Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Found Data Sci. Author manuscript; available in PMC 2021 September 02.

Published in final edited form as:

Found Data Sci. 2021 March ; 3(1): 67–97. doi:10.3934/fods.2021006.

HERMES: PERSISTENT SPECTRAL GRAPH SOFTWARE

Rui Wang,
Department of Mathematics, Michigan State University, MI 48824, USA

Rundong Zhao,
Department of Computer Science and Engineering, Michigan State University, MI 48824, USA

Emily Ribando-Gros,
Department of Computer Science and Engineering, Michigan State University, MI 48824, USA

Jiahui Chen,
Department of Mathematics, Michigan State University, MI 48824, USA

Yiying Tong*,
Department of Computer Science and Engineering, Michigan State University, MI 48824, USA

Guo-Wei Wei*
Department of Mathematics, Department of Electrical and Computer Engineering, Department of Biochemistry and Molecular Biology, Michigan State University, MI 48824, USA

Abstract

Persistent homology (PH) is one of the most popular tools in topological data analysis (TDA), while graph theory has had a significant impact on data science. Our earlier work introduced the persistent spectral graph (PSG) theory as a unified multiscale paradigm to encompass TDA and geometric analysis. In PSG theory, families of persistent Laplacian matrices (PLMs) corresponding to various topological dimensions are constructed via a filtration to sample a given dataset at multiple scales. The harmonic spectra from the null spaces of PLMs offer the same topological invariants, namely persistent Betti numbers, at various dimensions as those provided by PH, while the non-harmonic spectra of PLMs give rise to additional geometric analysis of the shape of the data. In this work, we develop an open-source software package, called highly efficient robust multidimensional evolutionary spectra (HERMES), to enable broad applications of PSGs in science, engineering, and technology. To ensure the reliability and robustness of HERMES, we have validated the software with simple geometric shapes and complex datasets from three-dimensional (3D) protein structures. We found that the smallest non-zero eigenvalues are very sensitive to data abnormality.

2020 Mathematics Subject Classification.

Primary: 55-04; Secondary: 92-08

*Corresponding authors: Yiying Tong, ytong@msu.edu and Guo-Wei Wei, weig@msu.edu.
(Communicated by Vasileios Maroulas)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Key words and phrases.

Persistent homology; persistent Laplacian; spectral graph theory; topological data analysis; spectral data analysis; simultaneous geometric; topological analyses

1. Introduction.

As a branch of discrete mathematics, graph theory focuses on the relations among vertices or nodes (0-simplices), edges (1-simplices), faces (2-simplices), and their high-dimensional extensions. Benefiting from the capability of graph formulations that encode inter-dependencies among constituents of versatile data into simple representations, graph theory has been regarded as the mathematical scaffold in the study of various complex systems in biology, material science, physical infrastructure, and network science. However, traditional graphs only represent the pairwise relationships between entries. Therefore, hypergraphs, a generalization of graphs that describe the multi-way relationships of mathematical structures have been developed to capture the high-level complexity of data [2, 6]. Mathematically, graphs and hypergraphs are intrinsically related to the simplicial complexes, which have broader use in computational topology. Moreover, many other areas such as algebra, group theory, knot theory, spectral graph theory (SGT), algebraic topology (AT), and combinatorics are closely related to graph theory. Among them, the applications of SGT have been driven by various real-life problems in chemistry, physics, and life science in the past few decades [37, 41].

In its early days, the spectral graph theory studied the properties of a graph by its graph Laplacian matrix and adjacency matrix. Later on, developments in the spectral graph theory involve some geometric flavor. The explicit constructions of expander graphs rely on studying the eigenvalues and isoperimetric properties of graphs. The discrete analog of Cheeger's inequality for graphs in Riemannian geometry is related to the study of manifolds [11]. Specifically, an eigenvalue of the Laplacian of a manifold is related to the isoperimetric constant of the manifold, which motivates the study of graphs by employing manifolds. Benefiting from the increasingly rich connections with differential geometry, the spectral graph theory has entered a new era [13]. One of the critical developments is the Laplacian on a compact Riemannian manifold in the context of the de Rham-Hodge theory [26, 48]. The harmonic part of the Hodge Laplacian spectrum contains the topological information, whereas the non-harmonic part of the Hodge Laplacian spectrum offers additional geometric information for shape analysis [12]. Indeed, the connectivity of a graph/topological space can be revealed from topological invariants. It is well-known that the number of the eigenvalues in the harmonic spectra of qth-order persistent Laplacian represents the dimension of persistent q-cohomology of a graph [22, 24, 44], which builds the connection between spectral graph theory and algebraic topology.

Homology and cohomology are key concepts in the algebraic topology, which were developed to analyze and classify manifolds according to their cycles. The traditional homology is genuinely metric-independent, indicating that the geometric information is barely considered [25]. Therefore, for practical computation, a new branch of algebraic

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

topology named persistent homology (PH) [9, 20, 49] is implemented to create a sequence of topological spaces characterized by a filtration parameter, such as the radius of a ball or the level set of a real-valued function. As the most important realization of topological data analysis (TDA) [7, 15, 17], topological persistence has had great success in computational chemistry [28, 42] and biology [8, 14, 29, 40, 46]. For instance, the superior performance of using PH features of protein-drug complexes in the free energy prediction and ranking at D3R Grand Challenges, a worldwide competition series in computer-aided drug design [38], was a remarkable success for TDA. Additionally, a weighted persistent homology is proposed as a unified paradigm for the analysis of the biomolecular data system [32].

Recently, we have introduced persistent spectral graph (PSG) theory to bridge persistent homology and spectral graph theory [44, 44]. The PSG theory extends the persistence notion or multiscale analysis to algebraic graph theory. A family of spectral graphs induced by a filtration overcome the difficulty of using traditional spectral graph theory in analyzing graph structures with a single geometry, giving rise to persistent spectral analysis (PSA). Additionally, the evolution of the null space dimension of the persistent Laplacian matrix (PLM) over the filtration offers the topological persistence. Therefore, PSG theory provides simultaneous TDA and PSA. Specifically, by varying a filtration parameter, a series of qth-order persistent Laplacians (or q-persistent Laplacian) provide persistent spectra. Notably, the persistent harmonic spectra of 0-eigenvalues span the null space of the q-th order persistent Laplacian and fully recover the persistent q-th Betti numbers or persistent barcodes [10] of the associated persistent homology. Specifically, the number of 0-eigenvalues of qth-order persistent Laplacian reveals the number of q-cocycles for a given point-cloud dataset. Moreover, the additional geometric shape information of the data will be unveiled in the non-harmonic spectra. For example, the spectral gap (the difference between the moduli of the first two smallest eigenvalues of a Laplacian) reveals the energy difference/density changes between the ground state and first excited state of a system/dataset. Additionally, the B-factor prediction performance can be significantly improved by using the non-harmonic spectra involved in the prediction model, as discussed in [44]. Recently, the theoretical properties and algorithms of PSGs have been further studied [31] and the application of PSG methods to drug discovery has been reported [33]. The de Rham-Hodge theory counterpart, called evolutionary de Rham-Hodge theory, has also been formulated [12].

Currently, many open-source packages have been developed for the applications of persistent homology, including Ripser [4], Dionysus [35], Gudhi [39], Perseus [34], DIPHA [5], Javaplex [1], CliqueTop [23], DioDe [36], Hera, Eirene, and “TDA” package in R [21]. These packages are able to construct a family of complexes with the point clouds data as input and calculate its corresponding Betti numbers, which are equivalent to the harmonic spectra of the persistent Laplacian. However, there is no software package for simultaneous TDA and PSA. While we developed the theoretical part of the persistent spectral graph in 2019, we have not constructed an efficient and robust software yet.

The objective of the present work is to provide the first open-source package, dubbed highly efficient robust multidimensional evolutionary spectra (HERMES), for evaluating both the harmonic and non-harmonic spectra of persistent Laplacian matrices, which enable broad

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and convenient applications of the PSG method. In the present release, we consider an implementation in both alpha complexes [19] and Vietoris–Rips complexes. To verify the reliability of HERMES, 15 complicated 3D structures of proteins as well as two fullerene structures are used to calculate the spectra of qth-order persistent Laplacians for $q = 0, 1, 2$. Moreover, as a validation, the persistent harmonic spectra generated by HERMES are compared with those obtained from Gudhi and DioDe. Furthermore, with the use of the spectra of PLMs, molecular data abnormality detection is also discussed. In a nutshell, HERMES provides a powerful tool in various applications such as drug discovery, protein flexibility analysis, and complex protein structures analysis. It can be potentially applied to various fields where persistent homology has had success.

2. Method.

As a powerful and versatile data representation that encodes inter-dependencies among constituents, graph theory has widely spread applications in various fields such as molecular sciences, engineering, physics, biology, algebra, topology, and combinatorics. In this section, we first briefly review the concepts of simplex, simplicial complex, chain complex, Delaunay complex, and alpha complex in topology, which can be regarded as generalizations of a graph into its higher-dimensional topological counterparts. Then, we review the qth-order Laplacian for simplicial complexes, which is a generalization of the graph Laplacian in graph theory. The topological and geometric information of a single configuration can be evaluated from the spectra of the qth-order Laplacian. Moreover, built upon these concepts, we will discuss persistent spectral graph [44, 44] for the analysis of topological invariants and geometric measurements of high-dimensional datasets. Instead of analyzing the spectra for only one configuration, the persistent spectral graphs can analyze a series of topological and geometric changes, which enriches the set of available representations for high-dimensional datasets.

2.1. Topological concepts.

In this section, we give a concise review of simplex, simplicial complex, and chain complex to provide essential background for persistent spectral graphs. More details can be found in the literature [9, 20, 49].

Simplex.—A q-simplex denoted as σ_q is the convex hull of q+1 affinely independent points in $\mathbb{R}^n$, having dimension $\dim(\sigma_q) = q$. For example, a vertex is a 0-simplex, an edge is a 1-simplex, a triangle is a 2-simplex, and a tetrahedron is a 3-simplex. We call the convex hull of each non-empty subset of $q + 1$ points a face of σ_q, and each of its corner points is also called one of its vertices.

Simplicial complex.—A set of simplices is a simplicial complex denoted as K if the following conditions are satisfied:

1. If all faces of any simplex in K are also in K, and
2. The non-empty intersection of any two simplices in K is a common face of the two simplices.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The dimension of simplicial complex K is defined as $\dim(K) = \max\{\dim\sigma_q : \sigma_q \in K\}$.

Chain complex.—A q-chain is a formal sum of q-simplices in simplicial complex K with $\mathbb{Z}_2$ coefficients. The set of all q-chains has a basis which the set of q-simplices in K, thus forming a finitely generated free abelian group denoted as $C_q(K)$. The boundary operator is a group homomorphism defined by $\partial_q : C_q(K) \rightarrow C_{q-1}(K)$ to relate the chain groups. More specifically, denoting q-simplex as $\partial_q : C_q(K) \rightarrow C_{q-1}(K)$ by its vertices v_i, the boundary operator is defined through its action on the basis,

$$\partial_q \sigma_q = \sum_{i=0}^{q} (-1)^i \sigma_{q-1}^i . \tag{1}$$

Here, $\sigma_{q-1}^i = \left[v_0, \cdots, \hat{v}_i, \cdots, v_q\right]$ is the $(q-1)$-simplex with v_i omitted. The following sequence of chain groups connected by boundary operators is a *chain complex* (defined as a set of abelian groups connected by homomorphisms such that the composite of any two consecutive homomorphisms is zero, $\partial_q \partial_{q+1} = 0$.)

$$\cdots \xrightarrow{\partial_{q+2}} C_{q+1}(K) \xrightarrow{\partial_{q+1}} C_q(K) \xrightarrow{\partial_q} C_{q-1}(K) \xrightarrow{\partial_{q-1}} \cdots$$

2.2. Combinatorial Laplacians.

Combinatorial Laplacians [18] offer both spectral analysis and topological analysis [24]. One central role played by the chain complex associated with a simplicial complex is to define its q-th homology group ($H_q = \ker\partial_q / \mathrm{im}\partial_{q+1}$), which is a topological invariant of the simplicial complex. The dimension of H_q is denoted by $\beta_q = \dim H_q$, the q-th Betti number, which, roughly speaking, measures the number of q-dimensional holes in the simplicial complex, or the geometric object tessellated into the simplicial complex.

A dual chain complex can be defined on any chain complex through the adjoint operator of ∂_q defined on the dual spaces $C^q(K) = C_q^*(K)$. The q-coboundary operator $\partial_q^* : C^{q-1}(K) \rightarrow C^q(K)$ is defined as:

$$\partial^* \omega^{q-1}\left(c_q\right) \equiv \omega^{q-1}\left(\partial c_q\right), \tag{2}$$

where $\omega^{q-1} \in C^{q-1}(K)$ is a $(q-1)$-*cochain*, which is a homomorphism mapping a chain to the coefficient group, and $c_q \in C_q(K)$ is a q-chain. The homology of the dual chain complex is often called *cohomology*.

If we denote by $\mathscr{B}_q$ the matrix representation of a q-boundary operator with respect to the standard basis for $C_q(K)$ and $C_{q-1}(K)$, the number of rows and the number of columns in $\mathscr{B}_q$ correspond to the number of $(q-1)$-simplices and that of q-simplices in K, respectively. Moreover, the matrix representation of q-coboundary operator is denoted $\mathscr{B}_q^T$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In de Rham-Hodge theory, homology and cohomology are often studied through their correspondences to the q-combinatorial Laplacian operator, defined as the linear operator $\Delta_q : C^q(K) \rightarrow C^q(K)$ as follows,

$$\Delta_q := \partial_{q+1}\partial_{q+1}^* + \partial_q^*\partial_q, \tag{3}$$

where the isomorphism $C^q(K) \cong C_q(K)$ is assumed, where each q-simplex is mapped to its own dual, i.e., the isomorphism keeps the coefficients of chains and cochains in the standard simplicial basis. Correspondingly, the matrix representation of Δ_q is the qth-order Laplacian, which is denoted $\mathscr{L}_q(K)$,

$$\mathscr{L}_q(K) = \mathscr{B}_{q+1}\mathscr{B}_{q+1}^T + \mathscr{B}_q^T\mathscr{B}_q. \tag{4}$$

Assume the number of q-simplices existing in K to be N_q, then $\mathscr{L}_q(K)$ is an $N_q \times N_q$-matrix. Since the qth-order Laplacian $\mathscr{L}_q(K)$ is symmetric and positive semi-definite, its spectrum consists of only real and non-negative eigenvalues. We denote the spectrum of $\mathscr{L}_q(K)$ as

$$\mathrm{Spec}\left(\mathscr{L}_q(K)\right) = \left\{\lambda_{1,q}, \lambda_{2,q}, \cdots, \lambda_{N_q,q}\right\}.$$

The multiplicity of zero in the spectrum (also called the harmonic spectrum) reveals the topological information β_q, whereas the non-harmonic spectrum encodes further geometric information. The correspondence between the multiplicity of zero spectra of $\mathscr{L}_q(K)$ and the qth Betti number defined in the homology is an important result in de Rham-Hodge theory, [12, 26, 48]

$$\beta_q = \dim\ker\partial_q - \dim\mathrm{im}\ \partial_{q+1} = \dim\ker\mathscr{L}_q(K) = \ \#\,0 \text{ eigenvalues of } \mathscr{L}_q(K). \tag{5}$$

Intuitively, β_0 represents the number of connected components in K, β_1 reveals the number of 1D noncontractible loops or circles in K, and β_2 shows the number of 2D voids or cavities in K.

