

Topological Data Analysis and Topological Deep Learning Beyond Persistent Homology—A Review

Zhe Su¹, Xiang Liu², Layal Bou Hamdan³, Vasileios Maroulas³, Jie Wu⁴, Gunnar Carlsson⁵ and Guo-Wei Wei^{*2,6,7}

¹Department of Mathematics and Statistics, Auburn University,
AL 36849, USA.

²Department of Mathematics, Michigan State University, MI
48824, USA.

³Department of Mathematics, University of Tennessee,
Knoxville, TN 37996-1320, USA.

⁴Beijing Institute of Mathematical Sciences and Applications,
Beijing 101408, China.

⁵Department of Mathematics, Stanford University, CA 94305,
USA.

⁶Department of Biochemistry and Molecular Biology, Michigan
State University, MI 48824, USA.

⁷Department of Electrical and Computer Engineering, Michigan
State University, MI 48824, USA.

Abstract

Topological data analysis (TDA) is a rapidly evolving field in applied mathematics and data science that leverages tools from topology to uncover robust, shape-driven insights in complex datasets. The main workhorse is persistent homology, a technique rooted in algebraic topology. Paired with topological deep learning (TDL) or topological machine learning, persistent homology has achieved tremendous success in a wide variety of applications in science, engineering, medicine, and industry. However, persistent homology has many limitations due to its high-level abstraction, insensitivity to non-topological changes, and reliance on

¹*Corresponding author. Email: weig@msu.edu

point cloud data. This paper presents a comprehensive review of TDA and TDL beyond persistent homology. It analyzes how persistent topological Laplacians and Dirac operators provide spectral representations to capture both topological invariants and homotopic evolution. Other formulations are presented in terms of sheaf theory, Mayer topology, and interaction topology. For data on differentiable manifolds, techniques rooted in differential topology, such as persistent de Rham cohomology, persistent Hodge Laplacian, and Hodge decomposition, are reviewed. For one-dimensional (1D) curves embedded in 3-space, approaches from geometric topology are discussed, including multiscale Gauss-link integrals, persistent Jones polynomials, and persistent Khovanov homology. This paper further discusses the appropriate selection of topological tools for different input data, such as point clouds, sequential data, data on manifolds, curves embedded in 3-space, and data with additional non-geometric information. A review is also given of various topological representations, software packages, and machine learning vectorizations. Finally, this review ends with concluding remarks.

MSC Classification: 62R40 , 55-08 , 57R19 , 57K18

Contents

1	Introduction	6
2	Algebraic topology approaches	9
2.1	Persistent homology	10
2.1.1	Simplicial homology	10
2.1.2	Filtration and persistence	11
2.1.3	Other variations	12
	Multiparameter persistent homology	12
	Zig-zag persistent homology	13
	Parametrized homology	13
	Localized homology	13
	Evolutionary homology	13
	Multiresolution persistent homology	14
	Multidimensional persistence	14
	Weighted persistent homology	14
	Persistent path homology	14
	Persistent hypergraph homology	14
	Persistent super-hypergraph homology	15
	Graph-complex-based persistent homology	15
	Extended persistent homology	15
	Join persistent homology	15
	Persistent cohomology	15
	Cayley-persistence	16
	Morse theory	16
	Conley index	16
	Persistent equivariant cohomology	17
	Persistence over posets	17
2.1.4	Additional perspectives	17
2.2	Persistent combinatorial Laplacians	18
2.2.1	Combinatorial Laplacians	19
2.2.2	Persistent combinatorial Laplacians	20
2.2.3	Other variations	21
	Persistent path Laplacians	21
	Persistent hyper(di)graphy Laplacians	21
	Persistent hopping path-Laplacian	21
	Persistent directed flag Laplacian	21
	Persistent Laplacian for simplicial maps	22
	Multiscale Hochschild Laplacian	22
2.3	Persistent Dirac operator	22
2.3.1	Dirac operator	23
2.3.2	Persistent Dirac operator	23
2.3.3	Other variations	24
	Persistent path Dirac and hypergraph Dirac operators	25
	Persistent Mayer Dirac operators	25

4 CONTENTS

	Khovanov Dirac	25
2.4	Other formulations	25
2.4.1	Persistent sheaf theory	25
2.4.2	Persistent Mayer topology	26
2.4.3	Persistent interaction topology	27
3	Differential topology approaches	27
3.1	De Rham-Hodge theory for manifolds with boundary	29
3.1.1	De Rham cohomology	30
3.1.2	Hodge Laplacians	30
3.1.3	Hodge decomposition	31
3.2	Persistent de Rham cohomology	32
3.3	Persistent Hodge Laplacians	33
3.3.1	Lagrangian formulation	34
3.3.2	Eulerian formulation	34
4	Geometric topology approaches	35
4.1	Multiscale Gauss linking integral	36
4.2	Localized Jones polynomials	37
	Multiscale Jones polynomial	37
	Persistent Jones polynomial	37
4.3	Persistent Khovanov homology	38
4.4	Khovanov Laplacians and Khovanov Dirac	39
5	Machine learning featurization	40
5.1	Input data types and the choice of topological modeling	40
5.1.1	Point-cloud data	40
5.1.2	Distance matrix data	41
5.1.3	Network data	42
5.1.4	Data on manifolds	42
5.1.5	Sequential data	43
5.1.6	Knots, links, tangles, and braids	45
5.1.7	Data with additional non-geometric information	45
5.2	Approaches for topological representations	46
5.2.1	Representations of persistent homology/cohomology	46
	Persistence barcode/diagram	46
	Persistent Betti numbers/ Betti curves	47
	Persistence landscape	47
	Persistence image	48
	Persistent Betti number image	48
	Colored persistence barcode	49
	Graded persistence barcode	49
5.2.2	Representations of persistent Laplacians	49
	Eigenvalue representation	50
	Eigenvector representation	50
	Hodge decomposition representation	50

5.2.3	Representations of persistent Dirac operator	51
5.2.4	Representations of multiscale Gauss linking integral . .	51
5.2.5	Representations of multiscale/persistent Jones polynomial	52
5.2.6	Software and code resources	52
5.3	Featurization methods	52
5.3.1	Functional formulations	53
5.3.2	Counting in bins	53
5.3.3	Statistical property	54
5.3.4	Algebraic formulations	54
5.3.5	Multidimensional formulations	55
5.3.6	Kernel-based models	56
5.3.7	Statistical Frameworks for Persistent Homology	56
6	Concluding remarks	58

1 Introduction

Topological data analysis (TDA), rooted in topological techniques, is an emerging approach in applied mathematics and data science. TDA extracts topological information from data that cannot be obtained using statistical, physical, or other mathematical methods. Since topology is one of the most abstract fields in mathematics, TDA inherently reduces the dimensionality of high-dimensional data. Additionally, it simplifies geometric complexity and provides high-order data representations, extending conventional zeroth-order algorithms. Unlike classical topology or homology, TDA employs multiscale analysis, bridging topology and geometry through filtration.

