$d_{\mathrm{ROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha = 1) \le d_{\mathrm{TSD}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; T = \infty) \le C(G) \cdot d_{\mathrm{ROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha = 1)$$

where $C(G)$ is a constant depending on the graph $G$.

# Outline

# Difference of Absolute Gradient (DAG) Pseudometric

- The idea of DAG is that we use the absolute gradient of each eigenvector as its feature vector describing its behavior.
- We define *DAG pseudometric* as:

$$d_{\text{DAG}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j) := \| |\nabla_G| \boldsymbol{\phi}_i - |\nabla_G| \boldsymbol{\phi}_j \|_2 = \| \operatorname{abs}.(Q^\top \boldsymbol{\phi}_i) - \operatorname{abs}.(Q^\top \boldsymbol{\phi}_j) \|_2$$

- Further, we derive the following equations:

$$\begin{aligned} d_{\text{DAG}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)^2 &= \langle |\nabla_G| \boldsymbol{\phi}_i - |\nabla_G| \boldsymbol{\phi}_j, |\nabla_G| \boldsymbol{\phi}_i - |\nabla_G| \boldsymbol{\phi}_j \rangle_E \\ &= \langle |\nabla_G| \boldsymbol{\phi}_i, |\nabla_G| \boldsymbol{\phi}_i \rangle_E + \langle |\nabla_G| \boldsymbol{\phi}_j, |\nabla_G| \boldsymbol{\phi}_j \rangle_E - 2 \langle |\nabla_G| \boldsymbol{\phi}_i, |\nabla_G| \boldsymbol{\phi}_j \rangle_E \\ &= \lambda_i + \lambda_j - \sum_{x \in V} \sum_{y \sim x} |\boldsymbol{\phi}_i(x) - \boldsymbol{\phi}_i(y)| \cdot |\boldsymbol{\phi}_j(x) - \boldsymbol{\phi}_j(y)| \end{aligned}$$

  in which $\langle \cdot, \cdot \rangle_E$ is the inner product over edges.
- The last term of the formula can be viewed as *a global average of absolute local correlation* between eigenvectors, which is close to the interpretation of HAD affinity measure.
- Given the eigenvectors, *the computational cost is $O(|E|)$*.

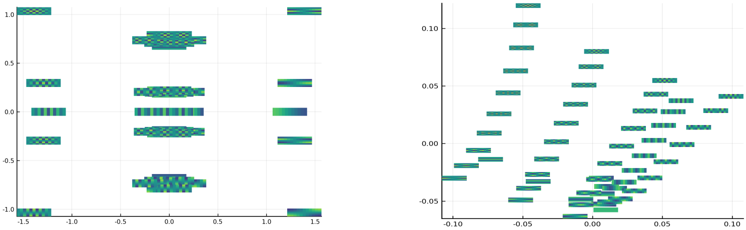
# Outline

# Numerical Experiments

- To evaluate the performance of those "metrics" for a given graph, we assemble the *distance matrix* by the mutual behavioral difference between the eigenvectors (or corresponding pmfs, e.g., $\boldsymbol{\phi}_i^2$ for $d_{\mathrm{ROT}}$) using each "metric".
- Then use the *classical MDS (Multidimensional Scaling)* on the distance matrix and embed the eigenvectors into the low dimensional Euclidean space, i.e., $\mathbb{R}^2$ or $\mathbb{R}^3$.
- By doing so, we can get the *visual arrangement* of eigenvectors organized by the corresponding "metric".

# Outline

# 2D Lattice $P_{11} \times P_5$: $d_{\mathrm{ROT}}$ and $a_{\mathrm{HAD}}$

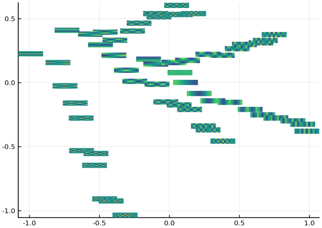



(a) $d_{\mathrm{ROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha = 1)$ (b) $a_{\mathrm{HAD}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)$

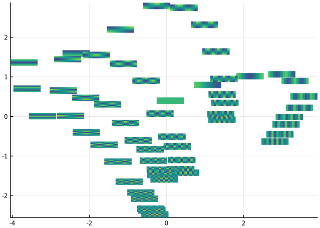
Figure: 2D-MDS embedding of the eigenvectors of $11 \times 5$ unweighted lattice graph based on the ROT and the HAD metrics: each small heatmap plot describes how the eigenvector looks like on the lattice graph.

- They both reveal the *two-dimensional ordering* of the eigenvectors.
- $a_{\mathrm{HAD}}$ is better but still has a little misordering in $y$ (vertical) direction.

