Metrics of graph Laplacian eigenvectors

Haotian Li and Naoki Saito

Department of Mathematics, University of California, One Shields Avenue, Davis, CA 95616, USA

ABSTRACT

The application of graph Laplacian eigenvectors has been quite popular in the graph signal processing field: one can use them as ingredients to design smooth multiscale basis. Our long-term goal is to study and understand the dual geometry of graph Laplacian eigenvectors. In order to do that, it is necessary to define a certain metric to measure the behavioral differences between each pair of the eigenvectors. Saito (2018) considered the ramified optimal transportation (ROT) cost between the square of the eigenvectors as such a metric. Clonginger and Steinerberger (2018) proposed a way to measure the affinity (or 'similarity') between the eigenvectors based on their Hadamard (HAD) product. In this article, we propose a simplified ROT metric that is more computational efficient and introduce two more ways to define the distance between the eigenvectors, i.e., the time-stepping diffusion (TSD) metric and the difference of absolute gradient (DAG) pseudometric. The TSD metric measures the cost of "flattening" the initial graph signal via diffusion process up to certain time, hence it can be viewed as a time-dependent version of the ROT metric. The DAG pseudometric is the l^2-distance between the feature vectors derived from the eigenvectors, in particular, the absolute gradients of the eigenvectors. We then compare the performance of ROT, HAD and the two new "metrics" on different kinds of graphs. Finally, we investigate their relationship as well as their pros and cons.

Keywords: Graph Laplacian eigenvectors, metrics between orthonormal vectors, dual geometry of graph Laplacian eigenvectors, multiscale basis dictionaries on graphs, heat diffusion on graphs, Wasserstein distance, optimal transport

1. INTRODUCTION

The graph Laplacian eigenvectors and the ordering of the eigenvectors have been used as the two key ingredients to design graph wavelets by the Littlewood-Paley type theory for graphs. For example, the Spectral Graph Wavelet Transform (SGWT)[5] orders the eigenvectors by the size of corresponding eigenvalues. However, this ordering may lead to unexpected problems if the underlying graph is more complicated than 1D paths or cycles as pointed out by Saito and his group.[7,12] The "metrics" in this article are designed to detect the "behavioral differences" between the eigenvectors on the graph so that we can order the eigenvectors more naturally than using the size of the corresponding eigenvalues. Furthermore, these metrics help us design smooth multiscale basis dictionaries that are quite important for many applications, e.g., efficiently approximating graph signals[8] and solving differential equations on graphs.[11,18]

In this article, we study five "metrics": the Ramified Optimal Transport (ROT) metric;[12,20] the simplified ROT (sROT) metric on trees; the affinity measure proposed by Cloninger and Steinerberger;[2] the Time-Stepping Diffusion (TSD) metric; and the Difference of Absolute Gradient (DAG) pseudometric, to quantify the difference between the graph Laplacian eigenvectors and assemble the corresponding distance matrix by computing the mutual distance between the eigenvectors. The sROT and latter two "metrics" are newly proposed in this article. In order to examine the quality of these distance matrices, we use the classical multidimensional scaling (MDS) method[1] to embed the eigenvectors into low dimensional Euclidean space, i.e., $\mathbb{R}^d$ $(d = 2, 3)$. Thus, we can visualize the arrangement of the eigenvectors organized by the corresponding "metrics".

Further author information: (Send correspondence to N.S.)
H.L.: E-mail: haotianl@math.ucdavis.edu
N.S.: E-mail: saito@math.ucdavis.edu; URL: `https://www.math.ucdavis.edu/~saito`

The TSD metric is based on the time evolution of the mass propagation via a diffusion process on a graph, provided with the difference between two eigenvectors as the initial condition. The ROT metric does not explicitly sense the "scale" information of the underlying graph: it only reflects the final and global transportation cost between two eigenvectors. However, we are also interested in the intermediate situation of the transportation process, i.e., where the mass "congestion" occurs during the transportation and how quickly the mass reaches a specific region, etc. Thus, we can think of TSD as a time-evolving optimal transport-like metric.

The DAG pseudometric is designed for characterizing oscillation patterns of the graph Laplacian eigenvectors. Intuitively speaking, we take the view that the gradient of the eigenvectors contains more direct information of the oscillations than the eigenvectors themselves. In addition, since we study undirected graphs $G = (V, E)$, it is natural to consider the absolute value of the gradient on each edge $e \in E$ as features. We then compute the ℓ^2-distance between the feature vectors as the behavioral distance between corresponding eigenvectors. Under this pseudometric, we expect the eigenvectors with similar oscillation pattern are close and those with distinct oscillation behaviors are far apart. For example, the eigenvectors oscillate in one direction and those oscillate in another direction on a 2D lattice graph have orthogonal absolute gradient feature vectors, which would lead to a larger DAG distance as expected.

The structure of this article is the following. First, we review the two existing "metrics", i.e., the ROT metric[12,20] and the affinity measure of Cloninger and Steinerberger.[2] We then introduce the sROT metric on tree graphs and propose our two new "metrics", i.e., TSD and DAG. We also analyze the relationship between some of the above "metrics". Finally, we conclude with the performance of all the "metrics" on different type of graphs and provide further discussion on them.

2. NOTATION AND REVIEWS

In this section, we will introduce some basic notation about graphs that will be used through out this article. We then review the two existing behavioral "metrics" of graph Laplacian eigenvectors and also introduce a method to simplify the ROT on trees for less computational cost.

2.1 Basics of Graphs

First, we review some background knowledge about graphs as discussed in papers.[7,10,12–14] A graph $G = (V, E, W)$ consists of a set of vertices (or nodes) $V = \{v_1, v_2, \cdots, v_n\}$, a set of edges $E = \{e_1, e_2, \cdots, e_m\}$ connecting some pairs of vertices in V and a weight matrix $W \in \mathbb{R}^{n\times n}$.

If the number of vertices is finite, i.e., $|V| < \infty$, then we call G a finite graph. If any $e \in E$ does not have a direction, then the graph is undirected. If any two vertices $v_i, v_j \in V$ are connected by a sequence of head-tail edges, then the graph is connected. Furthermore, if there is no edge connecting a vertex to itself or there are no multiple edges between any pair of vertices, then we call G a simple graph. In this article, we only deal with *finite undirected connected simple graphs.*

Next, we introduce the associated matrices on graphs.

Definition 2.1 (Graph Laplacian matrix). *Let $G = (V, E, W)$, $n = |V|$. Denote its weighted adjacency matrix $W = W(G) = (w_{ij}) \in \mathbb{R}^{n\times n}$, its degree matrix $D = D(G) = \operatorname{diag}(d_1, d_2, \cdots, d_n) \in \mathbb{R}^{n\times n}$, and its unnormalized Laplacian matrix $L = L(G) \in \mathbb{R}^{n\times n}$, whose entries are defined by the following,*

$$w_{ij} := \begin{cases} W(i,j) & \text{if } e = (v_i, v_j) \in E(G); \\ 0 & \text{otherwise.} \end{cases} \qquad d_i = d(v_i) := \sum_{j=1}^{n} w_{ij} \qquad L(G) := D(G) - W(G)$$

Also, for unweighted graphs $G = (V, E)$, W is automatically defined by $W(i,j) := \begin{cases} 1 & \text{if } e = (v_i, v_j) \in E(G); \\ 0 & \text{otherwise.} \end{cases}$

Observe that L is a real symmetric positive semi-definite matrix, so the eigenvalues of L are nonnegative. Moreover, thanks to the connectivity of graphs, $\lambda_0 = 0$ is an eigenvalue of L with multiplicity 1 and its corresponding eigenvector $\boldsymbol{\phi}_0$ is a constant vector, which is usually 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.[7,19]

Also, the eigenvectors $\{\boldsymbol{\phi}_l\}_{l=0}^{n-1}$ form an orthonormal basis (ONB) of $\mathcal{L}^2(V)$. If the multiplicity of the eigenvalue is more than 1, the choice of corresponding eigenvectors is not unique. So for simplicity, we only deal with the case when L has different eigenvalues, i.e., $0 = \lambda_0 < \lambda_1 < \lambda_2 < \cdots < \lambda_{n-1}$. In the following contexts when we talk eigenvectors, we mean the eigenvectors of the unnormalized graph Laplacian L, denoted as $\{\boldsymbol{\phi}_l\}_{l=0}^{n-1}$.

Definition 2.2 (Incidence matrix). *The incidence matrix of a* directed *graph $G = (V, E)$ is a $n \times m$ matrix Q where $n = |V|$ and $m = |E|$, such that*

$$Q(i,j) := \begin{cases} -1 & e_j \in E \text{ leaves vertex } v_i \in V; \\ 1 & e_j \in E \text{ enters vertex } v_i \in V; \\ 0 & \text{otherwise.} \end{cases}$$

If $G = (V, E)$ is undirected, we randomly assign a direction for each edge.