A cornerstone of TDA is persistent homology [112, 333], an algebraic topology tool for analyzing point clouds or discrete data. Persistent homology constructs a family of topological spaces across spatial scales, capturing topological invariants of varying dimensions. For further details, see [44, 116]. Its standard representations, i.e., persistence barcodes [132] and persistence diagrams [112], encode the birth, death, and persistence of features (e.g., connected components, holes, cavities) during filtration. Initially used for qualitative analysis, persistent homology has evolved into a robust tool for quantitative analysis and machine learning [42, 313, 317]. Applications span image processing [17, 72], neuroscience [84], computational chemistry [291], biology [37, 38, 128, 178], nanomaterials [184, 317], crystalline materials [161], and complex networks [156].

Despite its utility, persistent homology has notable constraints [310]. It primarily captures topological invariants, often overlooking geometric shape evolution during filtration when no topological changes occur. Moreover, it struggles with localized topological information in data and cannot directly analyze data on differential manifolds or 1D curves embedded in 3-space (e.g., knots, links).

Efforts have been made to develop persistent topological Laplacians for various data structures, providing representations that can be integrated into machine learning models. For point cloud data and graphs, persistent combinatorial Laplacians were introduced by Wang et al. [302] and have been extensively studied in recent years [143, 195, 217, 303]. The harmonic spectra of persistent combinatorial Laplacians capture changes in topological invariants, which could recover the topological information captured by persistent homology, while the non-harmonic spectra encode geometric shape changes in the filtration process [217]. In addition to persistent combinatorial Laplacians, a variety of persistent topological Laplacians have been proposed for other topological domains. These include persistent sheaf Laplacians [310] for labeled point cloud data, persistent path Laplacians [300] for directed graphs and networks, persistent hypergraph Laplacians [202] and hyperdigraphy Laplacians [60] for graphs and networks with complex relationships, persistent directed flag Laplacian for flag complexes [163, 329]. Persistent topological Laplacians have been successfully applied to protein-ligand binding prediction [58, 219, 282], interactomic network modeling [110], gene expression

analysis [81], deep mutational scanning [66], and protein engineering [253]. In particular, one of the most impressive achievements of persistent topological Laplacians is the prediction of emerging dominant SARS-COV-2 variants BA4 and BA5 [65].

Another important spectral approach is quantum persistent homology or persistent Dirac, first introduced by Ameneiro et al. [9]. In quantum theory, the Dirac operator is for momentum, whereas the Laplacian is for kinetic energy, which reveals their relationship. Persistent Dirac operators have been studied in many topological domains with various applications [287, 308].

For data on (differential) manifolds, a natural approach is persistent de Rham cohomology, rooted in differential topology or differential geometry. However, persistent de Rham cohomology has rarely been implemented independently. In contrast, evolutionary de Rham-Hodge theory, which includes persistent de Rham cohomology, was first proposed by Chen et al. for data on differential manifolds in Lagrangian representation, i.e., tetrahedral meshes [64]. This approach draws on differential geometry or differential topology. Recently, persistent Hodge Laplacians have been developed in the Eulerian representation, i.e., Cartesian grid defined by level-set functions [282]. The similarity and difference between combinatorial Laplacians and Hodge Laplacians have been thoroughly discussed in [260]. However, the term “Hodge Laplacians” has also been used in the combinatorial setting, which can be somewhat misleading. Essentially, although the orthogonal decomposition can be carried out for both combinatorial Laplacians and Hodge Laplacians, only the Hodge decomposition of vector fields on differential manifolds can give rise to curl-free, divergence-free, and harmonic components, i.e., the so-called Helmholtz-Hodge decomposition [181, 324]. Persistent Sheaf theory [309, 319] and the persistent interaction topology [193] can effectively describe the local topology in the data. For manifolds with boundary, a more refined topology-preserving 5-component Hodge decomposition was developed to further split the harmonic component into three subcomponents under certain boundary conditions [248, 249, 281, 283, 324]. Hodge decomposition is an important tool in analyzing vector fields by extracting different dynamic features, and has many applications in computational fluid dynamics [318, 320], geometric modeling [304], and spectral data analysis [171]. Recently, it has been applied to the study of single-cell RNA velocities [281] and has been integrated with deep learning models for medical image analysis [207].

For data on 1D curves embedded in 3-space, natural approaches stem from low-dimensional topology. Multiscale Gauss linking integral method was proposed to deal with knot-like data [270]. Additionally, multiscale Jones polynomial and persistent Jones polynomial have been introduced [277]. Recently, persistent Khovanov homology has been developed [194, 268]. While Khovanov homology bridges geometric topology and algebraic topology, persistent Khovanov homology extends TDA to 1D data embedded in 3-space.

Cang and Wei introduced the first integration of persistent homology and deep neural networks, including topological deep learning (TDL), in

2017 [37]. By leveraging the high-level abstraction of topology, TDL dramatically reduces dimensionality, simplifies geometric complexity, captures high-order interactions, and offers an interpretable learning framework. Persistent homology-based machine learning models have been successfully applied in various fields in science, engineering, medicine, industry, defense, etc. [151, 243]. However, topology inherently simplifies data, which implies the irreversible loss of certain information. Therefore, competitive performance from TDL mostly involves intrinsically complex data, such as those found in biological sciences [40]. Perhaps some of the most compelling examples of applications in which TDL consistently demonstrates its advantages over other competing methods are the victories of TDL in the D3R Grand Challenges [233, 234], the discovery of SARS-CoV-2 evolution mechanisms [63, 301], and the successful forecasting of SARS-CoV-2 variants BA.2 [62] and BA.4/BA.5 [65] about two months in advance.

Machine learning typically requires all samples to have a uniform input vector. Therefore, featurization or vectorization of TDA information is often needed to convert topological representations, such as persistence barcodes, diagrams, images, landscapes, Betti curves, etc., into fixed-length feature vectors. Many attempts have been made to define kernels [53, 180, 183, 255]. Others focus on developing featurization techniques, such as statistical summaries [6, 14, 15, 37, 40, 42, 167, 330, 331], persistence bins [37], topological images [40], and Wasserstein metrics [40].