2.3. Persistent spectral graphs.

Both topological and geometric information can be derived from analyzing the spectra of qth-order Laplacian. However, the information is restricted to those pieces contained in the connectivity of the simplicial complex. A single simplicial complex produces insufficient information for practical problems such as feature extraction for machine learning analysis. To enrich the spectral information, persistent spectral graph (PSG) is proposed by creating a sequence of simplicial complexes induced by varying a filtration parameter, which is inspired by persistent homology as well as our earlier multiscale graph Laplacians [45].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

First, we consider a filtration of simplicial complex K which is a nested sequence of subcomplexes $\left(K_t\right)_{t=0}^{m}$ of the final complex K:

$$\varnothing = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_m = K. \tag{6}$$

For each subcomplex K_t, we denote its corresponding chain group to be $C_q(K_t)$, and the q-boundary operator will be denoted by $\partial_q^t : C_q\left(K_t\right) \to C_{q-1}\left(K_t\right)$. As conventionally done, we define $C_q(K_t)$ for $q < 0$ as the zero group $\{0\}$ and ∂_q^t as a zero map. [1] If $0 < q \leq \dim K_t$, then

$$\partial_q^t\left(\sigma_q\right) = \sum_i^q (-1)^i \sigma_{q-1}^i, \quad \forall \sigma_q \in K_t, \tag{7}$$

with $\sigma_q = \left[v_0, \cdots, v_q\right]$ being any q-simplex, and $\sigma_{q-1}^i = \left[v_0, \cdots, \hat{v}_i, \cdots, v_q\right]$ being the $(q-1)$-simplex constructed by removing v_i. The adjoint operator of ∂_q^t is the coboundary operator $\partial_q^{t\,*} : C^{q-1}\left(K_t\right) \to C^q\left(K_t\right)$, which can be regarded as a map from $C_{q-1}(K_t)$ to $C_q(K_t)$ through the isomorphisms $C^q\left(K_t\right) \cong C_q\left(K_t\right)$ between cochain groups and chain groups.

Similar to the persistent homology, a sequence of chain complexes can be defined as below:

$$\begin{array}{ccccccccccc}
\cdots & C_{q+1}^1 & \underset{\partial_{q+1}^{1\,*}}{\overset{\partial_{q+1}^1}{\rightleftarrows}} & C_q^1 & \underset{\partial_q^{1\,*}}{\overset{\partial_q^1}{\rightleftarrows}} & \cdots & \underset{\partial_2^{1\,*}}{\overset{\partial_2^1}{\rightleftarrows}} & C_1^1 & \underset{\partial_1^{1\,*}}{\overset{\partial_1^1}{\rightleftarrows}} & C_0^1 & \underset{\partial_0^{1\,*}}{\overset{\partial_0^1}{\rightleftarrows}} & C_{-1}^1 = \{0\} \\
 & \mid\cap & & \mid\cap & & & & \mid\cap & & \mid\cap & & \\
\cdots & C_{q+1}^2 & \underset{\partial_{q+1}^{2\,*}}{\overset{\partial_{q+1}^2}{\rightleftarrows}} & C_q^2 & \underset{\partial_q^{2\,*}}{\overset{\partial_q^2}{\rightleftarrows}} & \cdots & \underset{\partial_2^{2\,*}}{\overset{\partial_2^2}{\rightleftarrows}} & C_1^2 & \underset{\partial_1^{2\,*}}{\overset{\partial_1^2}{\rightleftarrows}} & C_0^2 & \underset{\partial_9^{2\,*}}{\overset{\partial_0^2}{\rightleftarrows}} & C_{-1}^2 = \{0\} \\
 & \vdots & & \vdots & & & & \vdots & & \vdots & & \vdots \\
 & \mid\cap & & \mid\cap & & & & \mid\cap & & \mid\cap & & \\
\cdots & C_{q+1}^m & \underset{\partial_{q+1}^{m\,*}}{\overset{\partial_{q+1}^m}{\rightleftarrows}} & C_q^m & \underset{\partial_q^{m\,*}}{\overset{\partial_q^m}{\rightleftarrows}} & \cdots & \underset{\partial_2^{m\,*}}{\overset{\partial_2^m}{\rightleftarrows}} & C_1^m & \underset{\partial_1^{m\,*}}{\overset{\partial_1^m}{\rightleftarrows}} & C_0^m & \underset{\partial_0^{m\,*}}{\overset{\partial_0^m}{\rightleftarrows}} & C_{-1}^m = \{0\}
\end{array} \tag{8}$$

For simplicity, we use C_q^t to denote the chain group $C_q(K_t)$.

Next, we introduce persistence to the Laplacian spectra. We define the subset of C_q^{t+p} whose boundary is in C_{q-1}^t as $\mathbb{C}_q^{t,p}$, assuming the natural inclusion map from C_{q-1}^t to C_{q-1}^{t+p}

[1] We define the boundary matrix $\mathscr{B}_0^t$ for the boundary operator ∂_0^t as a zero matrix. The number of columns of $\mathscr{B}_0^t$ is the number of 0-simplices in K_t, the number of rows will be 1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\mathbb{C}_q^{t,p} := \left\{ \beta \in C_q^{t+p} \middle| \partial_q^{t+p}(\beta) \in C_{q-1}^t \right\}. \tag{9}$$

On this subset, one may define the p-persistent q-boundary operator denoted by $\partial_q^{t,p}: \mathbb{C}_q^{t,p} \to C_{q-1}^t$. Its corresponding adjoint operator is $\left(\partial_q^{t,p}\right)^*: C_{q-1}^t \to \mathbb{C}_q^{t,p}$, again through the identification of cochains with chains. We then define the q-order p-persistent Laplacian operator $\Delta_q^{t,p}: C_q^t \to C_q^t$ associated with the filtration as

$$\Delta_q^{t,p} = \partial_{q+1}^{t,p}\left(\partial_{q+1}^{t,p}\right)^* + \partial_q^{t^*}\partial_q^t. \tag{10}$$

The matrix representation of $\Delta_q^{t,p}$ in the simplicial basis is

$$\mathscr{L}_q^{t,p} = \mathscr{B}_{q+1}^{t,p}\left(\mathscr{B}_{q+1}^{t,p}\right)^T + \left(\mathscr{B}_q^t\right)^T \mathscr{B}_q^t, \tag{11}$$

where $\mathscr{B}_{q+1}^{t,p}$ is the matrix representation of $\partial_{q+1}^{t,p}$.

We denote the spectrum of $\mathscr{L}_q^{t,p}$ as

$$\mathrm{Spec}\left(\mathscr{L}_q^{t,p}\right) = \left\{ \lambda_{1,q}^{t,p}, \lambda_{2,q}^{t,p}, \cdots, \lambda_{N_q^t,q}^{t,p} \right\},$$

where $N_q^t = \mathrm{dim} C_q^t$ is the number of q-simplices in K_t, and the eigenvalues are listed in the ascending order. Thus, the smallest non-zero eigenvalue of $\mathscr{L}_q^{t,p}$ is denoted as $\lambda_{2,q}^{t,p}$. We may recognize the multiplicity of zero in the spectrum of $\mathscr{L}_q^{t,p}$ as the qth order p-persistent Betti number $\beta_q^{t,p}$, which counts the number of (independent) q-dimensional holes in K_t that still exists in K_{t+p}. The relation can be observed in

$$\beta_q^{t,p} = \mathrm{dim}\,\mathrm{ker}\,\partial_q^t - \mathrm{dim}\,\mathrm{im}\,\partial_{q+1}^{t,p} = \mathrm{dim}\,\mathrm{ker}\,\mathscr{L}_q^{t,p} = \#\,0 \text{ eigenvalues of } \mathscr{L}_q^{t,p}. \tag{12}$$

In this paper, we focus on the 0, 1, 2th-order persistent Laplacians, which depict the relations among vertices, edges, triangles, and tetrahedra, as we target 3D real-world applications.

For instance, given a set of vertices $V = \left\{ v_0, v_1, \cdots, v_{N_0-1} \right\}$, N_0 embedded in $\mathbb{R}^3$, we consider a nested family of simplicial complexes that may be created for a positive real number a. Denoting the simplicial complex generated for a by K_a, the traditional qth-order Laplacian is just a special case of qth-order 0-persistent Laplacian at K_a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\mathscr{L}_q^{\alpha,0} = \mathscr{B}_{q+1}^{\alpha,0}\left(\mathscr{B}_{q+1}^{\alpha,0}\right)^T + \left(\mathscr{B}_q^{\alpha}\right)^T \mathscr{B}_q^{\alpha}. \tag{13}$$

The spectrum of $\mathscr{L}_q^{\alpha,0}$ is simply associated with a snapshot of the filtration,

$$\mathrm{Spec}\left(\mathscr{L}_q^{\alpha,0}\right) = \left\{\lambda_{1,q}^{\alpha,0}, \lambda_{2,q}^{\alpha,0}, \cdots, \lambda_{N_q^{\alpha},q}^{\alpha,0}\right\}. \tag{14}$$

Correspondingly, the q-th 0-persistent Betti number $\beta_q^{\alpha,0} = \beta_q^{\alpha}$. In addition to the traditional homology information, and persistent homology information, our proposed persistent spectral graph theory, through the nonzero eigenvalues in the spectrum of the persistent Laplacian operator, provide richer spatial information induced by varying the filtration parameters. Thus it provides a powerful tool to encode high-dimensional datasets into various topological and geometric features in a coherent fashion.[2]

2.4. Delaunay triangulation and alpha shape.

In this section, we provide the details on a practical construction of filtration for persistent spectral graph theory based on the alpha complex. The alpha complex can be regarded as a simplicial complex, which is a homotopy equivalent to the nerve of balls around data points. Its geometric realization built as the union of convex hulls of points in each simplex is called the alpha shape. The alpha shape was first proposed in 1983, which defined the shape associated with a finite set of points in the plane controlled by one parameter [19].

In the following, we first describe how to construct the alpha shape, and then provide some necessary concepts for the implementation of the alpha complex in PSG theory. Let P be a finite set of points in qD Euclidean space $\mathbb{R}^q$ ($q = 2$ or 3 in most applications), and α be a positive real number. Denote an open ball with radius α as an alpha ball (α-ball). We say that an α-ball is empty if it contains no point of P, and the alpha hull (α-hull) of P is the set of points that do not belong to any empty α-ball. For any subset $T \subseteq P$ with size $|T| = k + 1$, $0 \le k \le q$, the geometric realization of k-simplex σ_T is the convex hull of T. We say that a k-simplex σ_T is α-exposed if there exists an empty α-ball $\mathbf{b}$ such that $T = \partial\mathbf{b} \cap P$ for $0 \le k \le q - 1$. Denoting the collection of α-exposed k-simplices as $F_{k,\alpha}$ for $0 \le k \le q - 1$, the alpha shape (α-shape) of P is the polytope whose boundary consists of the k-simplices in $F_{k,\alpha}$. The alpha complex is just the simplicial complex that is the collection of the simplices in the alpha shape.

There are two structures that are closely related to the alpha shape and helpful in efficient implementation of alpha shape and alpha complex. One is the Voronoi diagram [43] and the other is its dual structure, the Delaunay tessellation [16]. The latter is the alpha complex for

[2]In this work, we use notations $\mathbb{C}_q^{t,p}$, $\partial_q^{t,p}$, $\Delta_q^{t,p}$, $\mathscr{L}_q^{t,p}$, and $\beta_q^{t,p}$ instead of $\mathbb{C}_q^{t+p}$, ∂_q^{t+p}, Δ_q^{t+p}, $\mathscr{L}_q^{t+p}$, and β_q^{t+p} used in Ref. [44].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sufficiently large α, e.g., when α is greater than the diameter of P. Thus, the Delaunay tessellation is the final complete simplicial complex in the filtration that we use.

For a given set of points $P = \{p_1, p_2, \cdots, p_n\} \subseteq \mathbb{R}^q$, the Voronoi cell V_i of a point $p_i \in P$ contains all of the points for which p_i is the closest among all the points in P,

$$V_i = \left\{ x \in \mathbb{R}^q \middle| \ \|x - p_i\| \leq \|x - p_j\|, \ \forall p_j \in P \right\}. \tag{15}$$

The Voronoi diagram of P is the set of Voronoi cells, which is defined as

$$\mathrm{Vor}P = \left\{ V_i \middle| \forall i \in \{1, 2, \cdots, |P|\} \right\}. \tag{16}$$

The Delaunay tessellation for a given set P in general position (i.e., no $q + 1$ ponits are in a (q−1)-D linear subspace, and no $q + 2$ points share the same circumsphere) is the dual simplicial complex to the Voronoi diagrams. For instance, a Delaunay tessellation for a given set P in 2D is a triangulation DT(P) such that no point in P is inside the circumcircle of any triangle in DT(P) [3, 30]. A formal way to define the Delaunay tessellation is to use the nerve of the collection of Voronoi cells (Nrv(Vor$P$)), which can be expressed as

$$\mathrm{DT}(P) = \mathrm{Nrv}(\mathrm{Vor}P) = \left\{ J \subseteq \{1, 2, \ldots, |P|\} \ \middle| \ \bigcap_{i \in J} V_i \neq \varnothing \right\}, \tag{17}$$

under the condition that the points in P are general position. Note that, in practice, a set of points that are not in general position can be symbolically perturbed to general position.

Next, we introduce the mathematical description of the construction of alpha complex through the union of balls centered at points in P, which is essentially a van der Waals surface for atoms positioned at P with the same radius α. For a given set of points $P = \{p_1, p_2, \cdots, p_n\}$ in $\mathbb{R}^q$ and a positive real number α, we can denote the closed ball centered at p_i as $B_i(\alpha) = p_i + \alpha\mathbb{B}^q$, where $\mathbb{B}^q$ is a qD unit ball around the origin. The union of these balls can be expressed as

$$U(\alpha) = \left\{ x \in \mathbb{R}^q \middle| \exists p_i \in P \text{ s.t. } \|x - p_i\| \leq \alpha \right\}. \tag{18}$$

To ensure that we obtain a subcomplex of the Delaunay tessellation, we intersect $B_i(\alpha)$ with its corresponding Voronoi cell,

$$R_i(\alpha) = B_i(\alpha) \cap V_i. \tag{19}$$

It can be observed that $U(\alpha) = \cup_{p_i \in P} R_i(\alpha)$, so the R_i's is a covering of $U(\alpha)$. The alpha complex K_α is the simplicial complex representing the nerve of this covering,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$K_\alpha = \left\{ J \subseteq \{1, 2, \ldots, |P|\} \;\middle|\; \bigcap_{i \in J} R_i(\alpha) \neq \varnothing \right\}. \tag{20}$$

The equivalence to the original definition can be readily checked. The union of all simplices in the alpha complex forms the alpha shape. Figure 1 illustrates the Voronoi diagram, Delaunay triangulation, and non-Delaunay triangulation. The point set is $P = \{A,B,C,D,E\}$, and the blue lines in the left chart of Figure 1 separate the plane into the Voronoi cells. The red circles are the empty circumcircles for triples of points in P. We can notice that no four points are on the same red circle, which satisfies the uniqueness condition for constructing the Delaunay triangulation. In the right chart of Figure 1, the green circumcircle of ACD contains E and the green circumcirlce of AEC contains D, indicating that those two triangles do not belong to the Delaunay triangulation.