Definition 2.3 (Graph gradient). *Given $G = (V, E)$ and $\boldsymbol{f} \in \mathcal{L}^2(V)$, the graph gradient denoted as $\nabla_G \boldsymbol{f}$(or df) $\in \mathcal{L}^2(E)$ is defined in the following way. For any edge $e = (v_i, v_j)$, $v_i, v_j \in V$, we have*

$$\nabla_G \boldsymbol{f}(e) = \boldsymbol{f}(v_j) - \boldsymbol{f}(v_i) = Q^\intercal \boldsymbol{f} \qquad |\nabla_G \boldsymbol{f}|(e) = |\boldsymbol{f}(v_j) - \boldsymbol{f}(v_i)| \qquad |\nabla_G \boldsymbol{f}| = \text{abs}.(Q^\intercal \boldsymbol{f}) \in \mathbb{R}^{|E|} \tag{1}$$

where abs. *is the operation of taking absolute value in a component-wise manner and Q is the incidence matrix of G.* The reason we are interested in $|\nabla_G \boldsymbol{f}|$ over $\nabla_G \boldsymbol{f}$ is because the absolute value will get rid of the randomness when we assign directions to each edge for undirected graphs.

2.2 Spectral Graph Theory

Let $G = (V, E, W)$ be a graph with $|V| = n$. The classical Fourier transform is the expansion of a function f in terms of the eigenfunctions of the Laplace operator: $\hat{f}(\xi) = \langle f, e^{2\pi \mathrm{i} \xi t} \rangle$. Analogously, *the graph Fourier transform*[16] of $\boldsymbol{f} \in \mathcal{L}^2(V)$ is defined by the eigenvectors of the unnormalized graph Laplacian L:

$$\hat{\boldsymbol{f}}(l) = \langle \boldsymbol{f}, \boldsymbol{\phi}_l \rangle \qquad \text{for } l = 0, 1, \cdots, n-1$$

where $\boldsymbol{\phi}_l \in \mathbb{R}^n$ is the l-th eigenvector of L.

If the underlying graph is a simple undirected path, then the eigenvectors of its Laplacian matrix are nothing but the Discrete Cosine Transform (DCT) type II basis vectors,[10] which has been widely used in classical Fourier theory and signal analysis.[17] This is one of the reasons why people often use the eigenvalues and eigenvectors of the graph (unnormalized) Laplacian L as an analysis tool on graphs, such as the spectral graph wavelet transform (SGWT)[5] and other graph wavelets discussed in the survey.[15]

When the underlying graph is more complicated (not an undirected path or cycle), one may encounter serious problems if the eigenvectors are ordered and organized based on the size of the corresponding eigenvalues. In complicated graphs, there is no well-defined notion of "frequency" unlike in the case of simple path graphs since some eigenvectors may not have a global oscillation structure. The relations between eigenvectors and eigenvalues become more subtle.[10,12,14] Thus, one solution of this problem is to come up with some "metrics" of the graph Laplacian eigenvectors so that the eigenvectors can be organized based on their *behaviors* on graphs. We note that the usual ℓ^2-distance between the graph Laplacian eigenvectors does not work since $\|\boldsymbol{\phi}_i - \boldsymbol{\phi}_j\|_2 \equiv \sqrt{2}\delta_{ij}$ where δ_{ij} is the Kronecker delta.

2.3 Ramified Optimal Transport (ROT) Metric

The ROT metric[12] is presented as follows. First, we convert each eigenvector to a probability mass function (pmf) on the input graph $G = (V, E, W)$ with $|V| = n$ and $|E| = m$ (e.g., squaring an eigenvector $\boldsymbol{\phi}_i$ elementwise turns it to a pmf $\boldsymbol{\phi}_i^2$, which can be interpreted as the energy distribution of the eigenvector), and define the metric between a pair of the eigenvectors by the minimal cost to move the probability mass from one pmf $\boldsymbol{p}$ to the other pmf $\boldsymbol{q}$.

$$\tilde{\tilde{Q}} \boldsymbol{w} = \boldsymbol{q} - \boldsymbol{p}, \qquad \boldsymbol{w} \in \mathbb{R}^{2m}_{\geq 0}, \tag{2}$$

where $\tilde{\tilde{Q}} \in \mathbb{R}^{n\times 2m}$ is the *incidence matrix* of the *bidirected graph* $\tilde{\tilde{G}}$ generated from the undirected original graph G, i.e., each edge in $E(G)$ becomes two directed edges in $E(\tilde{\tilde{G}})$, so that the probability "mass" can move in either directions. Note that any $\boldsymbol{w}$ satisfying Eq. (2) represents a transportation path (or plan) from $\boldsymbol{p}$ to $\boldsymbol{q}$, and there may be multiple solutions. Hence, we first define the cost of a transport path $P \in \text{Path}(\boldsymbol{p}, \boldsymbol{q})$ as:

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

where length(e) is the "length" of the edge $e \in E(P)$, which may be the Euclidean distance between the two nodes associated with e, or the *inverse* of the original edge weight of the input graph G if the original edge weight represents the affinity between those two nodes. Now, we can define the *minimum transportation cost*

$$d_{\text{ROT}}(\boldsymbol{p}, \boldsymbol{q}; \alpha) := \min_{P\in\text{Path}(\boldsymbol{p},\boldsymbol{q})} \boldsymbol{M}_\alpha(P).$$

Xia proved that this is a metric on the space of pmfs.[20]

2.4 Simplified ROT (sROT) Metric

If the underlying graph is a tree (connected graph without any loop), we can develop a computational efficient simplified ROT (sROT) metric. Notice that there are only three types of vertices in a tree: terminal vertices (degree 1); internal vertices (degree 2); and junction vertices (degree greater than 2). When we consider the pmf's of the eigenvectors (i.e., $\Phi.^2 = [\boldsymbol{\phi}_0^2, \boldsymbol{\phi}_1^2, \cdots, \boldsymbol{\phi}_{n-1}^2] \in \mathbb{R}^{n\times n}$) for the ordering and organization purposes, we are mainly analyzing how the probability mass distributed on different branches (consisting of terminal and internal vertices) and junctions.

Therefore, we decompose the tree into branches and junctions. In order to do that, we first find all the junction vertices of the tree by checking the degree of each vertex. Then, we use the junction vertices to chop the tree into several branches and junctions (i.e., subgraphs). In particular, the junction vertices corresponds to a bunch of one-vertex subgraphs. After this process, we get J disconnected subgraphs $G^l = (V^l, E^l)$ $(l = 1, 2, \cdots, J)$ including all the branches and junctions of the tree. We also get a graph of the subgraphs, denoted as $G_s = (V_s, E_s)$, which describes how the subgraphs related with each other. Intuitively speaking, we compress the branches of the tree graph G as one vertex in G_s.

Next, we define sROT metric between eigenvectors. Like the first step of ROT, we convert the eigenvectors into its energy form $\Phi.^2$ (elementwise square). But then instead of computing the ROT distance directly between $\boldsymbol{\phi}_i^2$ and $\boldsymbol{\phi}_j^2$, we perform a preprocessing step. We compute the mass of $\boldsymbol{\phi}_i^2$ on each subgraph $G^l = (V^l, E^l)$ $(l = 1, 2, \cdots, J)$. In other words, if G^l is a junction one-vertex (i.e., $V = \{v\}$) subgraph, we just preserve the value of $\boldsymbol{\phi}_i^2$ at vertex v; if G^l corresponds to a branch subgraph, we sum probability mass of $\boldsymbol{\phi}_i^2$ over the subgraph vertices. In the end, we get a J dimensional vector for each eigenvector $\boldsymbol{\phi}_i$ $(i = 0, 1, \cdots, n-1)$. These vectors can be viewed as energy distribution feature vectors over J different subgraphs. Denote these feature vectors by $\Theta := [\theta_0, \theta_1, \cdots, \theta_{n-1}] \in \mathbb{R}^{M\times n} (J \ll n)$ where $\theta_i(l) := \sum_{v\in V^l} \boldsymbol{\phi}_i^2(v)$ for $l = 1, 2, \cdots, J$. Notice that each θ_i is a low dimensional pmf representation of $\boldsymbol{\phi}_i$. We then compute the ROT distance between θ_i and θ_j on the graph of subgraphs, i.e., $d_{\text{ROT}}(\theta_i, \theta_j; \alpha)$ on G_s, which will reduce the computational cost a lot compared to the original ROT, i.e., $d_{\text{ROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha)$ on G. We call this distance as the *sROT metric* between eigenvectors $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$, denoted as

$$d_{\text{sROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha) := d_{\text{ROT}}(\theta_i, \theta_j; \alpha) \tag{3}$$

2.5 Hadamard (HAD) Product Affinity Measure

The affinity measure between eigenvectors is introduced by Cloninger and Steinerberger,[2] which deals with the general setting, i.e., on a compact Riemannian manifold $(\mathcal{M}, g)$ as:

$$a_{\text{HAD}}(\phi_i, \phi_j)^2 := \|\phi_i\phi_j\|_2^{-2} \int_{\mathcal{M}} \Big(\int_{\mathcal{M}} p(t,x,y)(\phi_i(y) - \phi_i(x))(\phi_j(y) - \phi_j(x))\mathrm{d}y\Big)^2 \mathrm{d}x, \tag{4}$$

where $(\lambda_i, \phi_i)_i$ is an eigenpair of the Laplace-Beltrami operator Δ on $\mathcal{M}$, $p(t, x, y)$ is the classical heat kernel,[4] 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. Further, it can be shown that for the same t above

$$a_{\mathrm{HAD}}(\phi_i, \phi_j) = \frac{\|e^{t\Delta}(\phi_i\phi_j)\|_{L^2}}{\|\phi_i\phi_j\|_{L^2}} \tag{5}$$

This works well for Cartesian product graphs[2] in terms of detecting the Cartesian product structure of such graphs and the oscillation patterns of the eigenvectors.