Numerous reviews have been conducted on persistent homology and its featurization [8, 21, 70, 77, 116, 250], as well as a dedicated survey on persistent topological Laplacians [310]. However, there is no review of the generalizations of persistent homology and their vectorization. The aim of this paper is to present a high-level overview of various TDA approaches beyond persistent homology and their machine learning featurization methods, without too many technical details or formulas. This work is intended to be accessible to new researchers in the field and to provide a broad perspective on various techniques and their applications.

The rest of the review is organized as follows. Section 2 is devoted to algebraic topology approaches and techniques defined on point-cloud data. These approaches include persistent homology and its generalizations to various topological domains, persistent topological Laplacians on various topological domains, persistent Dirac operators, persistent sheaf theory, persistent Mayer topology, and persistent interactions. Section 3 reviews methods that originate in differential topology or differential geometry. We discuss the de Rham-Hodge theory for data on manifolds with boundary, a commonly used setting in various applications. Persistent de Rham-Hodge theory, including persistent de Rham cohomology and persistent Hodge Laplacian, is discussed. Section 4 is devoted to various geometric topology-rooted TDA methods, such as multiscale Gauss linking integrals, persistent Jones polynomial, persistent Khovanov homology, and Khovanov Laplacian. Vectorization techniques are reviewed in Section 5. We discuss commonly used methods, such as statistical

summaries, Betti curves, persistence landscapes, and persistence images. We also provide relevant featurization techniques applicable to spectral methods, including eigenvalues, eigenvectors, and Hodge decompositions. In addition, we discuss the choice of TDA methods for effective analysis of different types of datasets. Finally, in Section 6, we provide concluding remarks.

2 Algebraic topology approaches

Topological data analysis for point cloud data has grown rapidly in recent years, providing a range of methods built upon tools in algebraic topology. These methods offer multiscale characterizations of point cloud data by extracting topological and geometrical features through a filtration process. The resulting features provide rich information on the underlying data and have been successfully integrated into many downstream analyses, including classification, clustering, and other machine learning tasks.

The most fundamental method, persistent homology [112, 333], captures the persistence of topological features, such as connected components, holes, and cavities, on different scales in a filtration process. However, it has several limitations as it focuses mostly on the topology without considering the geometric evolution in the filtration or additional structures associated with data points, and it can be computationally expensive for large datasets, see Section 2.1.4.

Recently developed methods, such as persistent combinatorial Laplacians [302] with their variations [60, 163, 202, 300, 329], and the persistent Dirac operator [9, 10], along with their variations [286, 287], generalize persistent homology and address some of its limitations. Persistent combinatorial Laplacians [302] integrate the multiscale analysis with combinatorial Laplacians, allowing for the analysis of both topological and geometric or combinatorial structures across different scales. Their harmonic spectra fully recover the topological information from persistent homology, while the non-harmonic spectra offer additional geometric or combinatorial features of the data. The persistent Dirac [9, 10], whose square yields the persistent combinatorial Laplacians, provides efficient algorithms based on quantum computing with exponential speed-up compared to classical methods in topological data analysis. There are also other formulations based on alternative homology theories, such as the persistent sheaf Laplacians [309], which further extend the persistent combinatorial Laplacians, allowing the analysis of labeled or weighted datasets; the Mayer topology [269, 287]; and the persistent interaction topology [193, 194]. These tools offer new perspectives for capturing topological features that are not accessible through standard homology.

In the following, we first review persistent homology with its variations and limitations, then turn to the persistent combinatorial Laplacians and their various variations for analyzing point cloud data with additional structures. We

then describe the persistent Dirac operator and its variations. Finally, we discuss other formulations, including the persistent sheaf topology, the persistent Mayer topology, and the persistent interaction topology.

2.1 Persistent homology

Persistent homology is a multiscale extension of simplicial homology, which is a tool in algebraic topology that classifies simplicial complexes based on their homology groups. The homology groups measure the topological features in the underlying simplicial complex, such as connected components, holes, and cavities. Their ranks, given by Betti numbers, count the number of these topological features. For example, the 0th Betti number provides the number of connected components, the 1st Betti number gives the number of circles or holes, and the 2nd Betti number represents the number of cavities.

However, simplicial homology does not directly apply to point-cloud data, as such data do not have a well-defined topological structure unless explicitly constructed. In addition, it is sensitive to scales, as different choices of scales can lead to different homology groups. Persistent homology addresses these limitations by introducing filtration and persistence, providing a more robust characterization of the underlying data.

2.1.1 Simplicial homology

The simplicial complexes are made of simplices, where a k -simplex is defined to be the convex hull of $k+1$ affinely independent points v_0, v_1, \dots, v_k , and has dimension k . Examples include a point (0-simplex), a line segment (1-simplex), a triangle (2-simplex), and a tetrahedron (3-simplex). These points v_0, v_1, \dots, v_k are called the vertices of the k -simplex, and the simplices formed by a subset of the set of all these vertices are called faces.

A simplicial complex K is then defined to be a finite set of simplices satisfying the gluing conditions, where all faces of a simplex are also in K , and the intersection of any two simplices in K is either empty or a common face of both. The dimension of K is given by the largest dimension of its simplices. A linear combination of all k -simplices of K with coefficients in a field (commonly \mathbb{Z}_2) defines a k -chain. The set of all k -chains, together with addition, forms an abelian group $\mathcal{C}_k(K)$, called the k -th chain group. The set of all k -simplices of K serves as a basis of $\mathcal{C}_k(K)$.

To define the boundary operator ∂ , each simplex of K needs to be oriented (not necessary when using \mathbb{Z}_2 coefficients as $-1 = +1$). An oriented k -simplex is represented as an ordered list of its vertices, denoted by $[v_0, v_1, \dots, v_k]$. The orientation is crucial for constructing the chain complex and defining homology, as it ensures that $\partial\partial = 0$. The k -th boundary operator $\partial_k : \mathcal{C}_k(K) \rightarrow \mathcal{C}_{k-1}(K)$ is a linear map that acts on each k -simplex by mapping it to an alternating sum of its $(k-1)$ -faces, where each is obtained by removing one vertex from

the k -simplex, i.e.,

$$\partial_k[v_0, v_1, \dots, v_k] = \sum_i (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_k]$$

The boundary operator satisfies $\partial_k \partial_{k+1} = 0$. A chain complex is given by a sequence of chain groups connected by the boundary operator:

$$\dots \xrightarrow{\partial_{k+2}} \mathcal{C}_{k+1} \xrightarrow{\partial_{k+1}} \mathcal{C}_k \xrightarrow{\partial_k} \mathcal{C}_{k-1} \xrightarrow{\partial_{k-1}} \dots$$

The k -th simplicial homology group of K is defined as the quotient of the kernel of the boundary operator ∂_k modulo the image of ∂_{k+1} :

$$H_k(K) = \ker \partial_k / \text{im } \partial_{k+1},$$

where $\ker \partial_k$ and $\text{im } \partial_{k+1}$ are called the group of k -cycles and the group of k -boundaries, respectively. The homology group $H_k(K)$ captures the k -cycles that are not k -boundaries, i.e., the k -dimensional holes in the simplicial complex. Its rank, known as the k -th Betti number β_k , counts the number of these holes.