Figure 2 illustrates the standard filtration of alpha complexes. The top left figure is the Delaunay triangulation of six 2D points A, B, C, D, E, and F. With an ever-growing radius α centered at these points, a family of sub-complexes of the Delaunay triangulation can be constructed. Figure 3 shows the persistence barcode of these 6 points. It can be seen that when $\alpha = 0.2$, all six points are disconnected, indicating that 6 0-cycles (connected components) existed, which matches with Figure 3, where there are a total of 6 bars when $\alpha = 0.2$. With the radius α continually increasing, a 1-cycle will be formed, and the associated alpha shape are shown in the bottom left chart of Figure 2. One can notice that in Figure 3, when $\alpha = 0.6$, $\beta_1^{\alpha,0} = 1$. When α reaches 0.83, the 1-cycle disappears and $\beta_1^{\alpha,0} = 0$ as shown in the bottom left panel of Figure 2. Table 1 and Table 2 show how we construct the qth-order persistent Laplacian $\mathscr{L}_q^{t,p}$ and calculate the harmonic $\left(\beta_q^{t,p}\right)$ and non-harmonic persistent spectra of $\mathscr{L}_q^{t,p}$ from the simplicial complexes $K_{0.2}$ to $K_{0.6}$ and $K_{0.6}$ to $K_{0.6}$.

2.5. Vietoris–Rips complex.

Vietoris-Rips complex is an abstract simplicial complex. It is commonly used in various applications. For a given set of points $P = \{p_1.p_2, \cdots, p_n\}$ in a metric space and a real value $r > 0$, a k-simplex $\sigma_k = [p_{i0}, \cdots, p_{ik}]$ is in the Vietoris–Rips complex if and only if $\mathbb{B}\left(p_{ij,r}\right) \cap \mathbb{B}\left(p_{ij',r}\right) \neq \varnothing, \forall j, j' \in [0, k]$.

3. Implementation.

3.1. Construction of alpha shape.

Recall that, given a set of points, the alpha shape with any α value is a subcomplex of Delaunay tessellation. Thus, to construct the filtration of alpha complexes, it is necessary to first compute the complete simplicial complex through the Delaunay tessellation formed by the set of points. A number of efficient implementations is available in existing software packages. Our implementation employs the Computational Geometry Algorithms Library (CGAL), an efficient and robust software package for many commonly used calculations. We then assign each simplex σ with an alpha value α_σ. Finally, the alpha shape given at an α value α_0 is constructed by union of convex hulls of all the simplices σ satisfying $\alpha_\sigma \leq \alpha_0$,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

which naturally forms the nerve of balls centered at the given points truncated by the Voronoi regions, i.e., the corresponding alpha complex.

We illustrate our implementation with point sets P in 3D, as it is the most common use scenario. We also assume that all the points are in general positions, which means that no 4 points of P lie on the same plane and no 5 points of P lie on the same sphere. Given a simplex σ, which can be a point, an edge, a triangle or a tetrahedron, denote the open ball bounded by its minimal circumsphere as B_σ. The simplex σ is called *Gabriel* ([27]) if $B_\sigma \cap P = \varnothing$. Note that for vertices (0-simplices) the circumradius is considered 0. The above discussion can be directly adapted for 2D implementation by replacing circumsphere with circumcircle and omitting tetrahedra.

The filtration parameter α for every simplex σ can be defined as follows. If the simplex is Gabriel, the filtration value is the corresponding circumradius (for efficiency, we actually store its square) because the corresponding ball can be considered as an empty α-ball touching all its vertices. If the simplex is not Gabriel, the filtration value is the minimum of all the filtration values of the cofaces of σ that contain the points making the simplex non-Gabriel. When α value reaches that number, we will have an empty α-ball making the simplex α-exposed.

3.2. Implementation details for alpha shape.

To ensure the valid calculation of the filtration parameter for non-Gabriel simplices, the filtration value are always computed from the highest dimension (tetrahedra) down to 0 (vertices). We initialize the filtration value for all the simplices to be positive infinity. For dimension k, we iterate through each k-simplex. If the current filtration value α_σ^2 is positive infinity, we assign the filtration value as the square of the corresponding circumradius. Then, we check every (k−1)-dimensional face τ in $\partial\sigma$. If the circumsphere of τ enclosed the other vertex of σ in the interior, it is not Gabriel, and does not correspond to an empty α-ball. In this case, α_σ^2 is assigned to α_τ^2 if $\alpha_\sigma > \alpha_\tau$.

With this procedure, we ensure that α_σ for every simplex σ corresponding to the filtration value α is α-exposed to an empty α-ball. In other words, we ensure that for each simplex represented by its vertex index set $J \subseteq \{1, 2, \ldots, |P|\}$ is in the nerve of R_i's, which are the intersections $R_i = V_i \cap B_i$ of Voronoi cells V_i's and balls B_i's around the points p_i's.

3.2.1. Boundary operator construction.—With α_σ assigned, we sort the k-simplices with increasing filtration parameter value. This allows us to construct a single boundary operator B_q^∞ (the matrix representation of ∂_q^∞) for the entire filtration, which is that of the Delaunay tessellation. For any given α, we can read of the top left block of the full boundary matrix B_q^∞, i.e.,

$$\left(B_q^\alpha\right)_{ij} = \left(B_q^\infty\right)_{ij}, \quad \forall 1 \le i \le N_{q-1}^\alpha, 1 \le j \le N_q^\alpha, \tag{21}$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where N_q^α is the number of q-simplices in the alpha complex with the filtration parameter α. Alternative, we can consider the $N_q^\alpha \times N_q^\infty$ projection matrix P_q^α from the Delaunay tessellation to the alpha complex, $\left(P_q^\alpha\right)_{ij} = \delta_{ij}$ (1 on the diagonal and 0 elsewhere), with which we have $B_q^\alpha = P_{q-1}^\alpha B_q^\infty \left(P_q^\alpha\right)^T$.

3.2.2. Persistent boundary operator.—The construction of p-persistent boundary matrix $B_q^{\alpha,p}$ (the representation of operator $\partial_q^{\alpha,p}$ is more involved than reading off B_q^∞.

We first construct the projection matrix $\mathbb{P}_q^{\alpha,p}$ from $C_q^{\alpha+p}$ to $\mathbb{C}_q^{\alpha,p}$. Then, the p-persistent boundary matrix can be assembled as $B_q^{\alpha,p} = P_{q-1}^\alpha B_q^\infty \left(\mathbb{P}_q^{\alpha,p}\right)^T$.

To construct the projection matrix, we first note that it is the projection to the kernel of an operator that measures the difference between the boundary operator mapped onto $C_{q-1}^{\alpha+p}$ and the boundary restricted to C_{q-1}^α, $\mathrm{Diff}_q^{\alpha,p} = \left(I_{q-1}^{\alpha+p} - R_{q-1}^{\alpha,p}\right)^T B_q^{\alpha+p}$, where $R_q^{\alpha,p} = P_q^{\alpha+p}\left(P_q^\alpha\right)^T P_q^\alpha \left(P_q^{\alpha+p}\right)^T$ is the restriction from $C_q^{\alpha+p}$ to C_q^α and $I_q^{\alpha+p}$ is the identity matrix on $C_q^{\alpha+p}$.

Instead of storing a dense matrix, we propose to use a procedural representation involving the inverse of persistent Laplacians with gauge ([47]) to reduce the storage as well as speed up the computation. More specifically, we construct the projection matrix as follows

$$\mathbb{P}_q^{\alpha,p} = I_q^{\alpha+p} - \left(\tilde{\mathrm{Diff}}_q^{\alpha,p}\right)^T \left(\tilde{L}_{q-1}^{\alpha,p}\right)^{-1} \tilde{\mathrm{Diff}}_q^{\alpha,p}, \tag{22}$$

where $\left(\tilde{L}_{q-1}^{\alpha,p}\right)^{-1}$ can be implemented through rank deficiency fixing in [47], and the restricted operator $\tilde{\mathrm{Diff}}_q^{\alpha,p}$ is defined below. Note that this sparse linear equation solving approach is essentially the graph version of the harmonic extension described in Ref. [48].

The reason that the projection matrix can be defined this way is that starting from an arbitrary element $\omega_q \in C_q^{\alpha+p}$, we can modify it into $\omega_q - \left(\mathrm{Diff}_q^{\alpha,p}\right)^T f_{q-1} \in \mathbb{C}_q^{\alpha,p}$, where f_{q-1} is nonzero only in the difference complex $\mathrm{Cl}\left(T_{\alpha+p} - T_\alpha\right)$, the closure of the difference between $T_{\alpha+p}$ and T_α. Denoting any chain f on the difference complex as $\tilde{f}$ and any operator B on it as $\tilde{B}^{\alpha,p}$, and the $\tilde{B}_q^{\alpha,p}\left(\tilde{B}_q^{\alpha,p}\right)^T \tilde{f}_{q-1} = \tilde{B}_q^{\alpha,p}\tilde{\omega}_q$. Noticing that $\tilde{f}_{q-1}$ is determined up to a gauge transform $f_{q-1} - \left(\tilde{B}_{q-1}^{\alpha,p}\right)^T \tilde{g}_{q-2}$ for some $(q-2)$-chain g_{q-2} in $\mathrm{Cl}(T_{\alpha+p} - T_\alpha)$, we introduce the gauge fixing term $\tilde{B}_{q-1}^{\alpha,p} f_{q-1} = 0$, which leads us to the sparse linear system $\tilde{L}_{q-1}^{\alpha+p}\tilde{f}_{q-1} = \tilde{\mathrm{Diff}}_q^{\alpha,p}\omega_q$ where the $\tilde{\mathrm{Diff}}$ operator is the above operator projected to the difference complex. Note that fixing the rank deficiency of persistent Laplacians (in the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

difference complex) is computationally efficient as its kernel dimension is far smaller than that of the corresponding boundary or coboundary operators.

3.2.3. Persistent spectrum computation.—The q-order p-persistent Laplacian operators can then be implemented by direct evaluation of $L_q^{\alpha,p} = B_{q+1}^{\alpha,p}\left(\beta_{q+1}^{\alpha,p}\right)^T + \left(B_q^{\alpha}\right)^T B_q^{\alpha}$. Their spectra can be evaluated through any off-the-shelf sparse matrix eigensolver.

Thus, the dimension of the null space of $L_0^{\alpha,p}$ is number of p-persistent connected components. The dimension of the null space of $L_1^{\alpha,p}$ is number of p-persistent handles or tunnels. Similarly, the dimension of the null space of $L_2^{\alpha,p}$ is the number of p-persistent cavities.

3.3. Implementation details for Vietoris–Rips complex.

The Vietoris–Rips complex at different filtration values is also considered in HERMES. Following the definition of the Vietoris–Rips complex, the implementation is straightforward. However, due to large number of simplices, the calculation of non-harmonic spectra of PLMs $L_q^{t,p}$ can be resource-intensive. Therefore, we may set a maximum cutoff distance for the filtration r and an upper limit for persistent p for practical applications.

4. Validation.

We construct the alpha complex at different filtration values from the finite cells of a Delaunay tessellation from the Computational Geometry Algorithms Library (CGAL). Moreover, the Vietoris–Rips complex at different filtration values is also constructed in the HERMES. Gudhi and DioDe are two of the most frequently applied open-source libraries that are able to compute the Betti numbers (harmonic persistent spectra) based on CGAL, while Ripser is based on the blazing fast C++ Ripser package. As shown in [44], the 0-persistent qth Betti numbers $\beta_q^{t,0}$ at filtration parameter t is the number of zero eigenvalues of qth-order 0-persistent Laplacian $\mathscr{L}_q^{t,0}$:

$$\beta_q^{t,0} = \dim\left(C_q^t\right) - \operatorname{rank}\left(\mathscr{L}_q^{t,0}\right) = \dim\ker\mathscr{L}_q^{t,0}, \tag{23}$$

where $t = \alpha$ if we choose to construct alpha complex, and $t = r$ if we choose to construct Vietoris–Rips complex.

In fact, $\beta_q^{t,0}$ counts the number of q-cycles in alpha complex K_t that persists in K_t. Although Gudhi and DioDe can calculate the number of zero eigenvalues, the non-harmonic persistent spectra also play an important role in applications as shown in our earlier work [44]. Therefore, we developed an open-source package HERMES, which not only tracks the topological changes from the persistent Betti numbers but also derives the geometric changes from the non-harmonic spectra of persistent Laplacians. In the following, we compare the Betti numbers $\beta_q^{t,p}$ that are calculated from HERMES with the Betti numbers

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that are derived from Gudhi and DioDe on a set of 2D and 3D points, aiming to validate the robustness and accuracy of HERMES.

4.1. Validation on fullerene structures.

In this section, we will validate the correctness of HERMES with simple systems such as C_{20} and C_{60} molecules with known persistent Betti numbers [46] for Rips complex. Moreover, the persistent Betti numbers for alpha complex are also included in this section.

C_{20} molecule.—C_{20} molecule is the smallest member of the fullerene family, which has a dodecahedral cage structure as illustrated in Figure 4 (a). Both C_{20} and C_{60} have the molecular symmetry of the full icosahedral point group I_h. Figure 5 illustrates the persistent Betti numbers for Rips complex $\beta_0^{r,0.05}$, $\beta_1^{r,0.05}$, and $\beta_2^{r,0.05}$ (green curves) and the smallest non-zero eigenvalue $\lambda_0^{r,0.05}$, $\lambda_1^{r,0.05}$, and $\lambda_2^{r,0.05}$ (yellow curves) of C_20 that are computed from HERMES. Similarly, Figure 6 illustrates the persistent Betti numbers for alpha complex $\beta_0^{\alpha,0.05}$, $\beta_1^{\alpha,0.05}$, and $\beta_2^{\alpha,0.05}$ (green curves) and the smallest non-zero eigenvalue $\lambda_0^{\alpha,0.05}$, $\lambda_1^{\alpha,0.05}$, and $\lambda_2^{\alpha,0.05}$ (yellow curves) of C_20 that are computed from HERMES.

Note that although the Rips complex and the alpha complex have similar Betti-0 and Betti-1 patterns, their Betti-2 patterns differ from each other over the filtration range. Additionally, the non-harmonic spectra of the Rips complex and the alpha complex differ much from each other. Moreover, the non-harmonic spectra of the Rips complex appear to carry more information than those of the alpha complex.

C_{60} molecule.—C_{60} molecule is a well-known structure that also called buckminsterfullerene. A total of 12 pentagon rings and 20 hexagon rings consist of C_{60}. Figure 4 (b) shows the 3D structure of and C_{60}. Figure 7 and Figure 8 demonstrate the 0.05-persistent Betti numbers for rips complex and alpha complex, respectively. Figure 5 – Figure 8 indicate the capacity of HERMES for the direct calculation of the persistent spectra of $\mathscr{L}_q^{r,p}$ and $\mathscr{L}_q^{\alpha,p}$ ($p > 0$).