3. OUR PROPOSED METRICS

3.1 Time-Stepping Diffusion (TSD) Metric

3.1.1 TSD metric on graphs

Given a graph $G = (V, E, W)$, consider the governing heat diffusion ODE system on the graph, which describes the evolution of the graph signal $\boldsymbol{f}_0 \in \mathbb{R}^n$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{f}(t) + L(G) \cdot \boldsymbol{f}(t) = \mathbf{0} \quad t \geq 0, \qquad \boldsymbol{f}(0) = \boldsymbol{f}_0 \in \mathbb{R}^n. \tag{6}$$

Since the graph Laplacian (i.e., $L(G)$) eigenvectors $\{\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{n-1}\}$ form an ONB of $\mathbb{R}^n$, we have:

$$\boldsymbol{f}(t) = \sum_{k=0}^{n-1} \langle \boldsymbol{f}_0, \boldsymbol{\phi}_k \rangle e^{-\lambda_k t} \boldsymbol{\phi}_k \tag{7}$$

At a certain time $T > 0$, we define the following functional:

$$K(\boldsymbol{f}_0, T) := \int_0^T \|\nabla_G \boldsymbol{f}(t)\|_1 \mathrm{d}t, \tag{8}$$

where $\nabla_G \boldsymbol{f} \in \mathbb{R}^m$ is the *graph gradient* of $\boldsymbol{f}$ defined in Definition 2.3. This functional can be viewed as the cost (or effort) to "flatten" the initial graph signal $\boldsymbol{f}_0$ via heat diffusion process up to the time T, and also as the time-accumulated "anisotropic total variation norm" [6] of $\boldsymbol{f}_0$. Also, 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 the following lemma.

LEMMA 3.1. *For any $T > 0$ (including $T = \infty$), $K(\cdot, T)$ is a norm on $\mathcal{L}_0^2(V) := \{\boldsymbol{f} \in \mathcal{L}^2(V) | \sum_{x\in V} \boldsymbol{f}(x) = 0\}$. Therefore, $(\mathcal{L}_0^2(V), K(\cdot, T))$ is a normed vector space. Furthermore, for any fixed T (including $T = \infty$), we can get a metric vector space $(\mathcal{L}_0^2(V), d_{\mathrm{TSD}})$ by defining*

$$d_{\mathrm{TSD}}(\boldsymbol{f}, \boldsymbol{g}) := K(\boldsymbol{f} - \boldsymbol{g}, T) \qquad \boldsymbol{f}, \boldsymbol{g} \in \mathcal{L}_0^2(V)$$

See Appendix A for the proof.

3.1.2 TSD metric on a compact Riemannian manifold $\mathcal{M}$

Since the heat diffusion system, Eq. (6), can be defined on a compact Riemannian manifold $\mathcal{M}$ by the Laplace-Beltrami operator, we can generalize TSD metric to a continuous setting.

We consider the heat diffusion on $\mathcal{M}$ with Neumann boundary conditions and the corresponding eigenfunctions $-\Delta\phi = \lambda\phi$ with $\|\phi\|_2 = 1$. So given a initial signal $f_0 \in \mathcal{L}_0^2(\mathcal{M}) := \{f \in \mathcal{L}^2(\mathcal{M}) : \int_{\mathcal{M}} f(x)\mathrm{d}\mu(x) = 0\}$, where $\mathrm{d}\mu$ is the measure on $\mathcal{M}$, the TSD functional can be defined by:

$$K(f_0, T) := \int_0^T \int_{\mathcal{M}} |\nabla_x u(x, t)| \mathrm{d}\mu(x)\mathrm{d}t$$

There is a natural upper bound of this functional as shown in the following theorem.

THEOREM 3.2. *For any $T > 0$ and $f_0 \in \mathcal{L}^2_0(\mathcal{M})$,*

$$K(f_0, T) \leq \sum_{k=1}^{\infty} \frac{1}{\sqrt{\lambda_k}} |\hat{f}_0(k)| \cdot \sqrt{\mathrm{Vol}(\mathcal{M})}, \tag{9}$$

where λ_k's are the positive eigenvalues of Laplace-Beltrami operator and $\hat{f}_0(k) = \langle \phi_k, f_0 \rangle$ are the Fourier coefficients of f_0.

See Appendix B.1 for the proof. Therefore, the TSD metric with parameter T (including $T = \infty$) on $\mathcal{M}$ between eigenfunctions (except the DC component) ϕ_i and ϕ_j is well defined by

$$d_{\mathrm{TSD}}(\phi_i, \phi_j; T) := K(\phi_i - \phi_j, T) \leq (\frac{1}{\sqrt{\lambda_i}} + \frac{1}{\sqrt{\lambda_j}}) \sqrt{\mathrm{Vol}(\mathcal{M})} < \infty$$

3.2 Difference of Absolute Gradient (DAG) Pseudometric

We use the absolute gradient vector of an eigenvector as its feature vector describing its behavior. More precisely, see Eq. (1) in Definition 2.3. We then define:

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

This quantity is a *pseudometric* since the identity of discernible in the axioms of metric is not satisfied (e.g., adding constants to $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ clearly does not change the absolute gradient values) but the other axioms, i.e., the non-negativity, symmetry and triangle inequality, are satisfied. One of the biggest advantages of the DAG metric is its lower computational cost than the others, because it only involves multiplications of the eigenvectors by the sparse matrix $Q^{\intercal}$.

4. RELATIONS BETWEEN METRICS

4.1 The ROT and the TSD Metrics

The purpose of the TSD metric is to construct time-evolving optimal transport-like metric. As $T \to \infty$, we expect $d_{\mathrm{ROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha = 1) \leq d_{\mathrm{TSD}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; T = \infty) \leq C(G) \cdot d_{\mathrm{ROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha = 1)$, where $C(G)$ is a constant depending on the graph G. Moreover, we observe that if f, g are pmfs on graphs, then $d_{\mathrm{ROT}}(f, g; \alpha = 1) = W_1(f, g)$, the 1^{st} Wasserstein distance on graphs. Thus, we can also generalize the d_{ROT} metric on graphs to a compact Riemannian manifold M by the generalized W_1 on manifolds.

CONJECTURE 4.1. *Given any two probability mass functions (pmfs) $\boldsymbol{p}, \boldsymbol{q}$ on a connected graph $G = (V, E, W)$ with graph geodesic distance metric $d : V \times V \to \mathbb{R}_{\geq 0}$, i.e., the minimum sum of edge weights over all the paths connecting two input vertices,*

$$W_1(\boldsymbol{p}, \boldsymbol{q}) \leq K(\boldsymbol{p} - \boldsymbol{q}, \infty) \leq C' \cdot W_1(\boldsymbol{p}, \boldsymbol{q})$$

where $W_1(\boldsymbol{p}, \boldsymbol{q}) := \inf_{\gamma \in \Gamma(\boldsymbol{p}, \boldsymbol{q})} \int_{V \times V} d(x, y) \mathrm{d}\gamma(x, y)$, where $\Gamma(\boldsymbol{p}, \boldsymbol{q})$ denotes the collection of all measures on $V \times V$ with marginals $\boldsymbol{p}$ and $\boldsymbol{q}$ in the first and second factors respectively and C' is a constant depends on G and $K(\cdot, \cdot)$ is defined in Eq. (8).

There is also the manifold version of this conjecture. If the underlying manifold is $\mathcal{M} = [0, 1]$ or $\mathcal{M} = \mathbb{T}$, where the explicit expression of W_1 is known,[9] then we can show the first inequality of the conjecture as follows.

THEOREM 4.2. *Given two probability density functions f, g on $[0, 2\pi]$,*

$$W_1(f, g) \leq K(f - g, \infty)$$

in which W_1 is the 1^{st} Wasserstein distance, a.k.a., the earth mover distance.

See Appendix B.2 for the proof. Since there is no explicit formula of W_1 on other complicated manifolds or discrete graphs, the proof even for the first inequality of the conjecture is hard to proceed.

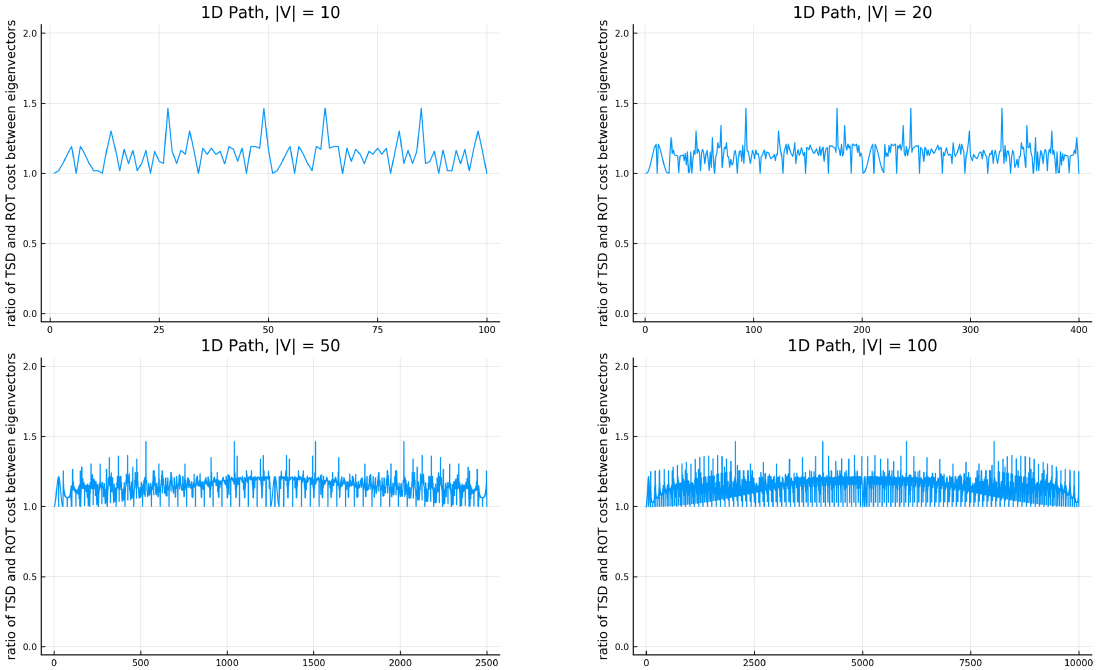


Figure 1: The 1D representations of the ratio matrix $R \in \mathbb{R}^{n\times n}$ between square of graph Laplacian eigenvectors, i.e., $R(i,j) = \rho(\boldsymbol{\phi}_{i-1}^2, \boldsymbol{\phi}_{j-1}^2)$, on different 1D path graphs. Every entry of R is lower bounded by 1 and the best upper bound constant $C(G) \approx 1.47$ for all the four cases.