2.1.2 Filtration and persistence

A filtration is given by a nested sequence of simplicial complexes. For point cloud data, one needs to construct such a filtration by building simplicial complexes to represent data points and their connections in various dimensions. The three most commonly used complexes are the Vietoris-Rips complex [295], the Čech complex [114], and the Alpha complex [113]. There are also other constructions such as witness complex [88], Cover complex [51], and Tangential complex [159].

The Vietoris-Rips complex [295] is constructed by considering pairwise intersections of balls centered at each data point within a fixed distance, leading to a simplex whenever its vertices have pairwise overlapping regions. The Čech complex [114] requires a nonempty intersection of all balls in a set, meaning that a simplex is included only if all corresponding balls intersect. The Alpha complex can be constructed in a similar manner to the Čech complex. However, each ball is restricted to the Voronoi cell associated with each point. The Alpha complex [113] is a subset of the Čech complex but has the same homology. By varying the radius of the balls, one obtains a filtration of simplicial complexes. The Čech complex involves checking for a large number of intersections and thus comes with significantly higher computational costs for large datasets, making the Vietoris-Rips complex the most common choice in real applications. An alternative to the Čech complex is the Alpha complex, which is particularly computationally more efficient when the ambient space has dimension 2 or 3 [114].

Let K be the largest simplicial complex constructed from a point cloud. Then a filtration of this point cloud is a nested sequence of the subcomplexes of K :

$$\emptyset \subset K^0 \subset K^1 \cdots \subset K^m = K.$$

Denote by $i^{i,j}$ the inclusion maps from K_i to K_j with $i < j$. Its induced map $f_k^{i,j} : H_k(K^i) \rightarrow H_k(K^j)$ on the homology groups tracks how the k -dimensional holes evolve across the filtration, more precisely, if the k -dimensional homology classes in $H_k(K^i)$ still exist in $H_k(K^j)$. Denote by $\partial_k^i : C_k(K^i) \rightarrow C_{k-1}(K^i)$ the k -th boundary operator on simplicial complex K^i . The k -th persistent homology group is defined to be the image of $f_k^{i,j}$:

$$H_k^{i,j} = \text{im } f_k^{i,j} = \ker \partial_k^i / \left(\text{im } \partial_{k+1}^j \cap \ker \partial_k^i \right),$$

and its rank leads to the corresponding k -th persistent Betti number $\beta_k^{i,j} = \text{rank } H_k^{i,j}$. The birth and death of a topological feature can then be defined by examining the elements in the persistent homology groups. A feature is said to be born at K^i if it first appears at K^i , and it is said to die at K^j if it merges with a feature at K^j that already exists before K^j . The persistence of γ , denoting the lifetime of a topological feature, is then given by $j - i$, and considered infinite if it never dies in the filtration.

These birth and death information of topological features gives rise to various representations of persistent homology, such as persistence barcodes [132], persistence diagrams [112], persistent Betti numbers, Betti curves, persistence landscapes [34], persistence images [4]. More details about these featurization techniques will be discussed in Section 5.

Persistent homology has been successfully applied in various fields, ranging from image processing to complex networks [17, 37, 38, 72, 84, 128, 156, 161, 178, 184, 291]. The most significant achievements would be its integration with deep learning models, which has led to exceptional performance in the D3R grand challenges [233, 234] and in predicting the evolutionary mechanism of SARS-CoV-2 [63, 301].

2.1.3 Other variations

In addition to the standard formulation of persistent homology for simplicial complexes, several variations have been proposed for analyzing data with different structures or for addressing specific application needs. We briefly review these methods below.

Multiparameter persistent homology

The multiparameter persistent homology was first explored in [46], while a variation in size theory was introduced in [127]. The framework extends classical homology by considering two or more parameters instead of one, and it is

suitable for analyzing data with outliers or variations in density, data equipped with real-valued functions and functional data with large local noise. For more introductory resources, see [31, 103, 169, 263]. Discussions on the computation and software can be found in [31].

Zig-zag persistent homology

The Zig-zag persistent homology was introduced in [45] to handle a more general setting where the sequence of spaces is connected with maps in arbitrary directions, in contrast to the classical persistent homology where the inclusion maps follow in the same direction. The framework is powerful since it allows simplicial complexes in the filtration to evolve with the insertion and deletion of simplices, which is suitable for analyzing dynamic point clouds or dynamic graphs with varying vertex connections [67, 95, 96, 154]. Various software for computing the zigzag persistence have been proposed [47, 97]. The framework has also been extended to the multiparameter setting [96, 98].

Parametrized homology

Parametrized homology [50] extends levelset zigzag persistent homology to a continuous-parameter setting. This framework leverages the theory of rectangle measures [56] to define four distinct continuous-parameter persistence diagrams, each corresponding to one of the four bar types in the levelset zigzag barcode. These diagrams encode homological features and capture detailed information about how each feature disappears at both endpoints of its defining interval.

Localized homology

Localized homology [334] is proposed for identifying the locations of topological attributes such as tunnels and voids. Given a topological space with a cover, the localized homology tool can be used to determine the location (the specific cover element) of topological features within the space. Specifically, it constructs a larger blowup complex [267] that is homotopy equivalent to the original space and has clear boundaries between cover elements. The homology bases are then computed on the blowup complex to localize the features to specific cover elements. An efficient algorithm is also provided for practical applications.

Evolutionary homology

Cang et al. [41] introduced Evolutionary Homology (EH), a method generating time-dependent topological invariants ("evolutionary barcodes") for individual components within physical systems. EH works by coupling differential equations or chaotic oscillators to model system interactions, creating coupled dynamics. On the resulting trajectories of these oscillators, they define simplices, simplicial complexes, algebraic groups, and topological persistence using a time evolution-based filtration. This process reveals the topology-function

relationship of individual components and is particularly effective for analyzing local topology within point cloud data. EH was successfully applied to realistic problems in biophysics.