4.2. Validation on proteins.

In this section, we further validate HERMES using 15 proteins. Their Protein Data Bank (PDB) IDs of these proteins are 1CCR, 1NKO, 1O08, 1OPD, 1QTO, 1R7J, 1V70, 1W2L, 1WHI, 2CG7, 2FQ3, 2HQK, 2PKT, 2VIM, and 5CYT. The 3D structures of these 15 proteins can be downloaded from the PDB). Here, only the alpha carbon atoms are considered in our calculations. The harmonic spectra of HERMES are compared with the persistent Betti numbers of Gudhi and DioDe. Figure 9 illustrates the network structures of 15 proteins. For each protein, color at atomic positions represents the normalized diagonal values of the accumulated 0th-order 0-persistent Laplacians: $\frac{1}{\max_i\left(\mathscr{L}_0^0\right)_{ii}}\left(\mathscr{L}_0^0\right)_{jj}$, with $\mathscr{L}_0^0 = \sum_\alpha \mathscr{L}_0^{\alpha,0}$. Here, the filtration α goes from $\sqrt{1.5}$ Å to $\sqrt{10}$ Å with the step size of 0.01 Å Figure 10 depicts the persistent Betti numbers $\beta_q^{\alpha,0}$ (blue curve) of PDB ID 5CYT that are

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

calculated from Gudhi, DioDe, and HERMES, together with the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) that are obtained only from HERMES.

It can be seen that all of these three packages return exactly the same persistent Betti numbers, suggesting that the calculation of our package HERMES is reliable. Additionally, the values of smallest non-zero eigenvalues $\lambda_0^{\alpha,0}$ and $\lambda_1^{\alpha,0}$ increase around 1.86 Å indicating the dramatic topological changes at this point. Similarly, with the increment of the α, the curve of $\lambda_2^{\alpha,0}$ also records the topological and geometric changes at a specific filtration value. The use of non-harmonic spectra for biophysical modeling was described in our earlier work [44].

To be noted, HERMES can also deal with the qth-order p-persistent Laplacians $\mathscr{L}_q^{\alpha,p}$. Figure 11 illustrates the persistent Betti numbers $\beta_0^{\alpha,0.5}$, $\beta_1^{\alpha,0.5}$, and $\beta_2^{\alpha,0.5}$ (green curves) and the smallest non-zero eigenvalue $\lambda_0^{\alpha,0.5}$, $\lambda_1^{\alpha,0.5}$, and $\lambda_2^{\alpha,0.5}$ (yellow curves) of 5CYT that are computed from HERMES, demonstrating the capacity of HERMES for the direct calculation of the persistent spectra of $\mathscr{L}_q^{\alpha,p}$ ($p>0$). Compared with the middle chart of Figure 10, the $\beta_1^{\alpha,0.5}$ in the middle chart of Figure 11 is always smaller than $\beta_1^{\alpha,0}$ at the same filtration α. Moreover, the $\lambda_1^{\alpha,0.5}$ also goes up around 1.86 Å, which has the same behavior as $\lambda_1^{\alpha,0}$. Similar behaviors can be also observed from the bottom charts of Figure 10 and Figure 11.

Furthermore, HERMES can be used to detect the abnormality of a protein structure. Figure 12 (a) shows a 3D secondary structure of PDB 1O08, where the balls represent the alpha carbon atoms. The light blue, purple, and orange colors represent helix, sheet, and random coils of PDB ID 1O08. Figure 12 (b) depicts its harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve). Notably, two unusual onset of $\beta_0^{\alpha,0}$ and $\beta_1^{\alpha,0}$ are detected when $\alpha \ll 1.9$ Å, indicating something is wrong with the structure data. Usually, the distance between the two alpha carbon atoms is around 3.8 Å. By examining the structure of PDB 1O08, we found that two pairs of alpha carbon atoms in PDB 1O08 have abnormal distances as marked with black frames. The distance of alpha carbon atoms in the upper box is 2.914 Å and that in the lower box is 2.996 Å which are too short. The plots of the other proteins can be found in the Appendix. Similar structural defects were detected for PDB IDs 1V70, 2HQK, 2PKT, and 2VIM.

Although our package provides additional geometric information by calculating the non-harmonic spectra of qth-order persistent Laplacians, there are two limitations of HERMES. First, the construction of the Vietoris–Rips complex is the primary bottleneck in the calculation of non-harmonic spectra of persistent Laplacian matrices (PLMs). Additionally, the input format of HERMES is point cloud data. Other input formats, such as pairwise distances, point cloud with van der Waals radii, and volumetric density are not supported. These limitations will be addressed in our future implementation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

5. Conclusion.

While spectral graph theory has had tremendous success in data science to capture the geometric and topological information, it is limited by representing a graph structure at a given characteristic length scale, which hinders its practical application in data analysis. Motivated by the persistent (co)homology in dealing with a given initial data by constructing a family of simplicial complexes to track their topological invariants, and the multiscale graphs by creating a set of spectral graphs aiming to extract rich geometric information, we proposed persistent spectral graph (PSG) theory as a unified multiscale paradigm for simultaneous geometric and topological analysis [44]. PSG theory has stimulated mathematical analysis and algorithm development [31], as well as applications to drug discovery [33], and protein flexibility analysis [44].

To enable broad and convenient applications of the PSG method, we present an open-source software package called highly efficient robust multidimensional evolutionary spectra (HERMES). For a given point-cloud dataset, HERMES creates persistent Laplacian matrices (PLMs) at various topological dimensions via a filtration. The spectrum of PLMs includes harmonic parts and non-harmonic parts. It turns out that the harmonic part spans the kernel spaces of PLMs and carries the full topological information of the dataset. As a result, HERMES delivers the same topological data analysis (TDA) as does persistent homology. The non-harmonic part of PLMs provides valuable geometric analysis of the shape of data at various topological dimensions. The smallest non-zero eigenvalues are found to be very sensitive to data abnormality. In the present HERMES, both the alpha complex and the Vietoris–Rips complex are implemented. Due to the potentially large number of simplicies, the eigenvalue problem of persistent Laplacian for the Vietoris–Rips complex becomes memory-intensive for large systems. This difficulty may be overcome with approximate eigenvalue solvers. We will continue improving the efficiency of HERMES. HERMES has been extensively validated for its accuracy, robustness, and reliability by standard test datasets and a large number of complex protein structures, including comparison with Gudhi and DioDe.

Acknowledgments

This work was supported in part by NIH grant GM126189, NSF grants DMS-2052983, DMS-1761320, and IIS-1900473, NASA grant 80NSSC21M0023, Michigan Economic Development Corporation, George Mason University award PD45722, Bristol-Myers Squibb 65109, and Pfizer.

Appendix A.

Figure 13 shows the harmonic spectra (under the construction of Vietoris–Rips complex) $\beta_q^{r,0}(q = 0, 1, 2)$ of C_{60} with shifting one of its atoms' position. It can be seen that an abnormality of distance between atoms are detected when the radius r is around 1.38Å. Figure 14 – Figure 26 illustrate the harmonic spectra (under the construction of alpha complex) $\beta_q^{\alpha,0}(q = 0, 1, 2)$ of PDB IDs 1CCR, 1NKO, 1OPD, 1QTO, 1R7J, 1V70, 1W2L, 1WHI, 2CG7, 2FQ3, 2HQK, 2PKT, and 2VIM at different filtration value α calculated from Gudhi, DioDe, and HERMES.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REFERENCES

[1]. Adams H, Tausz A and Vejdemo-Johansson M, JavaPlex: A research software package for persistent (co) homology, in International Congress on Mathematical Software, Lecture Notes in Computer Science, 8592, Springer, 2014, 129–136.

[2]. Aksoy SG, Joslyn C, Marrero CO, Praggastis B and Purvine E, Hypernetwork science via high-order hypergraph walks, EPJ Data Science, 9 (2020).

[3]. Aurenhammer F, Klein R and Lee D-T, Voronoi Diagrams and Delaunay Triangulations, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2013.

[4]. Bauer U, Ripser: A lean C++ code for the computation of Vietoris–Rips persistence barcodes, 2017. Software available from: https://github.com/Ripser/ripser.

[5]. Bauer U, Kerber M and Reininghaus J, DIPHA (A distributed persistent homology algorithm), 2014. Software available from: https://github.com/DIPHA/dipha.

[6]. Bressan S, Li J, Ren S and Wu J, The embedded homology of hypergraphs and applications, Asian J. Math, 23 (2019), 479–500.

[7]. Bubenik P and Kim PT, A statistical approach to persistent homology, Homology Homotopy Appl, 9 (2007), 337–362.

[8]. Cang Z and Wei G-W, TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions, PLoS Computational Biology, 13 (2017).

[9]. Carlsson G, De Silva V and Morozov D, Zigzag persistent homology and real-valued functions, in Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry, ACM, 2009, 247–256.

[10]. Carlsson G, Zomorodian A, Collins A and Guibas L, Persistence barcodes for shapes, International J. Shape Modeling, 11 (2005), 149–187.

[11]. Cheeger J, A lower bound for the smallest eigenvalue of the Laplacian, in Problems in Analysis, Princeton Univ. Press, Princeton, NJ, 1970, 195–199.

[12]. Chen J, Zhao R, Tong Y and Wei G-W, Evolutionary de Rham-Hodge method, Discrete Contin. Dyn. Syst. Ser. B, (2020).

[13]. Chung FR, Spectral Graph Theory, CBMS Regional Conference Series in Mathematics, 92, American Mathematical Society, Providence, RI, 1997.

[14]. Ciocanel M-V, Juenemann R, Dawes AT and McKinley SA, Topological data analysis approaches to uncovering the timing of ring structure onset in filamentous networks, Bull. Math. Biol, 83 (2021), 21pp. [PubMed: 33452960]

[15]. de Silva V and Ghrist R, Coverage in sensor networks via persistent homology, Algebr. Geom. Topol, 7 (2007), 339–358.

[16]. Delaunay B, Sur la sphère vide, Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk, 7 (1934), 793–800.

[17]. Dey TK, Fan F and Wang Y, Computing topological persistence for simplicial maps, in Computational Geometry (SoCG'14), ACM, New York, 2014, 345–354.

[18]. Eckmann B, Harmonische funktionen und Randwertaufgaben in einem Komplex, Comment. Math. Helv, 17 (1945), 240–255.

[19]. Edelsbrunner H, Alpha shapes - A survey, Tessellations in the Sciences, 27 (2010), 1–25. Available from: https://pub.ist.ac.at/~edels/Papers/2011-B-03-AlphaShapes.pdf.

[20]. Edelsbrunner H and Harer J, Persistent homology - A survey, in Surveys on Discrete and Computational Geometry, Contemp. Math., 453, Amer. Math. Soc., Providence, RI, 2008, 257–282.

[21]. Fasy BT, Kim J, Lecci F, Maria C, Millman DL and Kim MJ, Package (TDA), 2019.

[22]. Friedman J, Computing Betti numbers via combinatorial Laplacians, Algorithmica, 21 (1998), 331–346.

[23]. Giusti C, Pastalkova E, Curto C and Itskov V, Clique topology reveals intrinsic geometric structure in neural correlations, Proc. Natl. Acad. Sci. USA, 112 (2015), 13455–13460. [PubMed: 26487684]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[24]. Hernández Serrano D, Hernaández-Serrano J and Sánchez Gómez D, Simplicial degree in complex networks. Applications of topological data analysis to network science, Chaos Solitons Fractals, 137 (2020), 21pp.

[25]. Kaczynski T, Mischaikow K and Mrozek M, Computational Homology, Applied Mathematical Sciences, 157, Springer-Verlag, New York, 2004.

[26]. Kamber FW and Tondeur P, De Rham-Hodge theory for Riemannian foliations, Math. Ann, 277 (1987), 415–431.

[27]. Kerber M and Edelsbrunner H, The medusa of spatial sorting: 3D kinetic alpha complexes and implementation, preprint, arXiv:1209.5434.

[28]. Lee Y, Barthel SD, D-lotko P, Mohamad Moosavi S, Hess K and Smit B, Quantifying similarity of pore-geometry in nanoporous materials, Nature Communications, 8 (2017).

[29]. Maroulas V, Micucci CP and Nasrin F, Bayesian topological learning for classifying the structure of biological networks, preprint, arXiv:2009.11974.

[30]. May J, Multivariate Analysis, Scientific e-Resources, 2018.

[31]. Mémoli F, Wan Z and Wang Y, Persistent Laplacians: Properties, algorithms and implications, preprint, arXiv:2012.02808.

[32]. Meng Z, Vijay Anand D, Lu Y, Wu J and Xia K, Weighted persistent homology for biomolecular data analysis, Scientific Reports, 10 (2020), 1–15. [PubMed: 31913322]

[33]. Meng Z and Xia K, Persistent spectral based machine learning (PerSpect ML) for drug design, preprint, arXiv:2002.00582.

[34]. Mischaikow K and Nanda V, Morse theory for filtrations and efficient computation of persistent homology, Discrete Comput. Geom, 50 (2013), 330–353.

[35]. Morozov D, Dionysus Software, 2012.

[36]. Morozov D and Skraba P, DioDe Software, 2017.

[37]. Nguyen D and Wei G-W, AGL-Score: Algebraic graph learning score for protein-ligand binding scoring, ranking, docking, and screening, J. Chemical Information Modeling, 59 (2019), 3291–3304.

[38]. Nguyen DD, Cang Z, Wu K, Wang M, Cao Y and Wei G-W, Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges, J. Comput. Aided Mol. Des, 33 (2019), 71–82. [PubMed: 30116918]

[39]. Gudhi Project, GUDHI User and Reference Manual, 2015.

[40]. Sgouralis I, Nebenfuhr A and Maroulas V, A Bayesian topological framework for the identification and reconstruction of subcellular motion, SIAM J. Imaging Sci, 10 (2017), 871–899.

[41]. Spielman DA, Spectral graph theory and its applications, 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), IEEE, 2007, 29–38.

[42]. Townsend J, Micucci CP, Hymel JH, Maroulas V and Vogiatzis KD, Representation of molecular structures with persistent homology for machine learning applications in chemistry, Nature Communications, 11 (2020).

[43]. Voronoi G, Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites, J. Reine Angew. Math, 133 (1908), 97–102.

[44]. Wang R, Nguyen DD and Wei G-W, Persistent spectral graph, Int. J. Numer. Methods Biomed. Eng, 36 (2020), 27pp.

[45]. Xia K, Opron K and Wei G-W, Multiscale Gaussian network model (mGNM) and multiscale anisotropic network model (mANM), J. Chem. Phys, 143 (2015).

[46]. Xia K and Wei G-W, Persistent homology analysis of protein structure, flexibility, and folding, Int. J. Numer. Methods Biomed. Eng, 30 (2014), 814–844.

[47]. Zhao R, Desbrun M, Wei G-W and Tong Y, 3D hodge decompositions of edge-and face-based vector fields. ACM Transactions on Graphics (TOG), 38 (2019), 1–13.

[48]. Zhao R, Wang M, Chen J, Tong Y and Wei G-W, The de Rham–Hodge analysis and modeling of biomolecules, Bull. Math. Biol, 82 (2020), 38pp. [PubMed: 32162119]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[49]. Zomorodian A and Carlsson G, Computing persistent homology, Discrete Comput. Geom, 33 (2005), 249–274.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

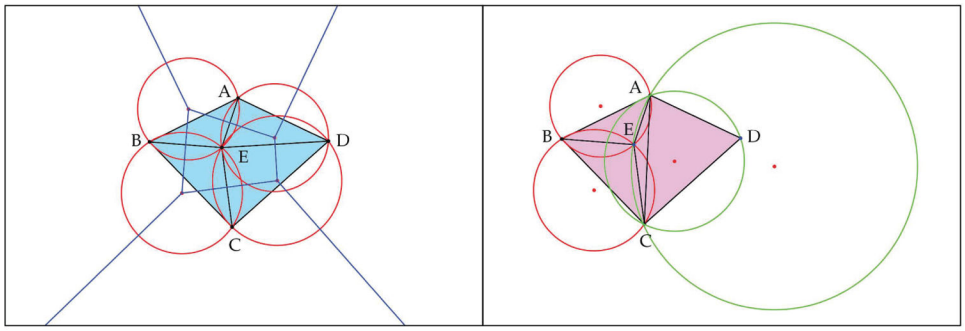

Figure 1.