On the other hand, empirically, we can always perform some numerical experiments on different graphs to test the conjecture. To make things easier to deal with, we introduce the ratio between $K(\boldsymbol{p}-\boldsymbol{q}, \infty)$ and $W_1(\boldsymbol{p}, \boldsymbol{q})$,

$$\rho(\boldsymbol{p}, \boldsymbol{q}) := \begin{cases} \frac{K(\boldsymbol{p}-\boldsymbol{q},\infty)}{W_1(\boldsymbol{p},\boldsymbol{q})}, & \text{if } \boldsymbol{p} \neq \boldsymbol{q}; \\ 1, & \text{if } \boldsymbol{p} = \boldsymbol{q}. \end{cases}$$

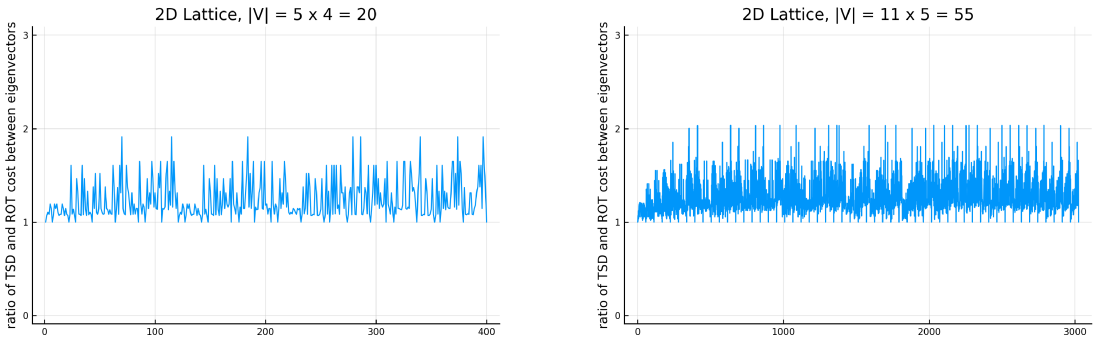
Now, the inequalities of Conjecture 4.1 are reformed to the following.

$$1 \leq \rho(\boldsymbol{p}, \boldsymbol{q}) \leq C'$$

Numerical Experiments of Conjecture 4.1 with Eigenvectors: Given a graph $G = (V, E)$, we can view the square of the eigenvector $\boldsymbol{\phi}_i^2$ as a pmf on the graph. We compute the ratio matrix $R \in \mathbb{R}^{n\times n}$ whose entry is defined by $R(i,j) := \rho(\boldsymbol{\phi}_{i-1}^2, \boldsymbol{\phi}_{j-1}^2)$. We hope every entry in R is lower bounded by 1, i.e., $\min_{i,j} R(i,j) \geq 1$, and there is some constant $C(G) > 1$ such that $\max_{i,j} R(i,j) \leq C(G)$. In other words, we want to verify the following inequalities in different graphs.

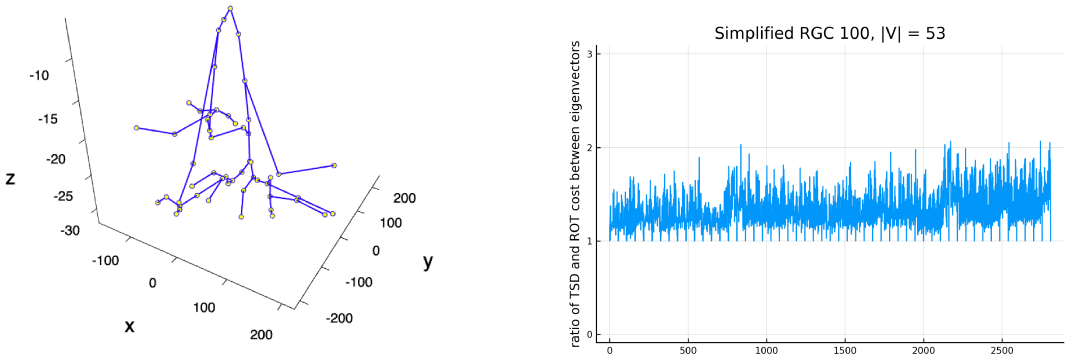
$$1 \leq R(i,j) \leq C(G) \qquad \forall i, j = 1, 2, \cdots, n \tag{10}$$

In order to visualize the results and find proper $C(G)$, we reshape the matrix R into a 1D vector and make the 1D plots for each G. The results of 1D path, 2D lattice and the simplified RGC#100 tree, are as shown in Figures 1, 2, & 3, respectively. As we can see from these figures, Eq. (10) is satisfied for different graphs with different $C(G)$. Thus, the conjecture is empirically verified in these graphs when the input pmfs are $\boldsymbol{\phi}_i^2$ $(i = 0, 1, \cdots, n-1)$.



(a) Best upper bound $C(G) \approx 1.92$ (b) Best upper bound $C(G) \approx 2.04$

Figure 2: The 1D representations of the ratio matrix $R \in \mathbb{R}^{n \times n}$ as in Eq. (10) on different 2D lattice graphs.



(a) The simplified RGC#100 graph (b) Best upper bound $C(G) \approx 2.07$

Figure 3: The 1D representation of the ratio matrix $R \in \mathbb{R}^{n \times n}$ as in Eq. (10) on the simplified RGC#100 graph.

4.2 The DAG and the HAD Affinity Measure

The DAG is closely related to the HAD affinity measure between eigenvectors introduced by Cloninger and Stenerberger.[2] Based on the definition of the DAG pseudometric, we derive the following equations:

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

The last equality follows from the discrete version of Green's first identity on graphs[3] and exchanging the sum over vertices with the sum over edges. From the above derivations, we can see that the last term of the formula can be viewed as a global average of absolute local correlation between these two eigenvectors, which is close to the interpretation of Eq. (4).

5. NUMERICAL EXPERIMENTS

In this section, to evaluate the performance of those five "metrics" discussed in Sections 2 & 3 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_{ROT} and d_{sROT}) using each "metric". We then use the classical MDS method[1] 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".

We first compare embeddings using the ROT and the sROT on a simple tree. We then present the eigenvector arrangements by different "metrics" in the embedded MDS space (i.e., $\mathbb{R}^2$ or $\mathbb{R}^3$) on two different graphs, i.e., a 2D lattice graph and a dendritic tree.

5.1 The ROT and the sROT on a Simple Tree

We generated a simple tree $G = (V, E)$ for demonstration purposes (See Fig. 4). In this graph, $|V| = 100$ and $|E| = 99$. Moreover, there are four junction vertices and four branches that we are interested, i.e., top left branch V^1, bottom left branch V^2, bottom right branch V^3 and top right branch V^4 (see Section 2.4 for more details). By computing the energy level of the eigenvectors on different branches, i.e.,

$$\mathrm{el}_k(i) := \frac{\sum_{v\in V^k} \boldsymbol{\phi}_i^2(v)}{\sum_{v\in V} \boldsymbol{\phi}_i^2(v)} = \sum_{v\in V^k} \boldsymbol{\phi}_i^2(v) \qquad k = 1, 2, 3, 4,$$

and by thresholding $\mathrm{el}_k(i) \geq 0.5$, we select and group the eigenvectors that concentrate on the k-th branch.

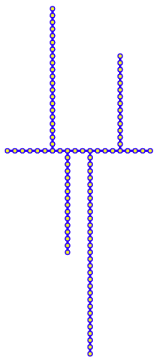


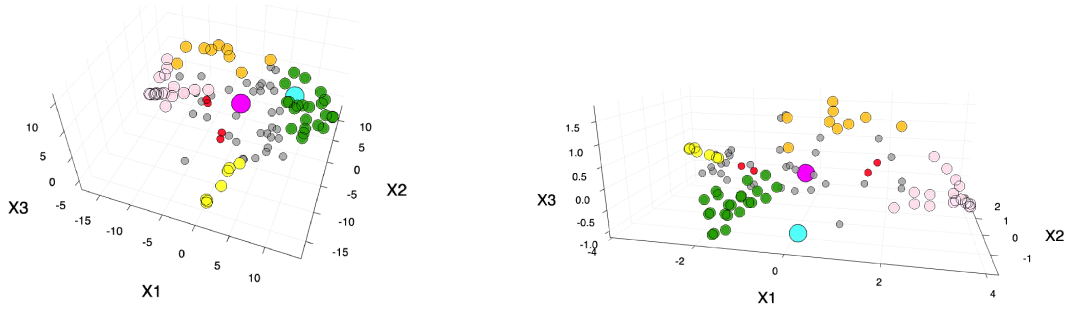
Figure 4: A simple tree with four branches

We use four colors, i.e., pink, orange, green and yellow, to represent the different group of eigenvectors that concentrated on different branches, i.e., V^1, V^2, V^3 and V^4, respectively. Similarly, we select and group the eigenvectors that focus on the junction vertices and we use red color for them. In Fig. 6, we show the five representatives of the eigenvectors in different groups. We consider eigenvectors within each group (i.e., eigenvectors with same color) have similar behavior on the tree.