Multiresolution persistent homology

Multiresolution persistent homology was proposed for analyzing excessively large datasets [315, 316]. The idea is to match the scale of interest in the data by employing the rigidity-density functions modulated through resolution parameters. By tuning these parameters, the original large-scale datasets can be represented at an appropriate lower resolution, enabling efficient topological analysis.

Multidimensional persistence

The multidimensional persistence was first introduced in 2007 for analyzing the topology of multifiltrations [46]. Later, another version of multidimensional persistence [314] was proposed for transforming both dynamic and single molecular data into image-based representations. For dynamical data, the image representation is constructed by repeatedly applying persistent homology to each data frame across time steps. For single molecular data, persistent homology is applied at multiple spatial scales to the volumetric density representation of the molecule, and the resulting topological features are aggregated to form the final image.

Weighted persistent homology

The weighted persistent homology is used to study the weighted data, such as weighted simplicial complexes, where each simplex is assigned a weight. There are primarily two types of methods for capturing the weight information: one approach uses weights to refine the filtration process [115, 247], while the other revises the boundary operators of the chain complexes by incorporating the weights [87, 257]. The weighted persistent homology has been applied to various biomolecular data analysis applications [220, 251].

Persistent path homology

Persistent path homology has been introduced for analyzing directed networks that incorporate asymmetric directional information [69], which integrates the filtration method in TDA with the path homology theory. Path homology [137], recently renamed GLMY homology [158], has been extensively developed in both theory [104, 138, 140, 186, 187] and applications [59, 124, 135, 312].

Persistent hypergraph homology

Persistent hypergraph homology [33] is a generalization of the standard persistent simplicial homology to hypergraphs. Using embedded homology theory, hypergraph homology is suitable for addressing systems with incomplete information [201], such as the coauthor network, where a three-person collaboration

exists but not all pairwise collaborations exist. A stability result [256] and a discrete Morse theory [258] for hypergraphs have been established, and an algorithmic implementation for persistent hypergraph homology [205] has also been developed.

Persistent super-hypergraph homology

Persistent super-hypergraph homology [136] provides a unified and generalized framework that encompasses persistent simplicial homology, persistent hypergraph homology, and persistent Δ -set homology as special cases. Compared with a simplicial complex and a hypergraph, a distinct characteristic of super-hypergraph is that each super-edge is not uniquely determined by its vertices, implying the presence of additional internal structure. This makes it a flexible framework to capture the complexity of various systems [123, 206].

Graph-complex-based persistent homology

In TDA, modeling data with an appropriate simplicial complex is as critical as data representation in machine learning. The most commonly used constructions are the Vietoris-Rips complex [332] and the Alpha complex [113], mainly due to the availability of well-established and user-friendly software packages. Several studies have explored the use of other graph-based complexes, such as the flag complex [156, 210], Dowker complex [203], neighborhood complex [200], Hom-complex [204], and others [136, 165]. Selecting the most suitable complex for a given application, rather than just using the popular choices, is worth further investigation.

Extended persistent homology

Distances between persistence diagrams are often used to compare topological features of different topological spaces. However, persistent diagrams of topological spaces with different Betti numbers can lead to infinite distances under any filtration. Extended persistent homology [74, 292] addresses this limitation by extending the filtration process so that all homology classes are assigned a birth and a death.

Join persistent homology

Join persistent homology [305] is based on the join operation in topology. The idea is that different topological features can be captured from multiple perspectives by employing various types of unit in the join process. This allows the method to reveal diverse structural information within the data. Join persistent homology also provides a solution to the issue of infinite distance between persistence diagrams of spaces with different Betti numbers.

Persistent cohomology

Persistent cohomology is a sister counterpart to persistent homology, offering computational advantages in time and space. In practice, people often

compute the persistent cohomology and subsequently derive persistent homology via the duality [90], as the former can be more efficient to compute. An example is the GUDHI package [211]. Many applications rely on the original persistent cohomology algorithm [89]. Several studies have explored incorporating domain-specific information into cohomology classes to enhance the characterization of background-related information in applications [39, 201].

Cayley-persistence

Cayley-persistence is formulated based on the group grading [27]. Similar to the standard \mathbb{Z} -graded persistence, it has the decomposition results. The key difference is its ability to capture more flexible persistence patterns. It allows the emergence of persistence not only on the intervals and grids, but also on more complex structures introduced by the underlying group grading.

Morse theory

Morse theory [231] provides a framework for analyzing and characterizing the topology of shapes. The central idea is to represent a shape as a topological space equipped with a Morse function that captures critical topological features through its critical points. Discrete Morse theory is integrated with persistent homology for unweighted and undirected networks [168]. Computational algorithm for Morse theory has been developed [145]. The persistent Homology of Morse decompositions has been applied to combinatorial dynamics [102]. Among the various descriptors derived from Morse theory, the Reeb graph is probably the most widely used [29, 130, 254, 323]. A key property of the Reeb graph is its ability to encode the topological structure of a shape into a one-dimensional graph, independent of the intrinsic dimension of the underlying manifold.

Conley index

Much of dynamical system analysis involves the existence and structure of invariant sets. Conley index theory, developed by Conley and his followers [61, 76, 225], is an important topological tool for studying such invariant sets. Rather than analyzing arbitrary invariant sets, Conley index theory focuses on isolated invariant sets, which are more robust to continuous perturbations than general invariant sets. For an isolated invariant set S , a quotient space can be constructed from its index pair, the homotopy Conley index of S is the homotopy type of this quotient space, and the cohomological Conley index of S is the Alexander-Spanier cohomology of the quotient space with integer coefficients. The Conley index is invariant under certain deformations of the dynamical system and can be used to establish the existence of fixed points and periodic orbits of the dynamical system. Recently, the Conley index has been extended to the persistence setting for data analysis [100, 101].

Persistent equivariant cohomology

For a dataset with a group action on it, persistent equivariant cohomology captures not only the underlying shape of the dataset but also properties of the group action, namely, the symmetries inherent in the data. In [5], the authors provide an explicit description of the persistent equivariant cohomology of the circle action on the Vietoris-Rips metric thickening of the circle. In particular, they derive explicit formulas for the persistent equivariant cohomology of these Vietoris-Rips thickening. When equivariant cohomology is based on a free group action, it can be viewed as a special case of twisted cohomology, with the twist arising from the group action. Various studies have explored the theory of twisted (co)homology in different contexts [139, 188, 322].