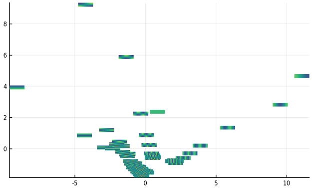
# 2D Lattice $P_{11} \times P_5$: $d_{\text{TSD}}$ with different $T$



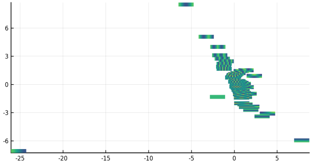

(a) $T = 0.1$




(b) $T = 1$

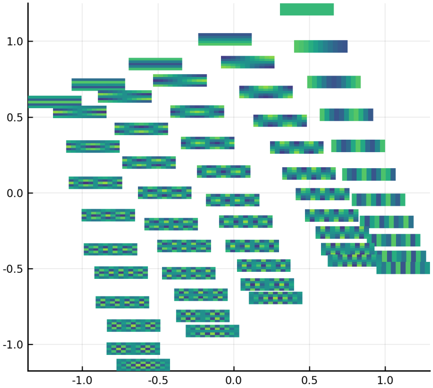



(c) $T = 10$




(d) $T = \infty$

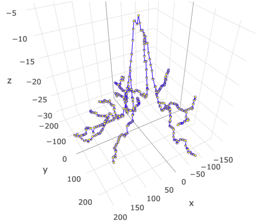
# 2D Lattice $P_{11} \times P_5$: $d_{\text{DAG}}$



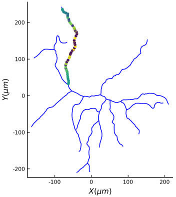

- $d_{\text{DAG}}$ nicely detect two directions of the oscillations. The eigenvectors are organized in *2D array*.
- For each column of the array, the eigenvectors have the same oscillation pattern in $y$ direction and oscillation in $x$ direction increases linearly. Vice versa.

# Outline

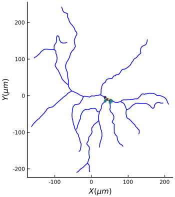
# RGC#100: Two Types of Eigenvectors




(a)




(b)




(c)

- (a): The 3D dendritic tree of RGC#100 graph.
- (b): The representative of eigenvectors with *semi-global oscillations* on the upper-left branch (projected in $\mathbb{R}^2$).
- (c): The representative of eigenvectors with much more *localized active support* around junctions/bifurcation vertices (projected in $\mathbb{R}^2$).
- The eigenvalues of two types eigenvectors can *be very close at* $4.0$.

# RGC#100: $d_{\mathrm{ROT}}$ with $\alpha = 0.5$

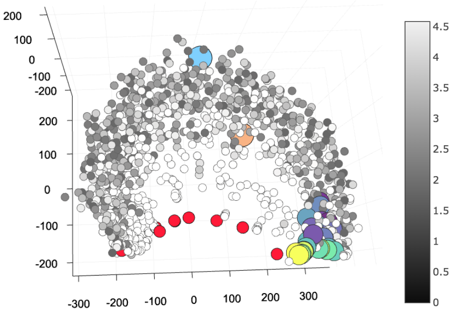
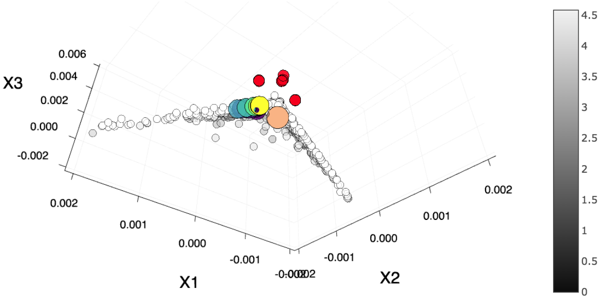



Figure: 3D-MDS embedding of the Laplacian eigenvectors of unweighted RGC #100 graph based on $d_{\mathrm{ROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha = 0.5)$: The large blue circle = the DC component and the big orange circle = the Fiedler vector; the small red circles = localized eigenvectors; the medium viridis circles = the semi-global oscillation eigenvectors. Grey scales represent the magnitude of the eigenvalues.

# RGC#100: $a_{\mathrm{HAD}}$




- $a_{\mathrm{HAD}}$ successfully separates the two types of eigenvectors, but everything is *too closely located*.
- The reason is that the Hadamard product will almost vanish on graphs, i.e., $\boldsymbol{\phi}_i \circ \boldsymbol{\phi}_j \approx \mathbf{0} \in \mathbb{R}^n$, if the active support of the concentrated part of $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ do not overlap.