Illustration of Voronoi diagram, Delaunay triangulation, and Non-Delaunay triangulation. **Left chart:** The Voronoi diagram and its dual Delaunay triangulation. The points set is $P = \{A,B,C,D,E\}$ and the Delaunay is defined as DT(P). The blue lines tessellate the plane into Voronoi cells. The red circle are the circumcircles of triangles in DT(P). **Right chart:** A Non-Delaunay triangulation. Vertices E and D are in the green circumcircles, implying the right chart is an example of Non-Delaunay triangulation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

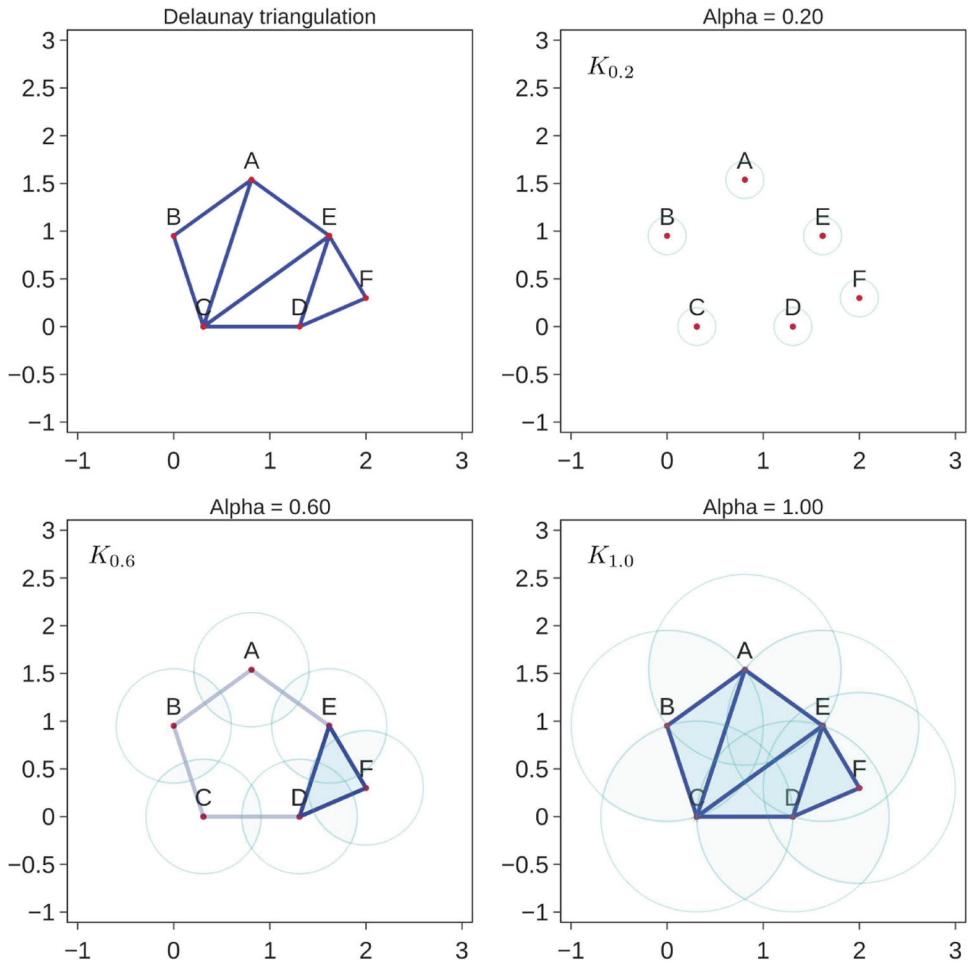

Figure 2.

Illustration of 2D Delaunay triangulation, alpha shapes, and alpha complexes for a set of 6 points A, B, C, D, E, and F. **Top left**: The 2D Delaunay triangulation. **Top right**: The alpha shape and alpha complex at filtration value α = 0.2. **Bottom left** : The alpha shape and alpha complex at filtration value α = 0.6. **Bottom right**: The alpha shape and alpha complex at filtration value α = 1.0. Here, we use dark blue color to fill the alpha shape.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

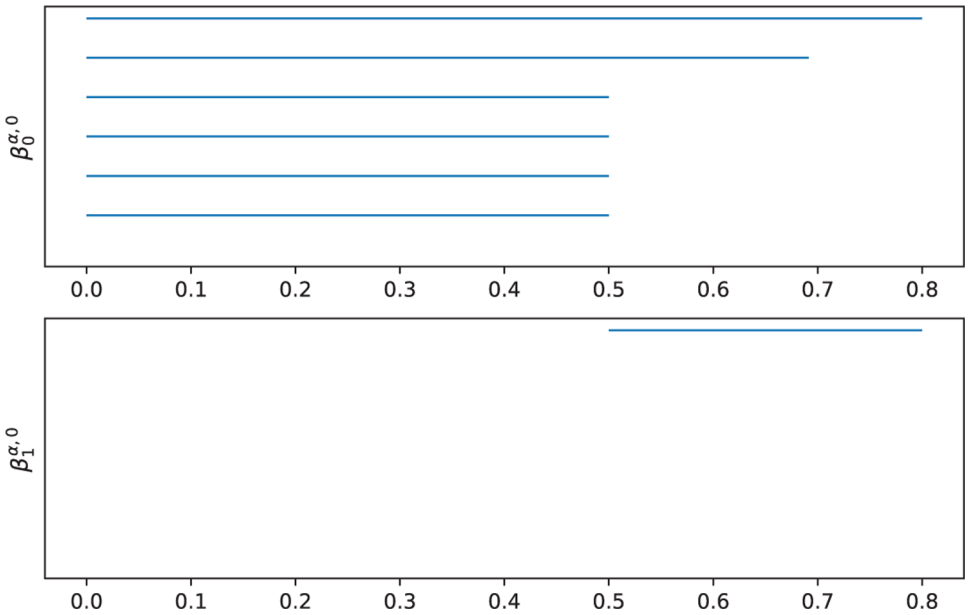

Figure 3.

The persistent barcode for a set of points as illustrated in Figure 2 that are generated from Gudhi and DioDe.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

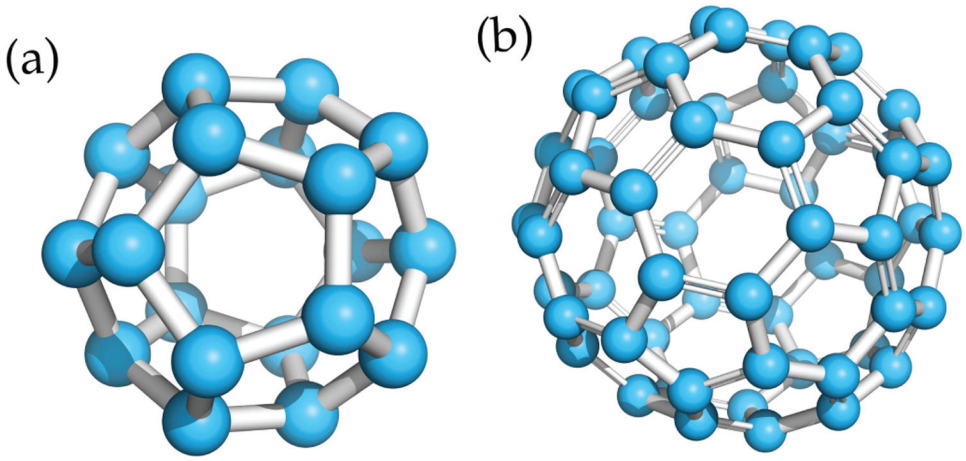

Figure 4.

The 3D structures of C_{20} and C_{60}. (a) C_{20} molecule. A total of 12 pentagon rings can be found in C_{20}. (b) C_{60} molecule. 12 pentagon rings and 20 hexoagon rings form the structure of C_{60}

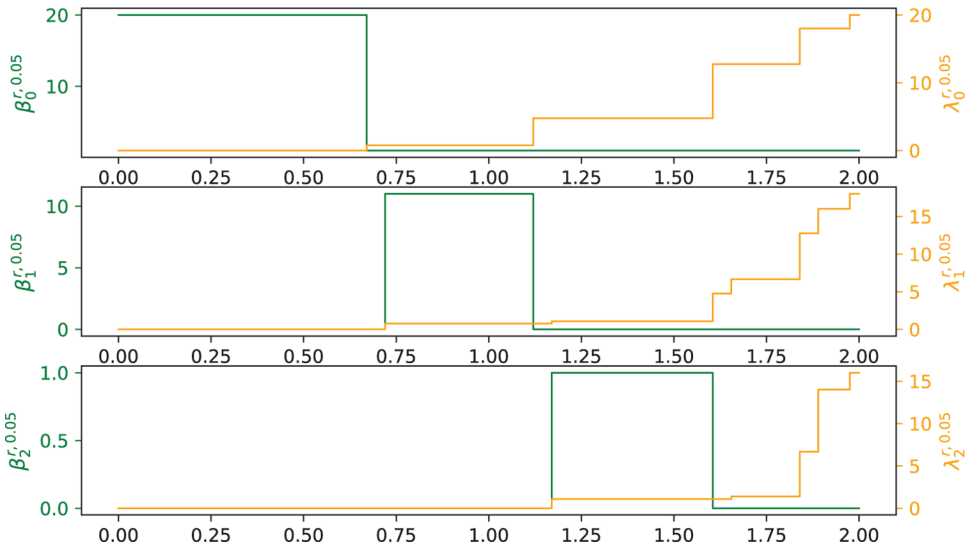

Figure 5.

Illustration of the harmonic spectra (for Rips complex) $\beta_0^{r,0.05}$, $\beta_0^{r,0.05}$, and $\beta_2^{r,0.05}$ (green curves from top chart to bottom chart) and the smallest non-zero eigenvalue $\lambda_0^{r,0.05}$, $\lambda_1^{r,0.05}$, and $\lambda_2^{r,0.05}$ (yellow curves from top chart to bottom chart) of C_{20} molecule (the bottom left chart in Fig. 9) at different filtration values α calculated from HERMES. Here, the x-axis represents the radius filtration value r (unit: Å), the left-y-axes represents the number of zero eigenvalues of $\mathscr{L}_0^{r,0.05}$, $\mathscr{L}_1^{r,0.05}$, and $\mathscr{L}_2^{r,0.05}$ from top to bottom, and the right-y-axes represents the first non-zero eigenvalue of $\mathscr{L}_0^{r,0.05}$, $\mathscr{L}_1^{r,0.05}$, and $\mathscr{L}_2^{r,0.05}$ from top to bottom.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

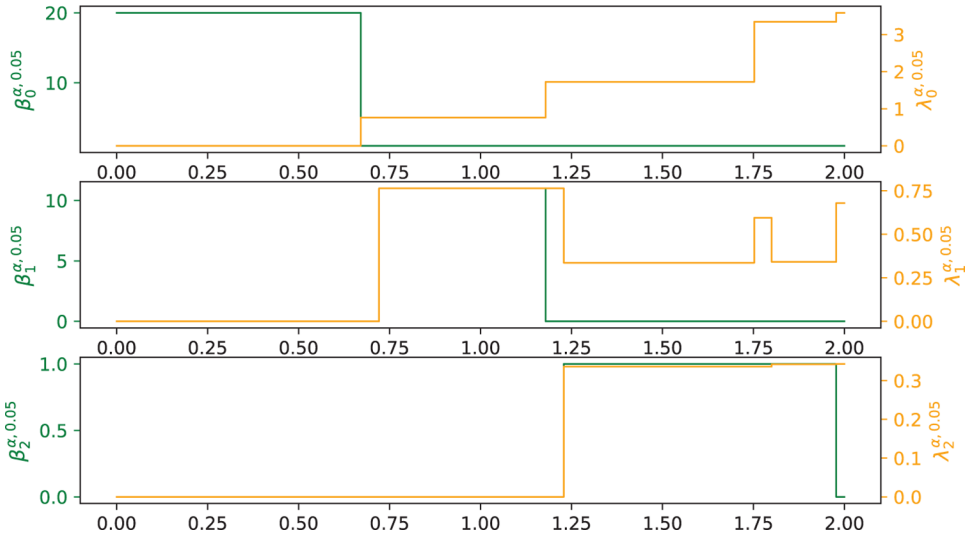

Figure 6.

Illustration of the harmonic spectra (for alpha complex) $\beta_0^{\alpha,0.05}$, $\beta_0^{\alpha,0.05}$, and $\beta_2^{\alpha,0.05}$ (green curves from top chart to bottom chart) and the smallest non-zero eigenvalue $\lambda_0^{\alpha,0.05}$, $\lambda_1^{\alpha,0.05}$, and $\lambda_2^{\alpha,0.05}$ (yellow curves from top chart to bottom chart) of C_{20} molecule (the bottom left chart in Fig. 9) at different filtration value α calculated from HERMES. Here, the x-axis represents the radius filtration value α (unit: Å), the left-y-axes represents the number of zero eigenvalues of $\mathscr{L}_0^{\alpha,0.05}$, $\mathscr{L}_1^{\alpha,0.05}$, and $\mathscr{L}_2^{\alpha,0.05}$ from top to bottom, and the right-y-axes represents the first non-zero eigenvalue of $\mathscr{L}_0^{\alpha,0.05}$, $\mathscr{L}_1^{\alpha,0.05}$, and $\mathscr{L}_2^{\alpha,0.05}$ from top to bottom.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

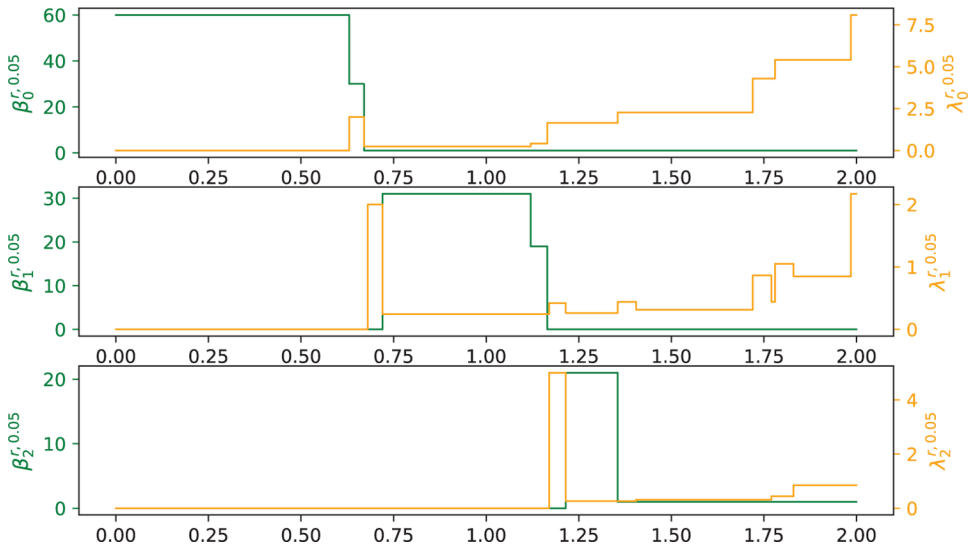

Figure 7.