The 3D-MDS results of $d_{\mathrm{ROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha = 1)$ and $d_{\mathrm{sROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha = 1)$ are as shown in Fig. 5. In Fig. 5, the magenta circle represents the DC vector $\boldsymbol{\phi}_0$; the cyan circle represents the Fiedler vector $\boldsymbol{\phi}_1$; the red circles represent eigenvectors concentrated on junctions; the pink circles represent eigenvectors concentrated on V^1; the orange circles represent eigenvectors concentrated on V^2; the green circles represent eigenvectors concentrated on V^3; the yellow circles represent eigenvectors concentrated on V^4. We can see from the figure that the similar behavior eigenvectors are clustered in the embedded space. On the contrary, if we use the eigenvalue to organize the eigenvectors, we will put $\boldsymbol{\phi}_{95}$ and $\boldsymbol{\phi}_{96}$ inevitably close even though their behaviors on the tree is very different, i.e., $\boldsymbol{\phi}_{95}$ has semi-global oscillation structure while $\boldsymbol{\phi}_{96}$ is extremely localized on the junctions.

Furthermore, $\boldsymbol{\phi}_{96}$ concentrates on the left most junction which is more closely related with semi-global oscillation eigenvectors like $\boldsymbol{\phi}_{94}$ (pink) and $\boldsymbol{\phi}_{92}$ (orange) compared to $\boldsymbol{\phi}_{95}$ (green) and $\boldsymbol{\phi}_{91}$ (yellow). In Fig. 5, if we discriminated the localized red eigenvectors at different junctions by different colors, we can get such structure for both the ROT and the sROT metrics.

The clustering patterns of embedded eigenvectors by these two metrics are similar, but the d_{sROT} is more computational efficient than d_{ROT}. In this graph, the energy feature vector of each eigenvector via d_{sROT} is $J = 13$ (4 branches, 4 junctions and 5 broken segments) comparing to $n = 100$ via d_{ROT}.



(a) $d_{\mathrm{ROT}}(\phi_i^2, \phi_j^2; \alpha = 1)$ (b) $d_{\mathrm{sROT}}(\phi_i^2, \phi_j^2; \alpha = 1)$

Figure 5: 3D-MDS results of the ROT and the sROT on the simple tree (Fig. 4).

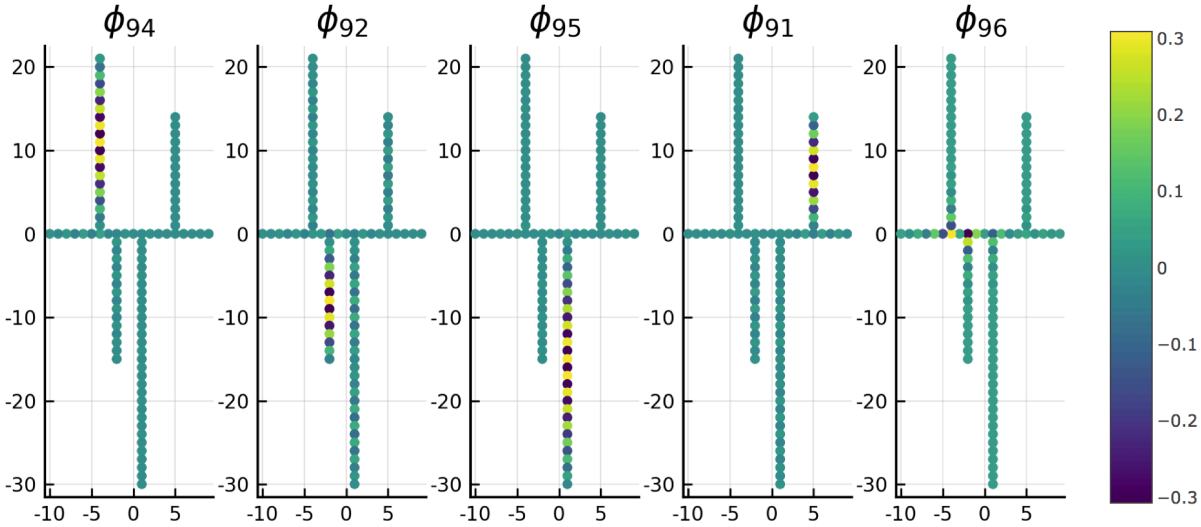


Figure 6: Representatives of five eigenvector groups: from left to right, pink-group ϕ_{94}; orange-group ϕ_{92}; green-group ϕ_{95}; yellow-group ϕ_{91}; red-group ϕ_{96} (concentrated on the junctions).

5.2 2D Lattice $P_{11} \times P_5$

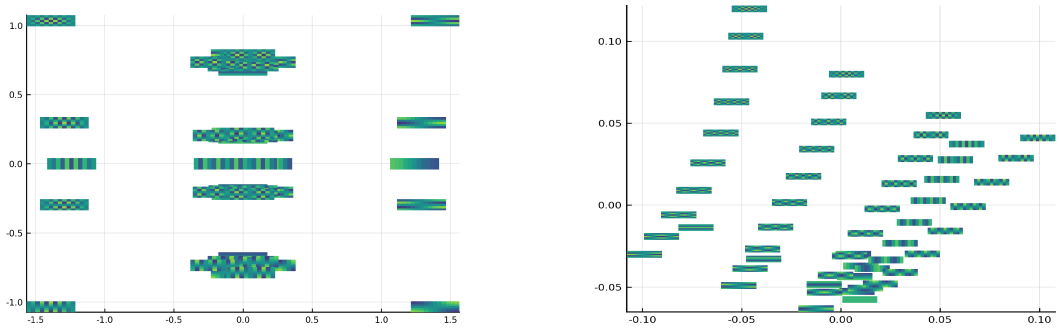
The 2D-MDS result of $d_{\mathrm{ROT}}(\phi_i^2, \phi_j^2; \alpha = 1)$ (Fig. 7a) reveals the two-dimensional ordering of the eigenvectors, but the eigenvectors with even or odd oscillations in either x (horizontal axis) or y (vertical axis) direction are embedded into a symmetric pattern around the DC vector ϕ_0. The reason is demonstrated in the paper.[12] It is mainly because we lose "half" of the information when we measure the difference between *the squared eigenvectors*.

The 2D-MDS result of a_{HAD} (Fig. 7b) recovers the two dimensional oscillation structure of the eigenvectors. It perfectly reflects the oscillation in x direction, but has a little misordering in y direction.

The 2D-MDS result of $d_{\mathrm{TSD}}(\phi_i, \phi_j; T)$ with different stopping times T are presented in Fig. 8. When $T = 0.1$ or $T = 1$, the structure of the oscillation is recovered as five curves with the same y direction oscillation pattern in each curve. As time increases, the Fiedler vector and DC vector becomes further away and other eigenvectors become more congested and the oscillation structures are not so obvious. Therefore, in this case, the result of $d_{\mathrm{TSD}}(\phi_i, \phi_j; T = \infty)$ is worse than the result of $d_{\mathrm{ROT}}(\phi_i^2, \phi_j^2; \alpha = 1)$ and they also look quite different. However, this does not conflict with the Conjecture 4.1, because we compared $d_{\mathrm{TSD}}(\phi_i^2, \phi_j^2; T = \infty)$ with $d_{\mathrm{ROT}}(\phi_i^2, \phi_j^2; \alpha = 1)$ in the conjecture instead of $d_{\mathrm{TSD}}(\phi_i, \phi_j; T = \infty)$ with $d_{\mathrm{ROT}}(\phi_i^2, \phi_j^2; \alpha = 1)$.