Persistence over posets

Persistence over posets [175] is a general framework where persistence modules are defined on posets, with multiparameter persistence as a special case where the indexing poset is \mathbf{R}^n . As the most known invariant for multiparameter persistence, the rank invariant [46] has evolved into the generalized rank invariant [244] for persistence over posets by naturally extending its domain. In particular, for any interval decomposable persistence module of finite-dimensional vector spaces, its barcode can be extracted from the generalized rank invariant by the principle of inclusion-exclusion [174]. This implies that the generalized rank invariant is a complete invariant for interval decomposable persistence modules, while the rank invariant is not a complete invariant for interval decomposable multiparameter persistence modules [32].

2.1.4 Additional perspectives

Despite the great utility of persistent homology, there are a number of directions in where it is desirable to extend it. These extensions parallel much of the work that has been done in mainstream algebraic topology, i.e., the topology of actual spaces and manifolds rather than point clouds. Here are a few of them.

- **Spectral analysis:** Note that persistent homology fails to capture the homotopic geometric shape changes within a filtration, meaning that it cannot distinguish between different shapes with the same topology [153]. Several frameworks have been developed to address this limitation based on the spectral analysis of Laplacians [310]. These methods not only recover topological information from persistent homology but also capture additional information that describes the homotopic geometric shape changes in the filtration.
- **Analytic methods:** From the early days of the subject, it was understood that in the case of manifolds, there are approaches to computing a version of homology (homology with \mathbb{R} coefficients) using differential forms. These methods rely on *de Rham cohomology* [306]. The advantage of these methods is that they tie the homological qualitative information together with the

analysis on the manifold. Another advantage is that the *Hodge theorem* [306] provides a choice, within a cohomology class, of a harmonic representative, which yields clarity about the significance of the class. A key ingredient within this analysis is the Hodge Laplacian, a partial differential operator on differential forms.

- **Sheaves:** After the initial development of homological methods in the early part of the 20th century, by 1950 it became understood that in order to gain more power in the use of the homological methods for the understanding of spaces, one needed the more complex notion of *cohomology with coefficients in a sheaf* pioneered by J. Leray [185]. This notion became critical to the extension of topological techniques to algebraic geometry and number theory [92, 93, 311].
- **Local to global computation:** The direct computation of topological invariants using the linear algebraic methods, is unmanageable except in the smallest toy examples. For this reason, algebraic topologists were forced to develop methods for parallelizing the computations, and for using structures inherent in the problem, such as a fibration structure. Examples of these ideas are the use of Mayer-Vietoris long exact sequences and spectral sequences, the Serre spectral sequence for a fibration, Eilenberg-Moore spectral sequences [216].
- **A priori knowledge:** Often a topological problem is equipped with a priori information, such as a map to a reference base space or some choices of subspaces within the space. Homological methods have been refined to capture this additional information in more subtle invariants.

2.2 Persistent combinatorial Laplacians

To address the limitation that persistent homology fails to capture the homotopic geometric changes during a filtration, recent advances have introduced persistent combinatorial Laplacians [302], and some other variations of persistent Laplacians [60, 163, 202, 300, 309, 329] that encode in spectra both the topological and the geometric or combinatorial structures of data. Compared to persistent homology, these persistent Laplacians enable deeper analysis of datasets. Their harmonic spectra (i.e., zero eigenvalues) fully recover the topological outputs from the persistent homology, while the non-harmonic spectra (i.e., nonzero eigenvalues) capture additional geometric or combinatorial information about the shape of data. This additional information is particularly useful in describing the homotopic geometric shape changes in a filtration, which the persistent homology fails to detect. A comprehensive survey of persistent Laplacians is given by Wei and Wei [310].

These persistent topological Laplacians can be viewed as multiscale formulations of combinatorial Laplacians [111], which originate from spectral graph theory and generalize the notion of graph Laplacians [176] to higher-dimensional simplicial complexes. The study of the spectra of combinatorial Laplacians is crucial in understanding the shape of data and has gained much attention in recent years [134, 144, 155].

Below, we begin with an overview of combinatorial Laplacians, then discuss the persistent combinatorial Laplacians, and finally cover several other variations of persistent topological Laplacians.

2.2.1 Combinatorial Laplacians

The combinatorial Laplacians extend the classical graph Laplacians to simplicial complexes [111]. Compared to graph Laplacians, which only encode connections between nodes, the combinatorial Laplacians are constructed using the boundary operators that capture the relationships between simplices even in higher dimensions, including vertices, edges, triangles, and higher-dimensional structures.

The k -th combinatorial Laplacian L_k on the k -th chain group $\mathcal{C}_k(K)$ is given as follows:

$$L_k = \partial_{k+1}\partial_{k+1}^* + \partial_k^*\partial_k,$$

where ∂_k^* is the adjoint operator of the boundary operator ∂_k with respect to an inner product defined on the chain group $\mathcal{C}_k(K)$. Note that here the combinatorial Laplacian L_k is defined on the chain group, a corresponding construction can also be made on the cochain group, see [155]. The combinatorial Laplacian L_k is positive semi-definite symmetric, and thus its spectrum consists of only non-negative real eigenvalues. In the case that K is a graph, $L_0 = \partial_1\partial_1^*$ reduces to the graph Laplacian. In addition, the combinatorial Laplacians lead to a combinatorial version of Hodge decomposition:

$$\mathcal{C}_k = \text{im } \partial_{k+1} \oplus \ker L_k \oplus \text{im } \partial_k^*,$$

see [191] for further details.

In the case where the inner product on the chain group $\mathcal{C}_k(K)$ is defined by the Kronecker delta $\delta_{i,j}$, i.e., all basis elements in $\mathcal{C}_k(K)$ are considered mutually orthogonal, the boundary operator ∂_k can be represented as a matrix of size $N_{k-1} \times N_k$ and its adjoint ∂_k^* given as the transpose of the matrix. Here N_k denotes the number of k -simplices.

It has been shown in [111] that the kernel of the k -th Laplacian on a simplicial complex K corresponds to its k -th simplicial homology group:

$$\ker L_k = H_k(K),$$

thereby capturing the topological information of the simplicial complex. Its dimension is given by the k -th Betti number. In addition, the non-harmonic spectra capture the geometrical or combinatorial features of the underlying structure. In particular, as noted in graph theory, the first nonzero eigenvalue describes the connectedness of a graph.

2.2.2 Persistent combinatorial Laplacians

To incorporate the ideas of persistence into the combinatorial Laplacians, Wang et al. [302] introduced the persistent combinatorial Laplacians, which is the first work, both theoretically and practically, to establish a connection between the persistent homology and the combinatorial Laplacians. A notion of persistent Laplacian also appeared in a seminar note [190]. A comprehensive theoretical study of persistent combinatorial Laplacians can be found in [217] and a stability result is found in [195].