Illustration of the harmonic spectra $\beta_0^{r,0.05}$, $\beta_0^{r,0.05}$, and $\beta_2^{r,0.05}$ (blue curves from top chart to bottom chart) and the smallest non-zero eigenvalue $\lambda_0^{r,0.05}$, $\lambda_1^{r,0.05}$, and $\lambda_2^{r,0.05}$ (red curves from top chart to bottom chart) of C_{60} molecule (the bottom left chart in Fig. 9) at different filtration value α calculated from HERMES. Here, the x-axis represents the radius filtration value α (unit: Å), the left-y-axes represents the number of zero eigenvalues of $\mathscr{L}_0^{r,0.05}$, $\mathscr{L}_1^{r,0.05}$, and $\mathscr{L}_2^{r,0.05}$ from top to bottom, and the right-y-axes represents the first non-zero eigenvalue of $\mathscr{L}_0^{r,0.05}$, $\mathscr{L}_1^{r,0.05}$, and $\mathscr{L}_2^{r,0.05}$ from top to bottom.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

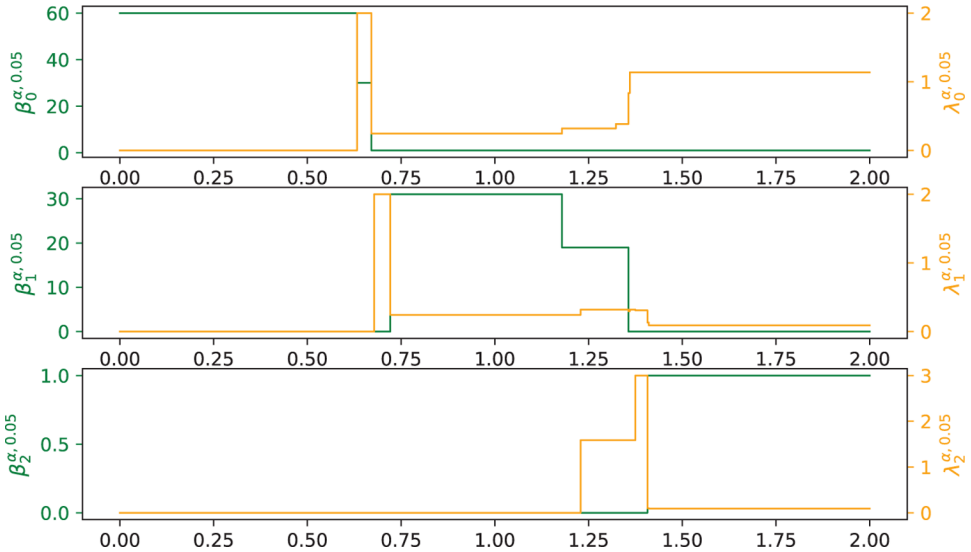

Figure 8.

Illustration of the harmonic spectra $\beta_0^{\alpha,0.05}$, $\beta_0^{\alpha,0.05}$, and $\beta_2^{\alpha,0.05}$ (green curves from top chart to bottom chart) and the smallest non-zero eigenvalue $\lambda_0^{\alpha,0.05}$, $\lambda_1^{\alpha,0.05}$, and $\lambda_2^{\alpha,0.05}$ (yellow curves from top chart to bottom chart) of C_{60} molecule (the bottom left chart in Fig. 9) at different filtration value α calculated from HERMES. Here, the *x*-axis represents the radius filtration value α (unit: Å), the left-*y*-axes represents the number of zero eigenvalues of $\mathscr{L}_0^{\alpha,0.05}$, $\mathscr{L}_1^{\alpha,0.05}$, and $\mathscr{L}_2^{\alpha,0.05}$ from top to bottom, and the right-*y*-axes represents the first non-zero eigenvalue of $\mathscr{L}_0^{\alpha,0.05}$, $\mathscr{L}_1^{\alpha,0.05}$, and $\mathscr{L}_2^{\alpha,0.05}$ from top to bottom.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

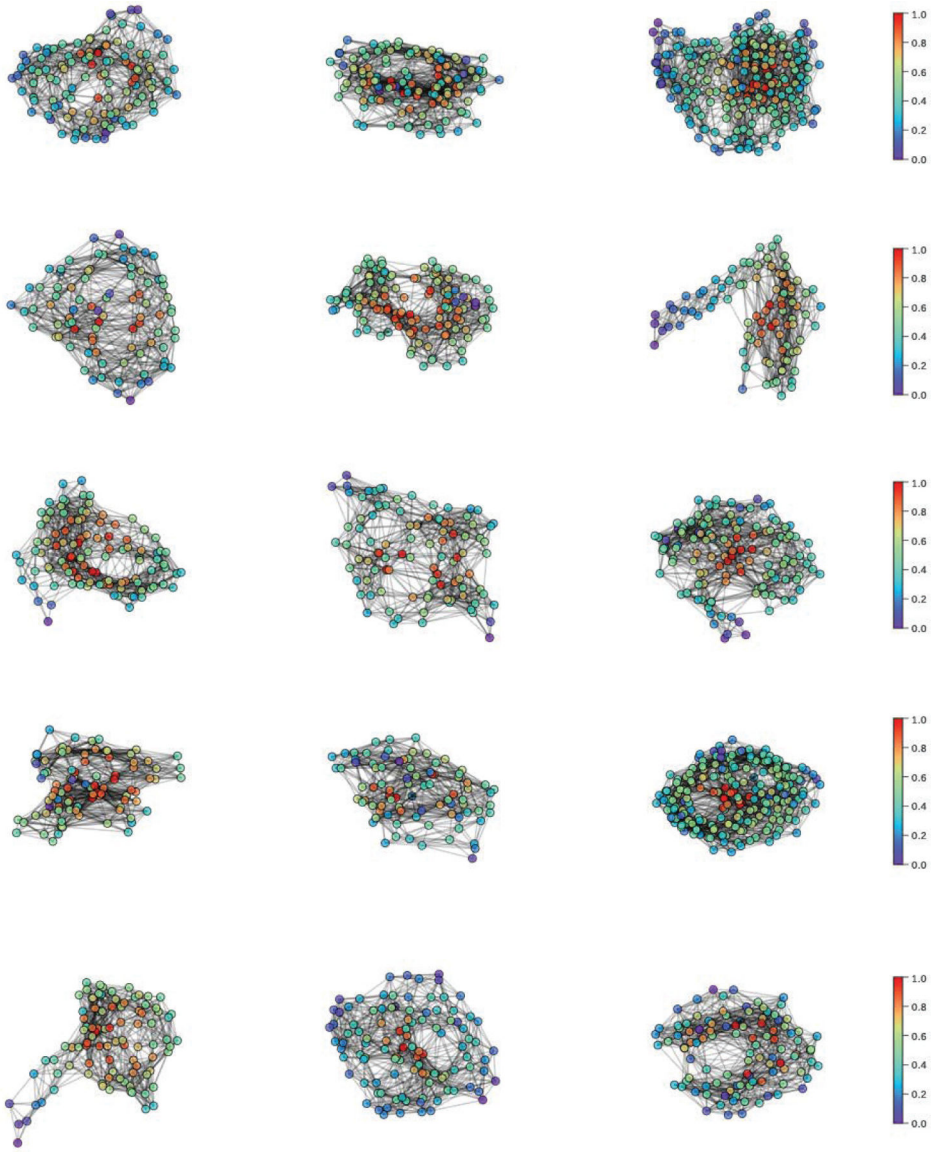

Figure 9.

The alpha carbon network plots of 15 proteins: PDB IDs 1CCR, 1NKO, 1O08, 1OPD, 1QTO, 1R7J, 1V70, 1W2L, 1WHI, 2CG7, 2FQ3, 2HQK, 2PKT, 2VIM, and 5CYT from left to right and top to bottom. The color represents the normalized diagonal element of the accumulated Laplacian at each alpha carbon atom.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

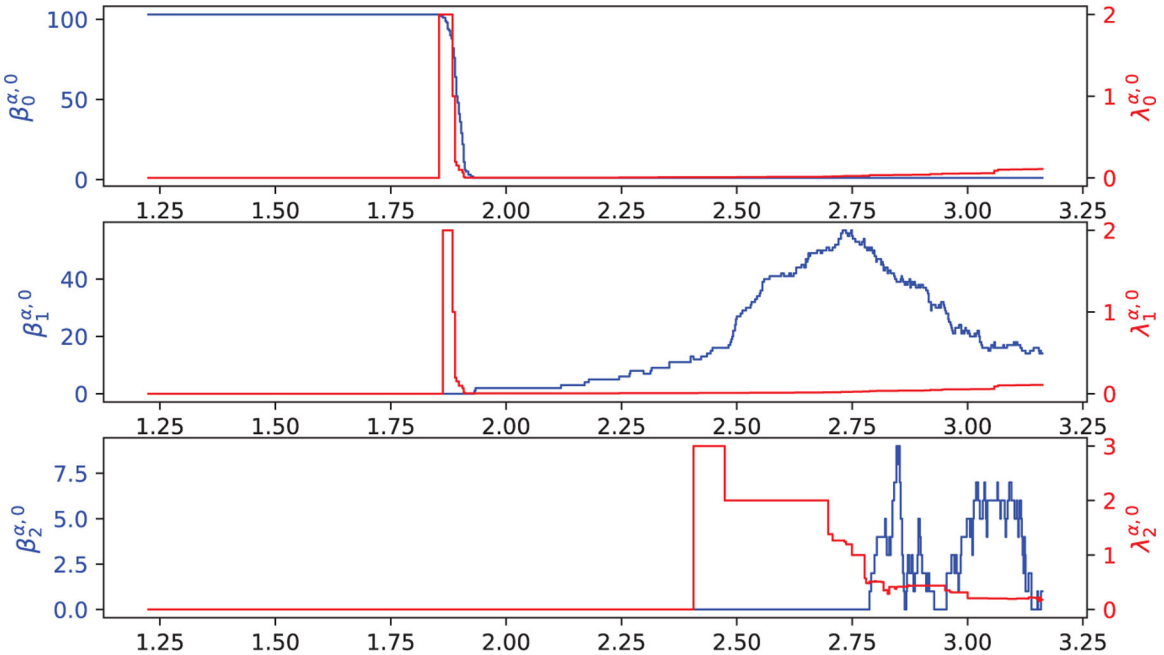

Figure 10.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 5CYT (the bottom left chart in Fig. 9) at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the x-axis represents the radius filtration value α (unit: Å), the left-y-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-y-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

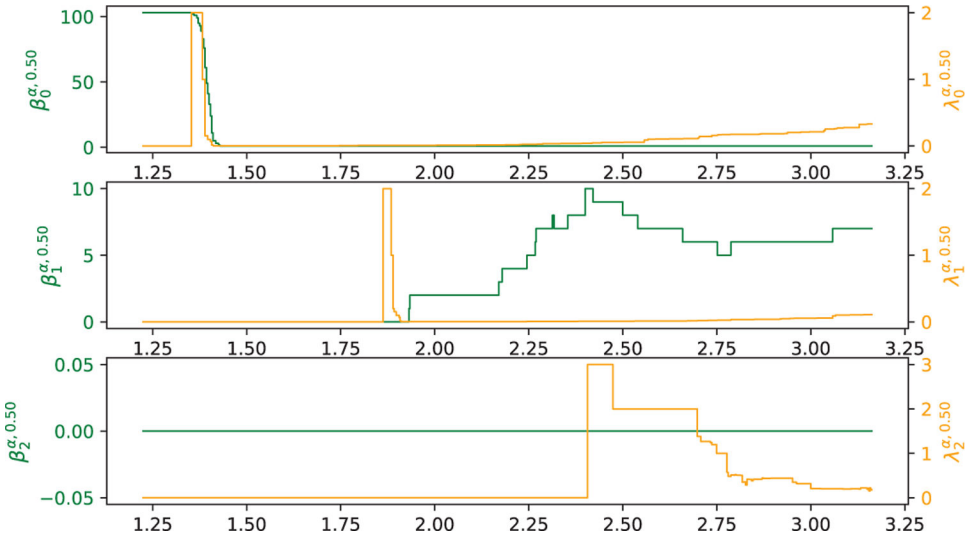

Figure 11.

Illustration of the harmonic spectra $\beta_0^{\alpha,0.5}$, $\beta_0^{\alpha,0.5}$, and $\beta_2^{\alpha,0.5}$ (green curves from top chart to bottom chart) and the smallest non-zero eigenvalue $\lambda_0^{\alpha,0.5}$, $\lambda_1^{\alpha,0.5}$, and $\lambda_2^{\alpha,0.5}$ (yellow curves from top chart to bottom chart) of PDB ID 5CYT (the bottom left chart in Fig. 9) at different filtration value α calculated from HERMES. Here, the *x*-axis represents the radius filtration value α (unit: Å), the left-*y*-axes represents the number of zero eigenvalues of $\mathscr{L}_0^{\alpha,0.5}$, $\mathscr{L}_1^{\alpha,0.5}$, and $\mathscr{L}_2^{\alpha,0.5}$ from top to bottom, and the right-*y*-axes represents the first non-zero eigenvalue of $\mathscr{L}_0^{\alpha,0.5}$, $\mathscr{L}_1^{\alpha,0.5}$, and $\mathscr{L}_2^{\alpha,0.5}$ from top to bottom.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

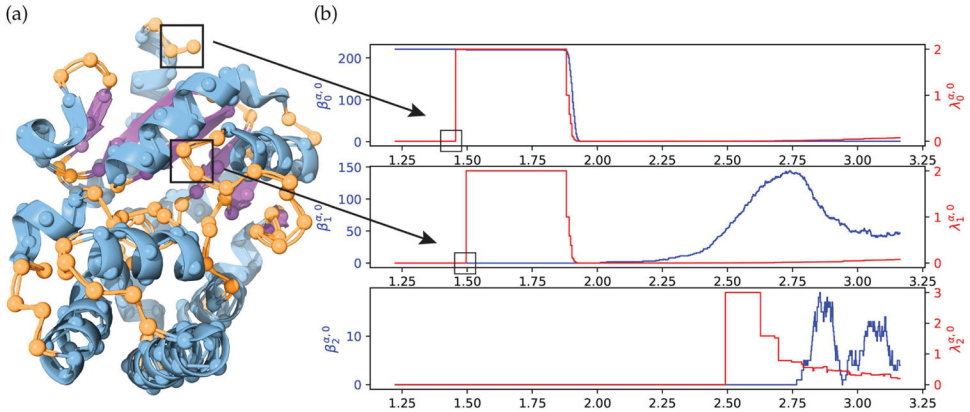

Figure 12.