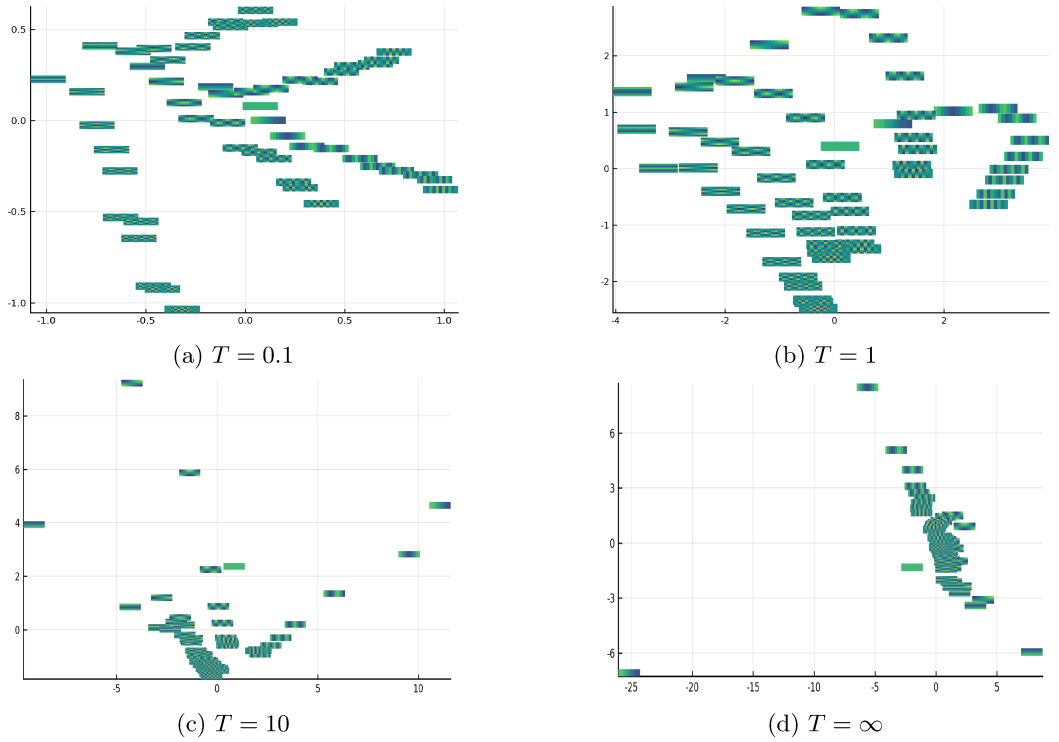
The 2D-MDS result of d_{DAG} is demonstrated in Fig. 9. We observe that d_{DAG} nicely detect two directions of the oscillations. The eigenvectors are organized in 2D array. For each column of the array, the eigenvectors have the same oscillation pattern in y direction and oscillation in x direction increases linearly. On the other hand,

for each row of the array, the eigenvectors have the same oscillation pattern in x direction and oscillation in y direction changes linearly.



(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 7: 2D-MDS embedding of the eigenvectors of 11×5 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.



(a) $T = 0.1$ (b) $T = 1$

(c) $T = 10$ (d) $T = \infty$

Figure 8: 2D-MDS embedding of the eigenvectors of 11×5 unweighted lattice graph based on $d_{\mathrm{TSD}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j; T)$ with different T: each small heatmap plot describes how the eigenvector looks like on the lattice graph.

5.3 Dendritic Tree of an RGC of a Mouse

Fig. 10a presents the conversion of the 3D dentritic tree to RGC #100 graph in $\mathbb{R}^3$ (see the reference[13] for the details of this RGC tree).

The 3D points cloud in Fig. 11 shows the 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 represents the DC vector $\boldsymbol{\phi}_0$, while the big orange circle shows the Fiedler vector $\boldsymbol{\phi}_1$. The small red circles indicates the localized eigenvectors as shown in Fig. 10c. The medium size viridis circles stand for the eigenvectors that concentrated on one of the

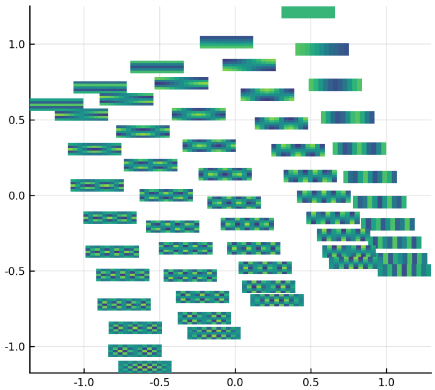


Figure 9: 2D-MDS embedding of the eigenvectors of 11×5 unweighted lattice graph based on $d_{\mathrm{DAG}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)$: each small heatmap plot describes how the eigenvector looks like on the lattice graph.

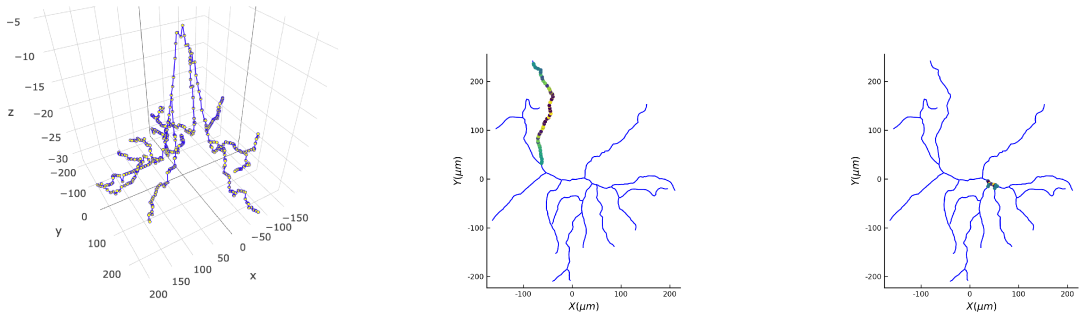
upper left branches as shown in Fig. 10b. Grey scales represent the size of corresponding eigenvalues. As we can see from the figure, the behavioral difference between two types of eigenvectors discussed in Fig. 10 are detected by this metric. If we order the eigenvectors by the size of their corresponding eigenvalues, we cannot distinguish such difference since the eigenvalues of two types eigenvectors can be very close around $\lambda = 4.0$.[14] Another observation is the DC vector and the Fiedler vector are far apart in the embedded space. In some cases, one might be interested in the behavior of the Fiedler vector but not the DC vector, d_{ROT} can emphasize the behavioral distance between them.

The 3D-MDS result of $a_{\mathrm{HAD}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)$ (see Fig. 12) looks good because it successfully separates the localized ones (red) and semi-global oscillation ones (viridis), but everything else is *too closely located* and the red ones do not always stay close with each other, i.e., there are three others outside the range in the picture. The big disadvantage of a_{HAD} is that it cannot handle the behavior of remotely-located localized eigenvectors very well. The reason is that the Hadamard product in Eq. (5) 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. This is also the reason why everything looks too congested in Fig. 12. There are many remotely-located localized eigenvectors in the RGC #100 graph which lead to extreme small HAD affinity measures.

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 (see Fig. 13 and Fig. 14). First, one of the reasons to choose $T = 0.1$ for d_{TSD} is that the coefficients of eigenvectors with large eigenvalues are diffused very fast in Eq. (7), i.e., $\exp(-\lambda_i t)$ decay fast for large λ_i. As we already demonstrated the long-term behavior of d_{TSD} in the Conjecture 4.1, we are also interested in the short-term behavior of d_{TSD}, i.e., every coefficient of the eigenvector has not diffused too much in Eq. (7). Since the maximum eigenvalue of the graph is $\lambda_{1153} \approx 4.58$, so the value of $\exp(-\lambda_{1153} T)$ will be not too small for $T = 0.1$. Another reason of using such small T is the large computational cost of d_{TSD} for large T. In these figures, they also successfully split the two types of eigenvectors, i.e., localized ones and those with semi-global oscillations on the upper left branch. Moreover, everything does not look too crowded. The relation between the DC vector and the Fiedler vector, however, is one of the big differences between these results and the d_{ROT} result. In these two figures, the DC vector and the Fiedler vector are too close to distinguish from each other. In other words, d_{TSD} with small T and d_{DAG} do not perform so well in terms of detecting behavioral difference between $\boldsymbol{\phi}_0$ and $\boldsymbol{\phi}_1$.

6. 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. On the other hand, 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



(a) (b) (c)

Figure 10: (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[12] (projected in $\mathbb{R}^2$).

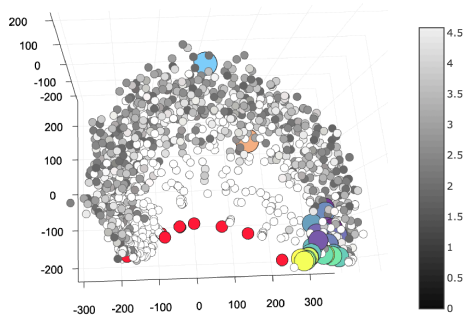


Figure 11: 3D-MDS embedding of the Laplacian eigenvectors of unweighted RGC #100 graph based on $d_{\mathrm{ROT}}(\boldsymbol{\phi}_i^2, \boldsymbol{\phi}_j^2; \alpha = 0.5)$: The large blue circle = the DC component and the big orange circle = the Fiedler vector; the small red circles = localized eigenvectors in Fig. 10c; the medium viridis circles = the semi-global oscillation eigenvectors in Fig. 10b. Grey scales represent the magnitude of the corresponding eigenvalues.

have semi-global oscillation structure on certain branches. Some "metrics" we described in this article work well to discriminate global oscillation characteristics while the others work better discriminating energy concentration pattern of eigenvectors. It is a hard but important question to ask which "metrics" is preferable for eigenvectors organization on a given graph. In the following, we discuss some empirical observations on different type of graphs.

For Cartesian graphs, the ROT metric does not perform well for detecting the oscillation patterns of the graph Laplacian eigenvectors in general. On the contrary, the DAG pseudometric and the HAD affinity measure reveal the directional oscillation patterns of the eigenvectors quite well. However, for tree graphs, the sROT and ROT metrics are better than the other two above in detecting the energy localization of the eigenvectors. For the TSD metric with small T, perceptually, it should be good for oscillation detection because it approximately computes the total variation of the difference between the eigenvectors (see Section 3.1.1), which behaves similarly as the DAG pseudometric. On the other hand, for the TSD with large T, it should be good for energy concentration detection because it behaves more like the ROT metric does. 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.