Given a pair of simplicial complexes $K^i \subset K^j$ in a filtration, the k -th persistent combinatorial Laplacian on the k -th chain group $\mathcal{C}_k(K^i)$ is defined as follows:

$$L_k^{i,j} = \partial_{k+1}^{i,j} \left(\partial_{k+1}^{i,j} \right)^* + (\partial_k^i)^* \partial_k^i,$$

where ∂_k^i is the k -th boundary operator on K^i , $\partial_{k+1}^{i,j}$ is called the persistent boundary operator, given by the restriction of ∂_{k+1}^j to a subgroup $\mathcal{C}_{k+1}^{i,j}$ of the $(k+1)$ -th chain group $\mathcal{C}_{k+1}(K^j)$ so that the images of its elements are in $\mathcal{C}_k(K^i)$, and $(\partial_k^i)^*$ and $\left(\partial_{k+1}^{i,j} \right)^*$ are their adjoint operators, respectively. In the case that $i = j$, the k -th persistent boundary operator equals the standard boundary operator, i.e., $\partial_{k+1}^{i,i} = \partial_k^i$, and thus the k -th persistent combinatorial Laplacian reduces to the standard k -th combinatorial Laplacian on K^i . The persistent combinatorial Laplacian $L_k^{i,j}$, analogous to the combinatorial Laplacian, is positive semi-definite symmetric, with its eigenvalues real and non-negative. This operator also results in a persistent Hodge decomposition [190] as follows:

$$\mathcal{C}_k(K^i) = \text{im } \partial_{k+1}^{i,j} \oplus \ker L_k^{i,j} \oplus \text{im } (\partial_k^i)^*.$$

The implementation of persistent combinatorial Laplacians has been made available as an open-source software package, called HERMES [303]. The algorithm is based on a practical construction of filtration of alpha complexes [113]. By constructing a single k -th boundary operator for the final complete alpha complex (i.e., the Delaunay tessellation [91]) and certain projection matrices, the matrix representations of $\left(\partial_{k+1}^{i,j} \right)^*$ and $(\partial_k^i)^*$ are both given simply by the transpose of the matrices representing $\partial_{k+1}^{i,j}$ and ∂_k^i . An alternative algorithm for computing persistent combinatorial Laplacians can be found in [217].

An immediate result is that the kernel of the k -th persistent combinatorial Laplacians is isomorphic to the k -th persistent homology group [190, 217]:

$$\ker L_k^{i,j} = H_k^{i,j},$$

and thus its dimension is the k -th persistent Betti number. The harmonic spectra of persistent combinatorial Laplacians, therefore, recover the topological

information from the persistent homology, while the non-harmonic spectra provide additional geometric or combinatorial evolutionary features of the shape of the data.

Persistent combinatorial Laplacians have been applied to various biological problems, such as protein thermal stability [302], protein-ligand binding [219], and protein-protein binding problems [307], etc. In particular, they have also been successfully used to forecast the SARS-CoV-2 variants BA.4 and BA.5 [65]. The advantage of persistent combinatorial Laplacians over persistent homology has been demonstrated in protein engineering [253].

2.2.3 Other variations

We now outline various variations of persistent combinatorial Laplacians proposed for analyzing point-cloud data with other additional structures, such as directional information or more complex relationships.

Persistent path Laplacians

Persistent path Laplacians [300] was introduced by Wang and Wei for analyzing point data with directional information, such as directed graphs and networks. The framework builds upon the path homology [137], which was later extended to a persistent framework on directed networks [69]. A weighted path homology for weighted digraphs was introduced in [192]. More recently, an efficient algorithm for computing the 1-dimensional persistent path homology was proposed in [99].

Persistent hyper(di)graphy Laplacians

Persistent hypergraphy Laplacians [202] and hyperdigraphy Laplacians [60] are proposed for complex systems with and without directional information, respectively. These constructions are based on the infimum and supremum chain complexes in embedded homology [33], enabling the definition of the corresponding infimum and supremum Laplacians. Particularly, when the hypergraph reduces to a simplicial complex, the infimum and supremum chain complexes coincide, recovering the simplicial Laplacian as a special case.

Persistent hopping path-Laplacian

Persistent hopping path Laplacian [199] generalizes the k -hopping path Laplacian of graphs [118] in two aspects: first, by extending from graphs to simplicial complexes and hypergraphs, and second, by incorporating persistence into the framework. By selecting different hopping parameters, this method provides a comprehensive description of pairwise, many-body, direct, and indirect interactions within the data structures.

Persistent directed flag Laplacian

Persistent directed flag Laplacian [163] is constructed based on the flag complex of directed graphs, which only includes those complete subgraphs that

satisfy specific order constraints. By integrating the directed flag complex with the standard simplicial Laplacian, it provides an efficient tool for analyzing complex networks that exhibits asymmetric directional relationships. Persistent directed flag Laplacian has been applied to complex biomolecular data [329].

Persistent Laplacian for simplicial maps

Persistent Laplacian for simplicial maps generalizes the standard persistent Laplacian defined on a filtration with inclusion maps [143]. It maintains a connection to persistent homology, where the nullity of the persistent Laplacian equals the persistent Betti numbers associated with the simplicial map-induced space sequence. A key difference is that its persistence depends not only on the sequence of simplicial complexes, as in the standard setting, but also on the structures of the simplicial maps themselves.

Multiscale Hochschild Laplacian

Multiscale Hochschild Laplacian [150] offers an efficient tool for analyzing digraph data by integrating Hochschild Laplacians with multiscale analysis. Particularly, the Hochschild Laplacian is based on the truncated path algebras of the digraphs, as the full path algebra of a digraph with directed cycles is infinite-dimensional, making it unsuitable for computation and practical applications. Similar to the combinatorial Laplacians, the kernel of the Hochschild Laplacian is isomorphic to the Hochschild cohomology in the corresponding dimension.

2.3 Persistent Dirac operator

Recently, quantum algorithms have been introduced for the efficient computation of topological and geometrical features to address the computational complexity of persistent homology, especially for large point-cloud data [9, 16, 28, 179, 208]. These methods allow the data to be accessed in quantum parallel on a quantum computer and the combinatorial simplicial complexes to be encoded as quantum states, enabling an exponential speed-up over the classical topological data analysis techniques. Central to these methods is the Dirac operator [106] on simplicial complexes, whose square yields exactly the combinatorial Laplacians of all degrees. It captures the same information as combinatorial Laplacians but provides additional information as it is more sensitive to local topological and combinatorial or geometrical features as a linear operator. The Dirac operator has been found particularly useful in analyzing topological signals [36], and has also been applied to molecular representations [308].