(a) The 3D secondary structure of PDB ID 1O08. The blue, purple, and orange colors represent helix, sheet, and random coils of PDB ID 1O08. The ball represents the alpha carbon of PDB ID 1O08. (b) Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 1O08 at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are calculated only from HERMES. Here, the x-axis represents the radius filtration value α (unit: Å), the left-y-axis represents for the number of zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$, and the right-y-axis represents for the non-zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

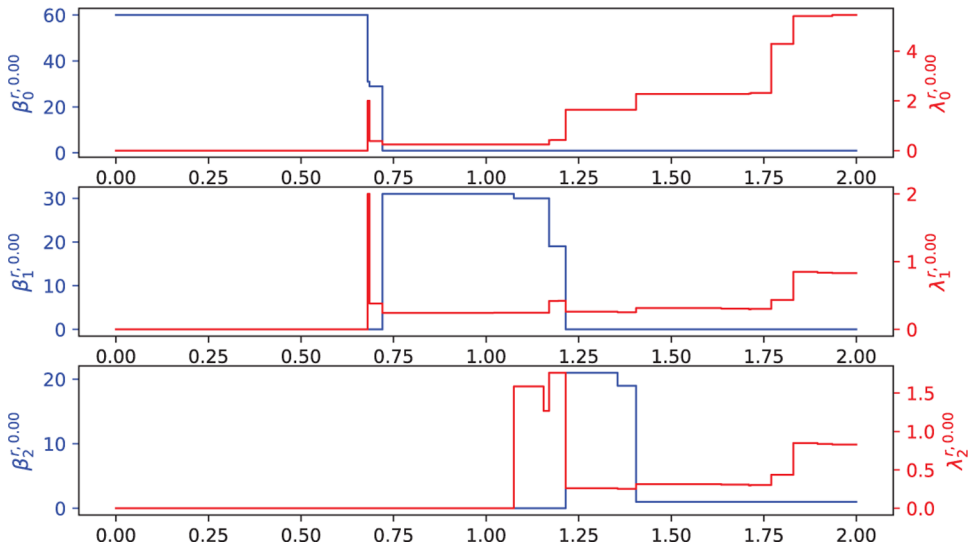

Figure 13.

Illustration of the harmonic spectra $\beta_0^{r,0}$, $\beta_0^{r,0}$, and $\beta_2^{r,0}$ (blue curves from top chart to bottom chart) and the smallest non-zero eigenvalue $\lambda_0^{r,0}$, $\lambda_1^{r,0}$, and $\lambda_2^{r,0}$ (red curves from top chart to bottom chart) of C_{60} molecule with one atom shifted (the bottom left chart in Fig. 9) at different filtration value a calculated from HERMES. Here, the *x*-axis represents the radius filtration value a (unit: Å), the left-*y*-axes represents the number of zero eigenvalues of $\mathscr{L}_0^{r,0}$, $\mathscr{L}_1^{r,0}$, and $\mathscr{L}_2^{r,0}$ from top to bottom, and the right-*y*-axes represents the first non-zero eigenvalue of $\mathscr{L}_0^{r,0}$, $\mathscr{L}_1^{r,0}$, and $\mathscr{L}_2^{r,0}$ from top to bottom.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

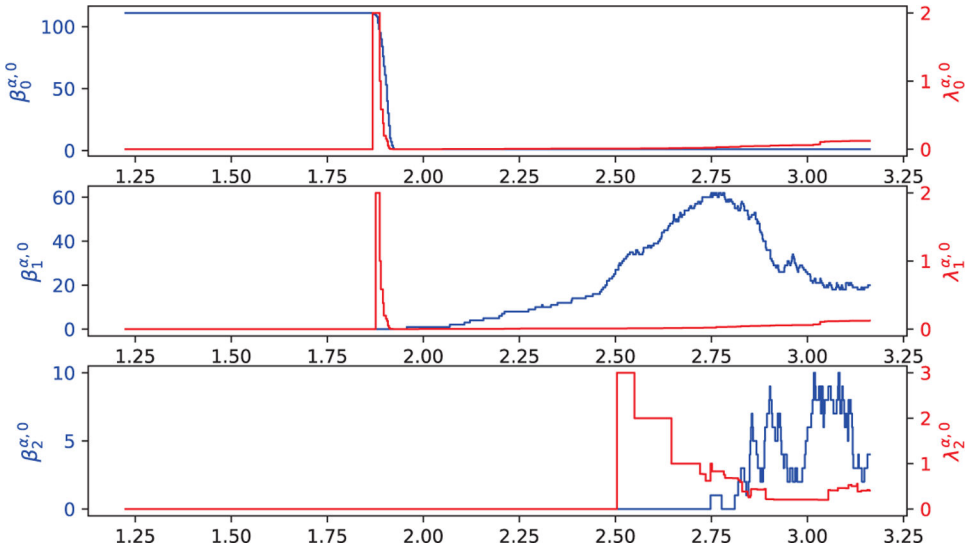

Figure 14.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 1CCR at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the *x*-axis represents the radius filtration value α (unit: Å), the left-*y*-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-*y*-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

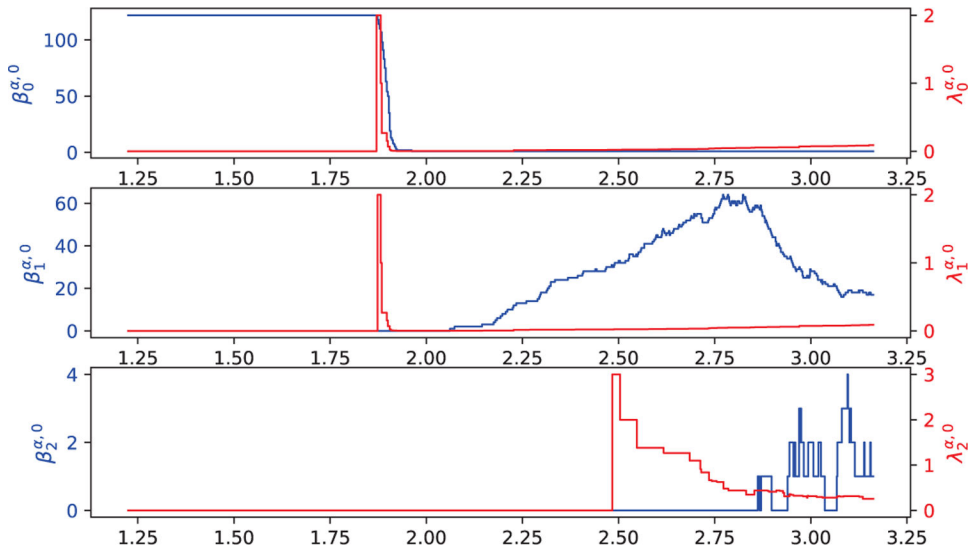

Figure 15.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 1NKO at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the *x*-axis represents the radius filtration value α (unit: Å), the left-*y*-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-*y*-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

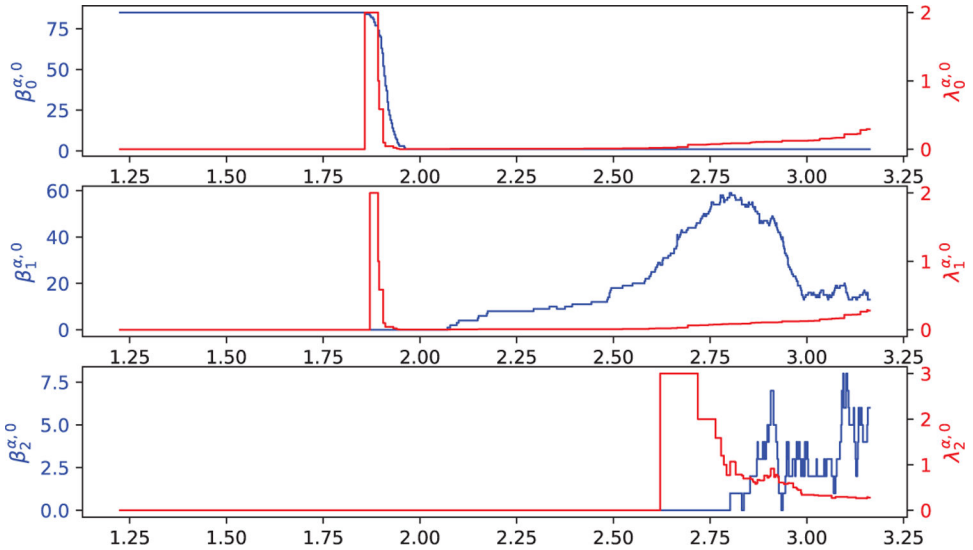

Figure 16.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 1OPD at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the *x*-axis represents the radius filtration value α (unit: Å), the left-*y*-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-*y*-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

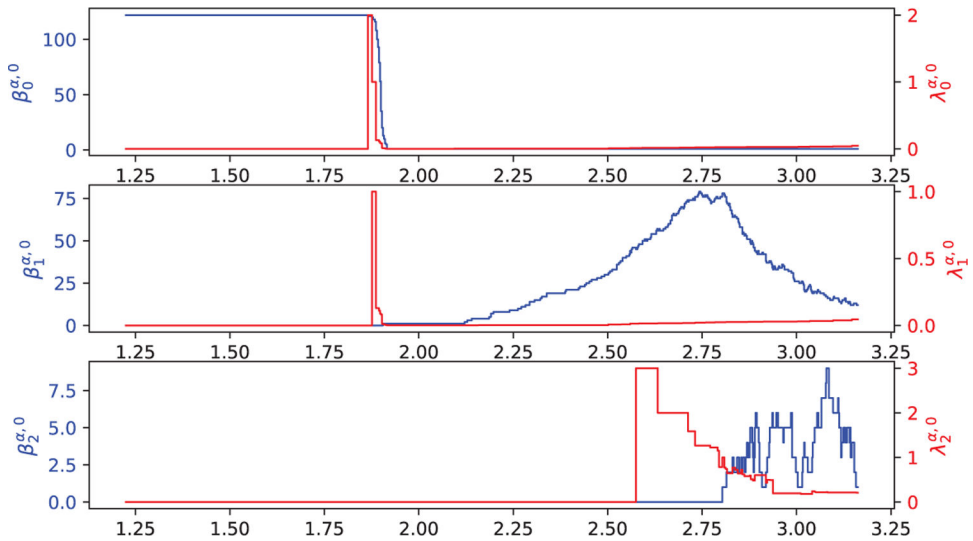

Figure 17.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 1QTO at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the *x*-axis represents the radius filtration value α (unit: Å), the left-*y*-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-*y*-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

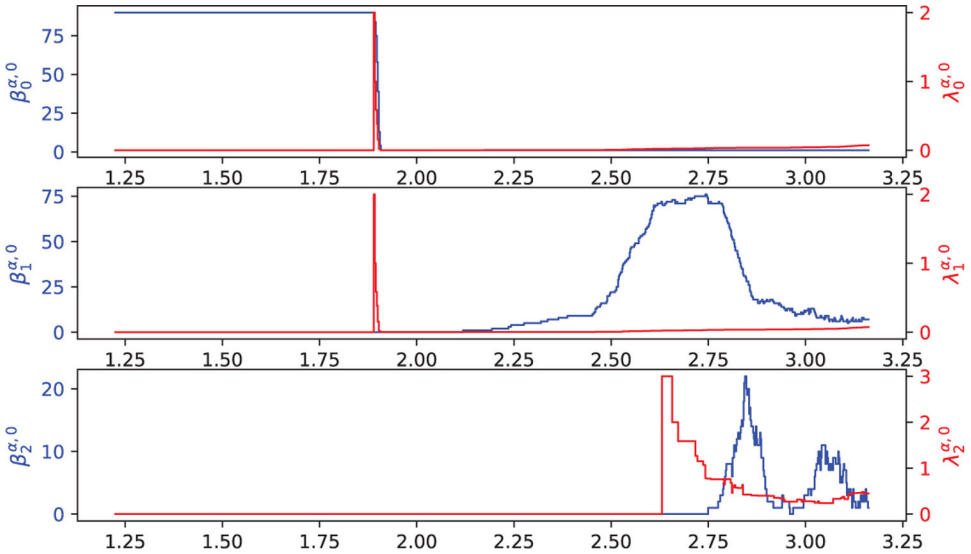

Figure 18.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 1R7J at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the *x*-axis represents the radius filtration value α (unit: Å), the left-*y*-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-*y*-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

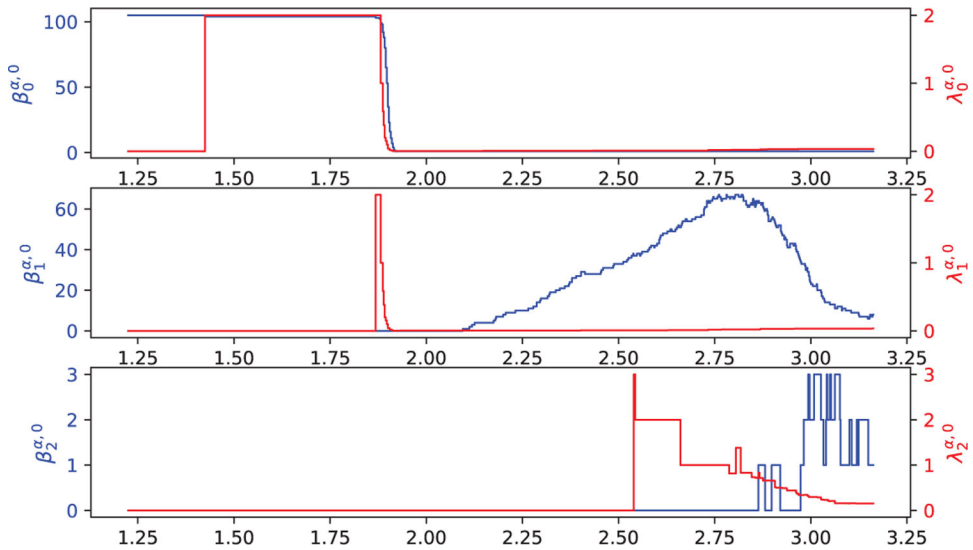

Figure 19.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 1V70 at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the x-axis represents the radius filtration value α (unit: Å), the left-y-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-y-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

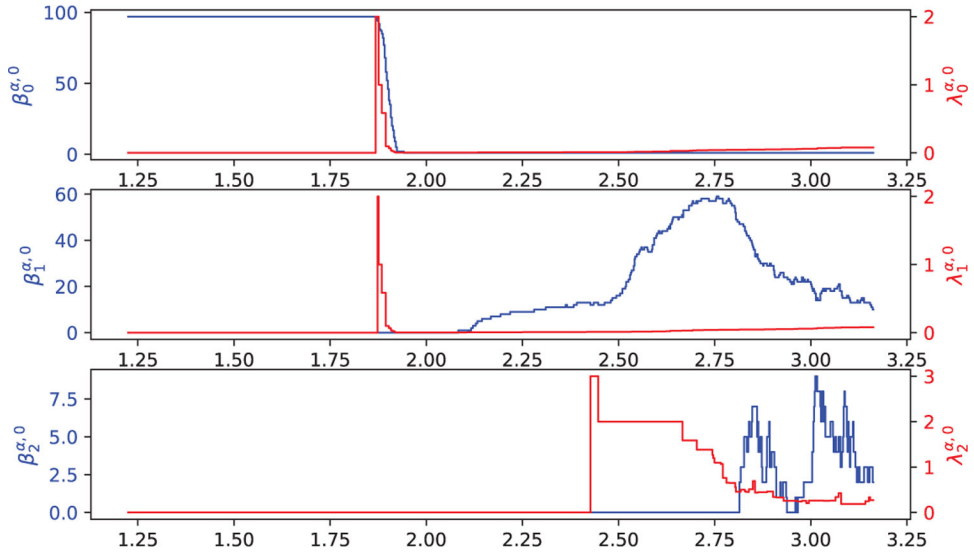

Figure 20.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 1W2L at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the x-axis represents the radius filtration value α (unit: Å), the left-y-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-y-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