We also observe that even though with the assumption of different eigenvalues, the choice of eigenvectors can still vary by signs, e.g., $\boldsymbol{\phi}_i$ can be substituted by $-\boldsymbol{\phi}_i$. Hence, it is important for the behavioral "metrics", i.e., d, of the eigenvectors to satisfy $d(\boldsymbol{\phi}_i, -\boldsymbol{\phi}_i) = 0$ for $i = 0, 1, \cdots, n-1$. Most of the metrics we mentioned above do not have this property, e.g., d_{TSD} and d_{DAG}. We will consider this property to design our future "metrics" between eigenvectors.

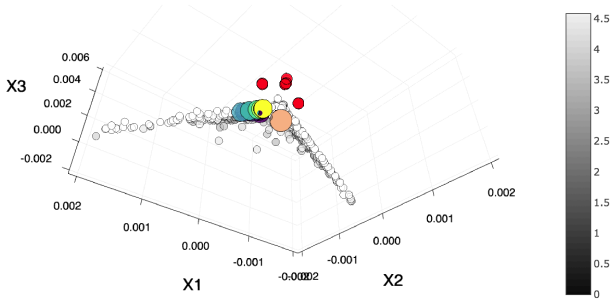


Figure 12: 3D-MDS embedding of the Laplacian eigenvectors of unweighted RGC #100 graph based on $a_{\mathrm{HAD}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)$: The large blue circle = the DC component and the big orange circle = the Fiedler vector; the small red circles = localized eigenvectors in Fig. 10c; the medium viridis circles = the semi-global oscillation eigenvectors in Fig. 10b. Grey scales represent the magnitude of the corresponding eigenvalues.

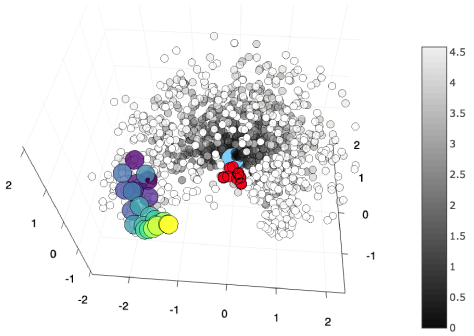


Figure 13: 3D-MDS embedding of the Laplacian eigenvectors of unweighted RGC #100 graph based on $d_{\mathrm{TSD}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j; T = 0.1)$: The large blue circle = the DC component and the big orange circle = the Fiedler vector; the small red circles = localized eigenvectors in Fig. 10c; the medium viridis circles = the semi-global oscillation eigenvectors in Fig. 10b. Grey scales represent the magnitude of the corresponding eigenvalues.

7. SUMMARY

In this article, we have proposed the simplified ROT metric i.e., d_{sROT}, and two new behavioral "metrics" of graph Laplacian eigenvectors, i.e., d_{TSD} and d_{DAG}. By comparing them to the two existing "metrics", i.e., d_{ROT} and a_{HAD}, we have demonstrated that the d_{sROT} is more computational efficient with good results on trees while the d_{TSD} behaves like a time-dependent $d_{\mathrm{ROT}}(\cdot\,, \cdot\,; \alpha = 1)$ metric. The d_{DAG} essentially considers global averages of absolute local correlation between eigenvectors, which is closely related to the interpretation of a_{HAD} as discussed in Section 4.2. We have compared the performance of all the four "metrics" on 2D lattice graph and the RGC #100 mouse neuronal dendritic tree. In the lattice graph, we have observed that d_{DAG} and a_{HAD} perform better than the others, i.e., they detect global and directional oscillation patterns of the eigenvectors more clearly. In the RGC #100 tree, we have demonstrated the result of a_{HAD} is not good because of the nature of the Hadamard product. On the other hand, the d_{ROT} metric has an interesting physical interpretation. Viewing two eigenvectors as two neuronal signals, the d_{ROT} can quantify the cost of transporting one of the signals along the branches of a dendritic tree and morphing it to the other signal.

APPENDIX A. PROOF OF LEMMA

Proof. [Lemma 3.1]

- (identity of discernible) If $\boldsymbol{f} = \mathbf{0} \in \mathcal{L}_0^2(V)$, then $K(\mathbf{0}, T) = 0$.

  On the other hand, if $K(\boldsymbol{f}, T) = 0$ $(T > 0)$, we denote Q as the incidence matrix as in Definition 2.2,

$$K(\boldsymbol{f}, T) = \int_0^T \|Q^{\intercal} \sum_{k=0}^{n-1} \langle \boldsymbol{f}, \boldsymbol{\phi}_k \rangle e^{-\lambda_k t} \boldsymbol{\phi}_k\|_1 \mathrm{d}t = 0$$

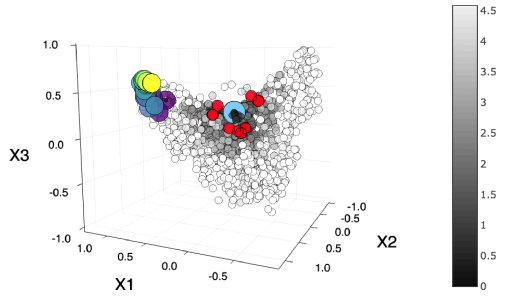


Figure 14: 3D-MDS embedding of the Laplacian eigenvectors of unweighted RGC #100 graph based on $d_{\mathrm{DAG}}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)$: The large blue circle = the DC component and the big orange circle = the Fiedler vector; the small red circles = localized eigenvectors in Fig. 10c; the medium viridis circles = the semi-global oscillation eigenvectors in Fig. 10b. Grey scales represent the magnitude of the corresponding eigenvalues.

$$
\begin{aligned}
\implies \quad & \|Q^\top \sum_{k=0}^{n-1} \langle \boldsymbol{f}, \boldsymbol{\phi}_k \rangle e^{-\lambda_k t} \boldsymbol{\phi}_k\|_1 = 0 && \text{for any } t \in [0, T] \\
\implies \quad & Q^\top \sum_{k=0}^{n-1} \langle \boldsymbol{f}, \boldsymbol{\phi}_k \rangle e^{-\lambda_k t} \boldsymbol{\phi}_k = \mathbf{0} && \text{for any } t \in [0, T] \\
\implies \quad & Q^\top \sum_{k=1}^{n-1} \langle \boldsymbol{f}, \boldsymbol{\phi}_k \rangle e^{-\lambda_k t} \boldsymbol{\phi}_k = \mathbf{0} && (Q^\top \boldsymbol{\phi}_0 = \mathbf{0} \in \mathcal{L}^2(E)) \\
\implies \quad & Q^\top \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_{n-1} \end{bmatrix} \cdot \begin{bmatrix} \langle \boldsymbol{f}, \boldsymbol{\phi}_1 \rangle e^{-\lambda_1 t} \\ \langle \boldsymbol{f}, \boldsymbol{\phi}_2 \rangle e^{-\lambda_2 t} \\ \vdots \\ \langle \boldsymbol{f}, \boldsymbol{\phi}_{n-1} \rangle e^{-\lambda_{n-1} t} \end{bmatrix} = \mathbf{0}
\end{aligned}
$$

Since the graph G is connected (implies $m \geq n-1$), it is easy to show that $\mathrm{rank}(Q) = n-1$. Denote $A := Q^\top[\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_{n-1}] \in \mathbb{R}^{m \times (n-1)}$, then $\mathrm{rank}(A) = n-1$. In other words, A is full column rank. Therefore, it implies

$$
\langle \boldsymbol{f}, \boldsymbol{\phi}_k \rangle e^{-\lambda_k t} = 0 \implies \langle \boldsymbol{f}, \boldsymbol{\phi}_k \rangle = 0 \qquad \text{for } k = 1, 2, \cdots, n-1.
$$

$$
\boldsymbol{f} \in \mathcal{L}_0^2(V) \implies \langle \boldsymbol{f}, \boldsymbol{\phi}_0 \rangle = 0
$$

Hence, $\boldsymbol{f} = \sum_{k=0}^{n-1} \langle \boldsymbol{f}, \boldsymbol{\phi}_k \rangle \boldsymbol{\phi}_k = \mathbf{0} \in \mathcal{L}_0^2(V)$.

- (absolutely homogeneous) For $\alpha \in \mathbb{R}$ and $\boldsymbol{f} \in \mathcal{L}_0^2(V)$, $K(\alpha \boldsymbol{f}, T) = |\alpha| K(\boldsymbol{f}, T)$.
- (Triangle inequality) For any $\boldsymbol{f}, \boldsymbol{g} \in \mathcal{L}_0^2(V)$ and $T > 0$,

$$
\begin{aligned}
K(\boldsymbol{f} + \boldsymbol{g}, T) &= \int_0^T \|Q^\top \sum_{k=0}^{n-1} \langle \boldsymbol{f} + \boldsymbol{g}, \boldsymbol{\phi}_k \rangle e^{-\lambda_k t} \boldsymbol{\phi}_k\|_1 \mathrm{d}t \\
&\leq \int_0^T \|Q^\top \sum_{k=0}^{n-1} \langle \boldsymbol{f}, \boldsymbol{\phi}_k \rangle e^{-\lambda_k t} \boldsymbol{\phi}_k\|_1 + \|Q^\top \sum_{k=0}^{n-1} \langle \boldsymbol{g}, \boldsymbol{\phi}_k \rangle e^{-\lambda_k t} \boldsymbol{\phi}_k\|_1 \mathrm{d}t \\
&= K(\boldsymbol{f}, T) + K(\boldsymbol{g}, T)
\end{aligned}
$$

Therefore, $K(\cdot, T)$ is a norm on $\mathcal{L}_0^2(V)$ and $(\mathcal{L}_0^2(V), K(\cdot, T))$ is a normed vector space.