# RGC#100: $d_{\text{TSD}}$ with $T = 0.1$

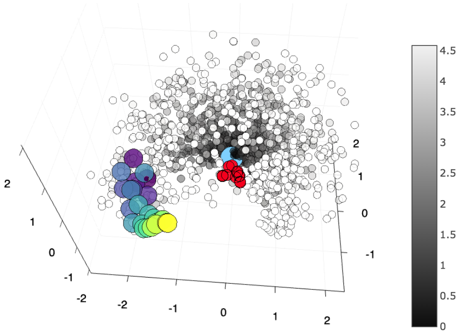
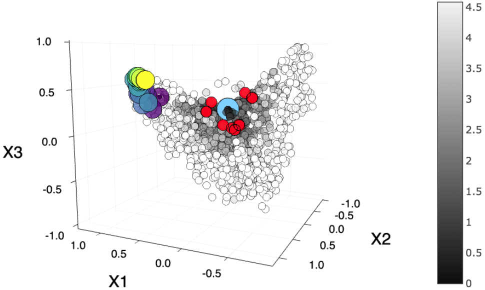



Figure: 3D-MDS embedding of the Laplacian eigenvectors of unweighted RGC #100 graph based on $d_{\text{TSD}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j; T = 0.1)$.

# RGC#100: $d_{\mathrm{DAG}}$




- The 3D-MDS result of $d_{\mathrm{TSD}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j; T = 0.1)$ and $d_{\mathrm{DAG}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)$ have similar structures.
- They also *successfully split* the two types of eigenvectors.
- But the DC vector and the Fiedler vector are *too close to distinguish* from each other in the two results.

# Outline

# Discussion

- In general, we are interested in two types of eigenvector behavior patterns on graphs: *global and directional oscillation pattern* and *energy concentration pattern*.
- Global and directional oscillation pattern represents *how the eigenvector globally oscillate on the graphs*, e.g., the DCT type II eigenvectors on 1D path graphs where the oscillation pattern is completely characterized by the eigenvalues; the eigenvectors of 2D lattice graphs or more general Cartesian product graphs where the oscillation patterns can be characterized by different directions.
- The energy concentration pattern of the eigenvector describes *which part of the graphs that the eigenvector is more active*, e.g., the tree graphs where eigenvectors may concentrated on the junctions or may have semi-global oscillation structure on certain branches.

# Discussion

- Empirically, the DAG pseudometric and the HAD affinity measure reveal the directional oscillation patterns of the eigenvectors quite well.
- The ROT metric works well on energy concentration detection.
- The TSD time-dependent metric behaves similar to DAG with small $T$ and similar to ROT with large $T$.
- However, the huge computational cost of TSD with large $T$ limit its performance on complicated graphs.
- In the future, we will work on designing better *auto-adaptive* and *cost efficient* "metrics" which expected to be *good for both types* of eigenvector behaviors on different graphs.

# Simplified ROT (sROT) Metric

- If the underlying graph $G$ is a *tree* (connected graph without loop), we can develop a computational efficient simplified ROT (sROT) metric.
- Notice that there are only two types of vertices in a tree: branch vertices (degree less than 2) and junction vertices (degree greater than 2).
- Therefore, one can simplified the tree graph *by treating the branch nodes on the same branch as one vertex* and get a simplified graph $G_s$.
- Correspondingly, instead of converting the eigenvectors $\boldsymbol{\phi}_i$ into pmfs $\boldsymbol{\phi}_i^2$ on $G$, we can convert them into *low-dimensional* pmfs $\boldsymbol{\theta}_i$ on $G_s$ by integrating the values of $\boldsymbol{\phi}_i^2$ over each branch of the tree.
- Define *sROT metric* as:

$$d_{\mathrm{sROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha) := d_{\mathrm{ROT}}(\theta_i, \theta_j; \alpha)$$

# References


- N. Saito, *"How can we naturally order and organize graph Laplacian eigenvectors?"* in Proc. 2018 IEEE Workshop on Statistical Signal Processing. (**2018**, pp. 483-487).
- A. Cloninger and S. Steinerberger, *"On the Dual Geometry of Laplacian Eigenfunctions."* Exp. Math., (**2018**), pp. 1-11.
- Q. Xia, *"Motivations, ideas and applications of ramified optimal transportation,"* ESAIM: Mathematical Modelling and Numerical Analysis, vol. 49, no. 6, pp. 1791-1832, **2015**, Special Issue Optimal Transport.
- D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, *"The emergingfield of signal processing on graphs"*, IEEE Signal Processing Magazine, 30 (**2013**), pp. 83-98.
- Y. Nakatsukasa, N. Saito, and E. Woei, *"Mysteries around the graph Laplacian eigenvalue 4,"* Linear Algebra Appl., 438, (**2013**), pp. 3231-3246.
- D. K. Hammond, P. Vandergheynst, and R. Gribonval. *"Wavelets on graphs via spectral graph theory."* Appl. Comput. Harm. Anal., 30, (**2011**), pp. 129-150.
- I. Borg and P. Groenen, *"Modern Multidimensional Scaling: Theory and Applications"*, Springer, New York, 2nd edition, **2005**.

Thanks for your attention!
Any questions?