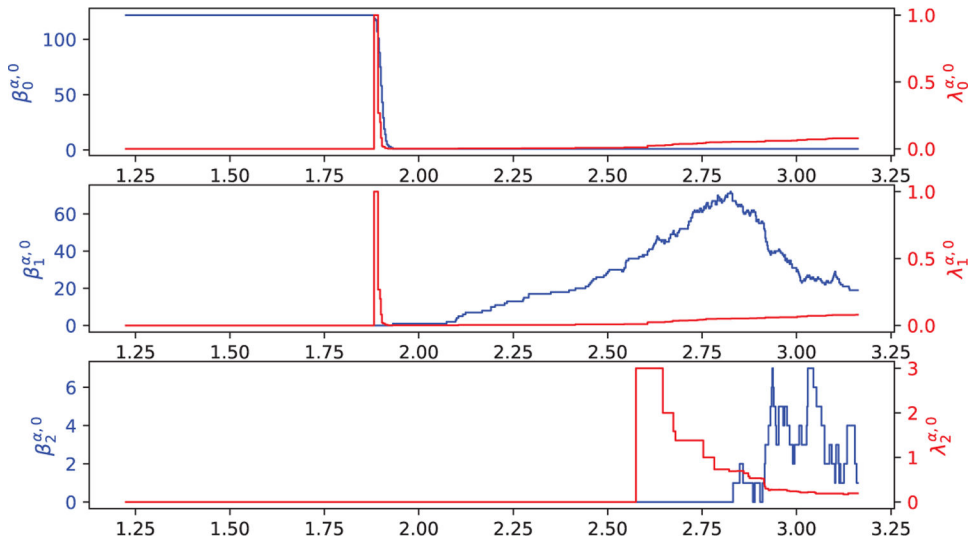

Figure 21.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 1WHI at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the x-axis represents the radius filtration value α (unit: Å), the left-y-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-y-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

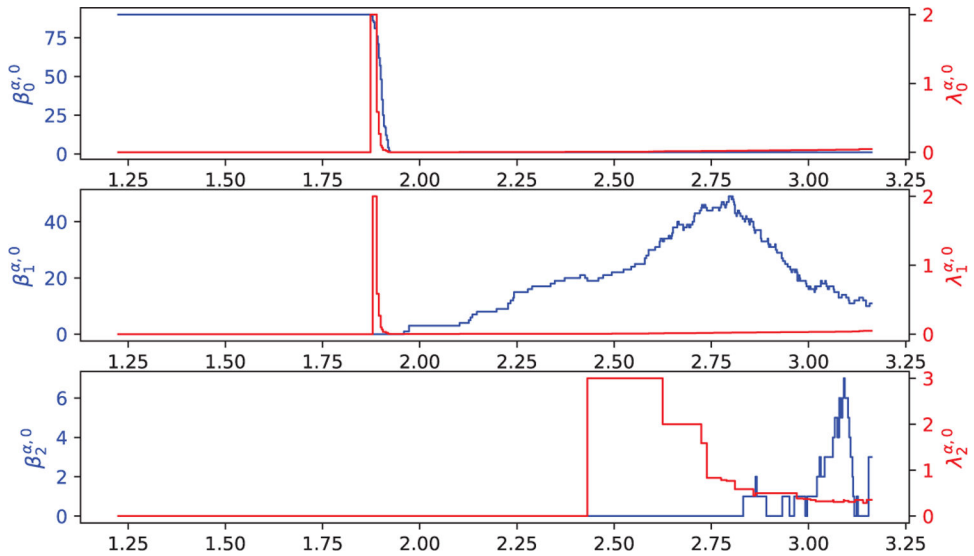

Figure 22.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 2CG7 at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the x-axis represents the radius filtration value α (unit: Å), the left-y-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-y-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

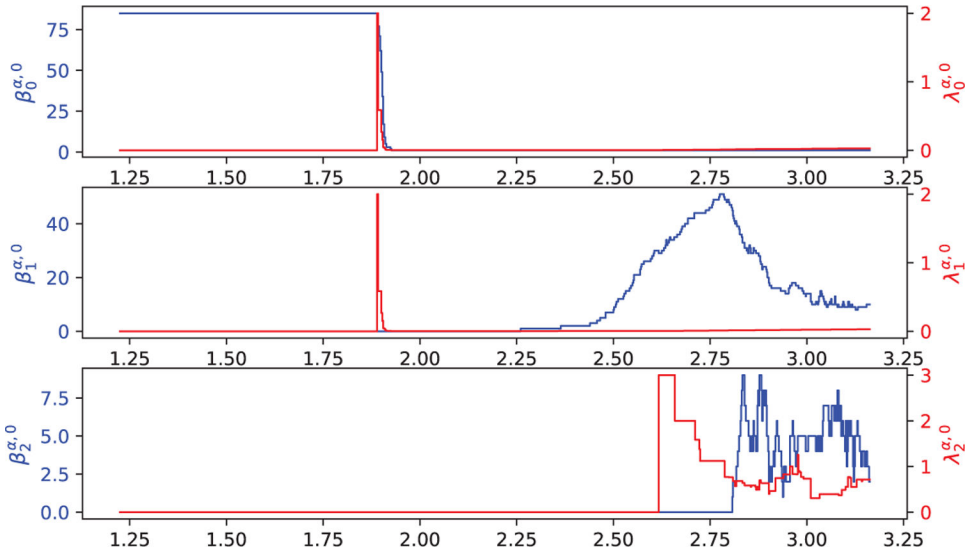

Figure 23.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 2FQ3 at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the *x*-axis represents the radius filtration value α (unit: Å), the left-*y*-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-*y*-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

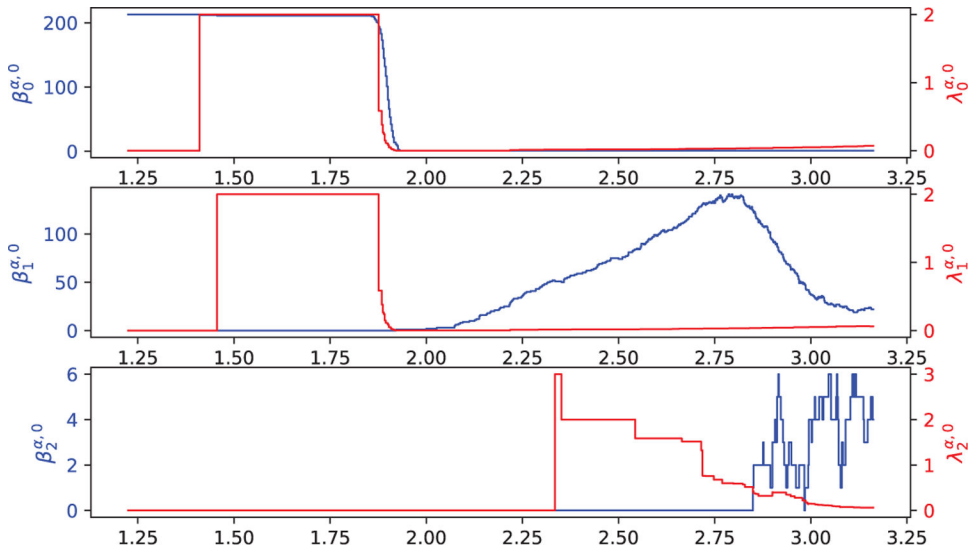

Figure 24.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 2HQK at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the x-axis represents the radius filtration value α (unit: Å), the left-y-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-y-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

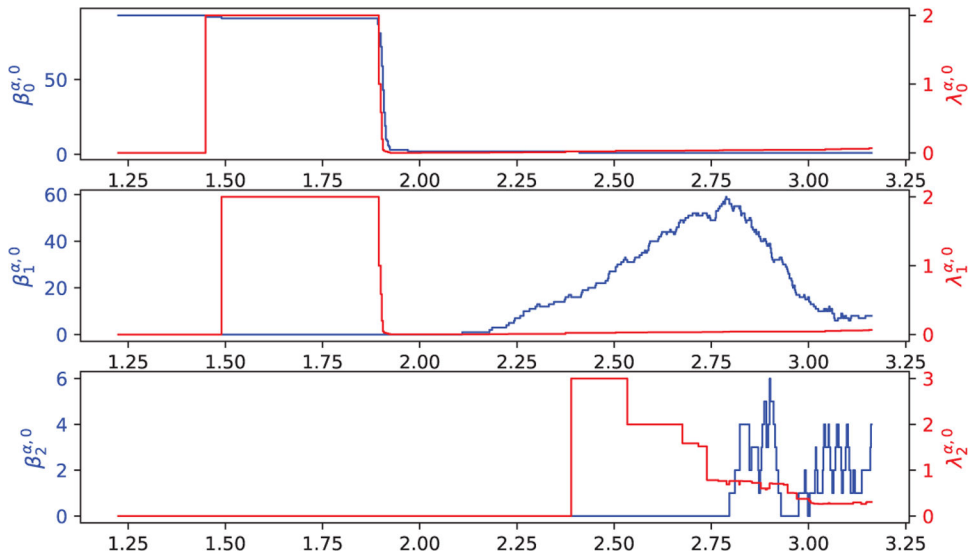

Figure 25.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 2PKT at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the x-axis represents the radius filtration value α (unit: Å), the left-y-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-y-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

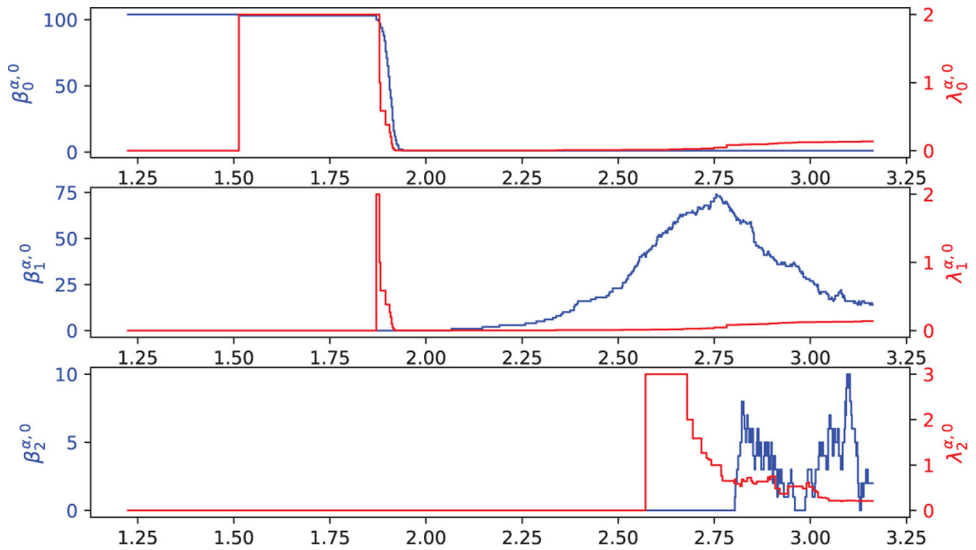

Figure 26.

Illustration of the harmonic spectra $\beta_q^{\alpha,0}$ (blue curve) and the smallest non-zero eigenvalue $\lambda_q^{\alpha,0}$ (red curve) of PDB ID 2VIM at different filtration value α when $q = 0, 1, 2$. The $\beta_q^{\alpha,0}$ are calculated from Gudhi, DioDe, and HERMES, and $\lambda_q^{\alpha,0}$ are obtained only from HERMES. Here, the *x*-axis represents the radius filtration value α (unit: Å), the left-*y*-axis represents the number of zero eigenvalues of $\mathscr{L}_q^{\alpha,0}$, and the right-*y*-axis represents the first non-zero eigenvalue of $\mathscr{L}_q^{\alpha,0}$. Note that the harmonic spectra from three methods are indistinguishable.

Author Manuscript

Table 1.

The matrix representation of q-boundary operator and its qth-order persistent Laplacian with corresponding dimension, rank, nullity, and spectra from alpha complex $K_{0.6} \rightarrow K_{0.6}$.

q	$q = 0$	$q = 1$	$q = 2$
$\mathscr{B}_{q+1}^{0.6,0}$	$\begin{array}{c c} & \begin{array}{ccccccc} AB & BC & CD & DE & EF & DF & AE \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \end{array} & \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array}$	$\begin{array}{c c} & DEF \\ \begin{array}{c} AB \\ BC \\ CD \\ DE \\ EF \\ DF \\ AE \end{array} & \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} \end{array}$	/
$\mathscr{B}_{q}^{0.6}$	$\begin{array}{c} A\ B\ C\ D\ E\ F \\ [0\ 0\ 0\ 0\ 0\ 0] \end{array}$	$\begin{array}{c c} & \begin{array}{ccccccc} AB & BC & CD & DE & EF & DF & AE \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \end{array} & \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array}$	$\begin{array}{c c} & DEF \\ \begin{array}{c} AB \\ BC \\ CD \\ DE \\ EF \\ DF \\ AE \end{array} & \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} \end{array}$
$\mathscr{L}_{q}^{0.6,0}$	$\begin{bmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & 0 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix}$	$\begin{bmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 1 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 3 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 3 & 0 & -1 \\ 0 & 0 & -1 & 0 & 0 & 3 & 0 \\ 1 & 0 & 0 & 1 & -1 & 0 & 2 \end{bmatrix}$	[3]
$\beta_{q}^{0.6,0}$	1	1	0
$\dim\left(\mathscr{L}_{q}^{0.6,0}\right)$	6	7	1
$\mathrm{rank}\left(\mathscr{L}_{q}^{0.6,0}\right)$	5	6	1
$\mathrm{nullity}\left(\mathscr{L}_{q}^{0.6,0}\right)$	1	1	0
$\mathrm{Spec}\left(\mathscr{L}_{q}^{0.6,0}\right)$	{0, 1, 1.5858, 3, 4, 4.4142}	{0, 1, 1.5858, 3, 3, 4, 4.4142}	{3}

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

The matrix representation of q-boundary operator and its qth-order persistent Laplacian with corresponding dimension, rank, nullity, and spectra from alpha complex $K_{0.2} \rightarrow K_{0.6}$.

q	$q = 0$	$q = 1$	$q = 2$
$\mathscr{B}_{q+1}^{0.2,0.4}$	$\begin{array}{c c} & \begin{array}{ccccccc} AB & BC & CD & DE & EF & DF & AE \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \end{array} & \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array}$	/	/
$\mathscr{B}_{q}^{0.2}$	$\begin{array}{c} A\ B\ C\ D\ E\ F \\ [0\ 0\ 0\ 0\ 0\ 0] \end{array}$	/	/
$\mathscr{L}_{q}^{0.2,0.4}$	$\begin{bmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & 0 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix}$	/	/
$\beta_{q}^{0.2,0.4}$	1	/	/
$\dim\left(\mathscr{L}_{q}^{0.2,0.4}\right)$	6	/	/
$\mathrm{rank}\left(\mathscr{L}_{q}^{0.2,0.4}\right)$	5	/	/
$\mathrm{nullity}\left(\mathscr{L}_{q}^{0.2,0.4}\right)$	1	/	/
$\mathrm{Spec}\left(\mathscr{L}_{q}^{0.2,0.4}\right)$	{0, 1, 1.5858, 3, 4, 4.4142}	/	/