□

APPENDIX B. PROOF OF THEOREMS

B.1 Proof of Theorem 3.2

Proof.

$$
\begin{aligned}
K(f_0, T) &= \int_0^T \int_{\mathcal{M}} |\nabla_x u(x,t)| \, \mathrm{d}x\mathrm{d}t \\
&= \int_0^T \int_{\mathcal{M}} \left| \nabla_x \sum_{k=0}^{\infty} \langle \phi_k, f_0 \rangle e^{-\lambda_k t} \phi_k(x) \right| \mathrm{d}x\mathrm{d}t \\
&= \int_0^T \int_{\mathcal{M}} \left| \nabla_x \sum_{k=1}^{\infty} \langle \phi_k, f_0 \rangle e^{-\lambda_k t} \phi_k(x) \right| \mathrm{d}x\mathrm{d}t \\
&\leq \int_0^T \int_{\mathcal{M}} \sum_{k=1}^{\infty} \left| \langle \phi_k, f_0 \rangle e^{-\lambda_k t} \nabla \phi_k(x) \right| \mathrm{d}x\mathrm{d}t && \text{(triangle ineq.)} \\
&= \sum_{k=1}^{\infty} \int_0^T e^{-\lambda_k t} \mathrm{d}t \cdot \int_{\mathcal{M}} |\langle \phi_k, f_0 \rangle \nabla \phi_k(x)| \, \mathrm{d}x \\
&\leq \sum_{k=1}^{\infty} \frac{1}{\lambda_k} |\hat{f}_0(k)| \cdot \int_{\mathcal{M}} |\nabla \phi_k(x)| \, \mathrm{d}x && (T \to \infty) \\
&\leq \sum_{k=1}^{\infty} \frac{1}{\lambda_k} |\hat{f}_0(k)| \cdot \|\nabla \phi_k\|_2 \cdot \sqrt{\mathrm{Vol}(\mathcal{M})} && \text{(Cauchy Schwarz ineq.)} \\
&= \sum_{k=1}^{\infty} \frac{1}{\sqrt{\lambda_k}} |\hat{f}_0(k)| \cdot \sqrt{\mathrm{Vol}(\mathcal{M})} && (\|\nabla \phi_k\|_2 = \sqrt{\lambda_k} \text{ by Green's formula})
\end{aligned}
$$

□

B.2 Proof of Theorem 4.2

The heat diffusion PDE on $\mathbb{T}$:

$$
\begin{cases}
\frac{\partial}{\partial t} u(x,t) - \frac{\partial^2}{\partial x^2} u(x,t) = 0 & x \in [0, 2\pi] \\
u(x,0) = f_0 \in L_0^2([0, 2\pi]) & \text{I.C.} \\
\frac{\partial}{\partial x} u(0,t) = \frac{\partial}{\partial x} u(2\pi, t) = 0 & \text{B.C.}
\end{cases}
$$

and its general solution using Laplacian eigenfunctions ϕ_n:

$$
u(x,t) = \sum_{n=0}^{\infty} \langle \phi_n, f_0 \rangle e^{-\lambda_n t} \phi_n(x) \qquad \text{in which } \phi_n(x) = \frac{1}{\sqrt{\pi}} \cos\left(\frac{n}{2} x\right) \text{ and } \lambda_n = \frac{1}{4} n^2.
$$

Now, Eq. (8) becomes

$$
K(f_0, T) = \int_0^T \int_0^{2\pi} \left| \frac{\partial}{\partial x} u(x,t) \right| \mathrm{d}x\mathrm{d}t
$$

Proof.

$$
\begin{aligned}
K(f-g,\infty) &= \int_0^\infty \int_0^{2\pi} \left| \sum_{n=0}^{\infty} \langle \phi_n, f-g\rangle e^{-\lambda_n t} \phi_n'(x) \right| \mathrm{d}x\mathrm{d}t \\
&= \int_0^\infty \int_0^{2\pi} \left| \sum_{n=1}^{\infty} \langle \phi_n, f-g\rangle e^{-\lambda_n t} \phi_n'(x) \right| \mathrm{d}x\mathrm{d}t \\
&= \int_0^{2\pi} \int_0^{\infty} \left| \sum_{n=1}^{\infty} \langle \phi_n, f-g\rangle e^{-\lambda_n t} \phi_n'(x) \right| \mathrm{d}t\mathrm{d}x && \text{(Fubini theorem)} \\
&\geq \int_0^{2\pi} \left| \int_0^{\infty} \sum_{n=1}^{\infty} \langle \phi_n, f-g\rangle e^{-\lambda_n t} \phi_n'(x) \mathrm{d}t \right| \mathrm{d}x && \text{(triangle ineq.)} \\
&= \int_0^{2\pi} \left| \sum_{n=1}^{\infty} \langle \phi_n, f-g\rangle \frac{1}{\lambda_n} \phi_n'(x) \right| \mathrm{d}x \\
&= \int_0^{2\pi} \left| \sum_{n=1}^{\infty} \langle \phi_n, f-g\rangle \int_0^x \phi_n(s) \mathrm{d}s \right| \mathrm{d}x && (-\phi_n'' = \lambda_n \phi_n) \\
&= \int_0^{2\pi} \left| \int_0^x f(s) - g(s) \mathrm{d}s \right| \mathrm{d}x = W_1(f,g)
\end{aligned}
$$

For the last equation, we used the explicit formula of W_1 in $\mathbb{R}$.[9] □

Acknowledgements

This research was supported in part by N.S.'s NSF grants DMS-1418779 and DMS-1912747. We also thank the `Julia` community who helped our computational and graphical issues we encountered.

REFERENCES

[1] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, Springer, New York, 2nd edition, (2005).

[2] A. Cloninger and S. Steinerberger, *On the dual geometry of Laplacian eigenfunctions*, Exp. Math., (2018), pp. 1–11. Published online: 27 Dec 2018.

[3] J. Dodziuk, *Difference equations, isoperimetric inequality and transience of certain random walks*, Trans. Amer. Math. Soc., 284 (1984), pp. 787–794.

[4] A. Grigor'yan, *Heat Kernel and Analysis on Manifolds*, vol. 47 of Studies in Advanced Mathematics, Amer. Math. Soc., Providence, RI, (2009).

[5] D. K. Hammond, P. Vandergheynst, and R. Gribonval, *Wavelets on graphs via spectral graph theory*, Appl. Comput. Harm. Anal., 30 (2011), pp. 129–150.

[6] H.Birkholz, *A unifying approach to isotropic and anisotropic total variation denoising models*, J. Comput. Appl. Math., 235 (2011), pp. 2502–2514.

[7] J. Irion and N. Saito, *Applied and computational harmonic analysis on graphs and networks*, in Wavelets and Sparsity XVI, Proc. SPIE 9597, M. Papadakis, V. K. Goyal, and D. Van De Ville, eds., 2015. Paper # 95971F.

[8] J. Irion and N. Saito, *Efficient approximation and denoising of graph signals using the multiscale basis dictionaries*, IEEE Trans. Signal and Inform. Process. Netw., 3 (2017), pp. 607–616.

[9] E. Levina and P. Bickel, *The earth mover's distance is the mallows distance: Some insights from statistics*, in Proc. Int'l Conf. Computer Vision, vol. 2, 2001, pp. 251–256.

[10] Y. Nakatsukasa, N. Saito, and E. Woei, *Mysteries around the graph Laplacian eigenvalue 4*, Linear Algebra Appl., 438 (2013), pp. 3231–3246.

[11] S. Nicaise, *Some results on spectral theory over networks, applied to nerve impulse transmission*, in Polynômes Orthogonaux et Applications, e. a. C. Brezinski, ed., vol. 1171 of Lecture Notes in Mathematics, New York, Springer-Verlag, 1985, pp. 532–541.

[12] 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.

[13] N. Saito and E. Woei, *Analysis of neuronal dendrite patterns using eigenvalues of graph Laplacians*, JSIAM Letters, 1 (2009), pp. 13–16.

[14] ——, *On the phase transition phenomenon of graph Laplacian eigenfunctions on trees (recent development and scientific applications in wavelet analysis)*, RIMS Kôkyûroku, 1743 (2011), pp. 77—-90.

[15] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, *The emerging field of signal processing on graphs*, IEEE Signal Processing Magazine, 30 (2013), pp. 83–98.

[16] D. I. Shuman, B. Ricaud, and P. Vandergheynst, *A windowed graph Fourier transform*, in Proc. 2012 IEEE Workshop on Statistical Signal Processing, 2012, pp. 133–136.

[17] G. Strang, *The discrete cosine transform*, SIAM Review, 41 (1999), pp. 135–147.

[18] A. I. Vol'pert, *Differential equations on graphs*, Math. USSR-Sb., 17 (1972), pp. 571–582.

[19] U. von Luxburg, *A tutorial on spectral clustering*, Stat. Comput., 17 (2007), pp. 395–416.

[20] Q. Xia, *Motivations, ideas and applications of ramified optimal transportation*, ESAIM: Mathematical Modelling and Numerical Analysis, 49 (2015), pp. 1791–1832. Special Issue – Optimal Transport.