The persistent Dirac operator [9, 10] extends the Dirac operator to the persistent setting for analyzing point-cloud data. Its square yields persistent combinatorial Laplacians and thus can be seen as a generalization of the persistent combinatorial Laplacian technique, capturing the same topological information while encoding more local topological and geometric structures of

the underlying data. The Dirac operator has also been extended for tracking the persistence of paths and hypergraphs [286]. The Persistent Mayer Dirac [287] has been developed based on N -chain complexes, expanding the scope and improving the applicability of persistent Dirac-based techniques.

In the following, we briefly review the Dirac operator and the persistent Dirac operator. We also discuss the other variations for data with additional structures.

2.3.1 Dirac operator

In the combinatorial case, the Dirac operator also connects simplices of consecutive dimensions in a simplicial complex K as the combinatorial Laplacians. It is defined as the sum of the boundary operator and its adjoint, given by $D = \partial + \partial^*$. It can be viewed as the square root of the combinatorial Laplacian defined in the full chain complex, i.e., the direct sum of all chain groups $\mathcal{C}(K) = \bigoplus_k \mathcal{C}_k(K)$. When expressed in the Hermitian representation, its square leads to a block matrix with diagonals given by the combinatorial Laplacians in different degrees:

$$D^2 = \begin{pmatrix} 0 & \partial_1 & 0 & \cdots \\ \partial_1^* & 0 & \partial_2 & \cdots \\ 0 & \partial_2^* & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}^2 = \begin{pmatrix} L_1 & 0 & 0 & \cdots \\ 0 & L_2 & 0 & \cdots \\ 0 & 0 & L_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Compared to the combinatorial Laplacian, which preserves the degree and maps a k -chain to a k -chain, the Dirac operator maps a k -chain to both $(k-1)$ - and $(k+1)$ -chains. In addition, the Dirac operator is a first-order differential operator, which means that it is more sensitive to local geometric and topological structures compared to the Laplacians, which are second-orders. The kernel of the Dirac operator encodes the Betti numbers of a simplicial complex across all degrees, and the nonzero eigenvalues are the square roots of those of the combinatorial Laplacians.

The key advantage of using the Dirac operator in topological data analysis lies in its suitability for efficient quantum computations. Lloyd et al. [208] introduced the first quantum algorithm for estimating Betti numbers by using Dirac operators, with an exponential speed-up over the classical algorithm in topological data analysis. The method constructs a quantum state that represents the simplicial complex at each scale, and then use a quantum phase algorithm [235] to find the dimension of the kernel of the Dirac operator D in Hermitian representation. The method provides a multiscale version of Betti numbers without explicitly tracking the persistence in filtration.

2.3.2 Persistent Dirac operator

The Dirac operator was extended by Ameneiro et al. [9] to the persistent setting for the first time in a quantum implementation for computing the

persistence of point clouds using the Dirac operator across different scales. The framework utilizes a single k -th boundary operator ∂_k encoded using 1-qubit Pauli X gates defined on the set of all possible k -simplices in the filtration, and also a projection matrix P_k^i from this set to the set of k -simplices in the simplicial complex K^i . The k -th persistent boundary operator is then given as $\partial_k^{i,j} = P_{k-1}^i \partial_k P_k^j$ for a pair of simplicial complexes $K^i \subset K^j$ stored as quantum states in the filtration. The k -th persistent Dirac operator as Hermitian is defined in [9] as

$$D_k^{i,j} = \begin{pmatrix} 0 & \partial_k^{i,i} & 0 \\ (\partial_k^{i,i})^* & 0 & \partial_{k+1}^{i,j} \\ 0 & (\partial_{k+1}^{i,j})^* & 0 \end{pmatrix} - \xi \begin{pmatrix} P_{k-1}^i & 0 & 0 \\ 0 & -P_k^i & 0 \\ 0 & 0 & P_{k+1}^j \end{pmatrix}$$

with $\xi \in \mathbb{R}$, and its square is given by

$$(D_k^{i,j})^2 = \begin{pmatrix} \partial_k^{i,i} (\partial_k^{i,i})^* & 0 & \partial_k^{i,i} \partial_{k+1}^{i,j} \\ 0 & L_k^{i,j} & 0 \\ (\partial_{k+1}^{i,j})^* (\partial_k^{i,i})^* & 0 & (\partial_{k+1}^{i,j})^* \partial_{k+1}^{i,j} \end{pmatrix} + \xi^2 \begin{pmatrix} P_{k-1}^i & 0 & 0 \\ 0 & P_k^i & 0 \\ 0 & 0 & P_{k+1}^j \end{pmatrix}$$

Based on its definition, the square of the persistent Dirac does not lead to a diagonal block matrix as in the standard case, but just one of the blocks provides the k -th persistent combinatorial Laplacian shifted by ξ^2 . It follows that the eigenvalues of the k -th persistent combinatorial Laplacian correspond to a ξ^2 shift of the squared eigenvalues of the k -th Dirac operator. In the computation of persistent Betti numbers, the shift ξ is chosen to be the positive eigenvalue of the Dirac operator. As noted in [9], setting $\xi = 0$ results in overcounting of the states in the kernel of the Laplacian and thus an incorrect estimate of Betti numbers.

The persistent Dirac operator has been applied for analyzing time series data [10]. Another quantum implementation of computing the persistent Betti numbers was proposed by Hayakawa [149], which is based on the block-encoding of persistent combinatorial Laplacians. While these methods provide efficient computations of the persistent topological features, their realization requires fault tolerant quantum computers, which are unlikely to be available in the near future for realistic big data applications.

2.3.3 Other variations

Here we provide a few variations of the persistent Dirac operator recently developed for data with different structures.

Persistent path Dirac and hypergraph Dirac operators

The path Dirac and hypergraph Dirac operators, along with their persistence extensions, were introduced in [286] for extracting valuable information from path complexes and hypergraphs. Their application to molecular structure analysis has also been explored.

Persistent Mayer Dirac operators

The Mayer Dirac and persistent Mayer Dirac operators on N -chain complexes were proposed in [287] and applied to molecular structure analysis. Note that Mayer homology [215] differs from the standard homology. In this context, the boundary operator satisfies $\partial^N = 0$ with $N \geq 2$ instead of the usual condition $\partial^2 = 0$. For a description of tools based on Mayer topology, see Section 2.4.2.

Khovanov Dirac

The Khovanov Dirac [162] was proposed for analyzing the global topological features of knots and links. It is based on Khovanov homology, a powerful topological invariant in geometric topology for studying Knots and links. This topic will be discussed further later in Section 4.3.

2.4 Other formulations

In this section, we outline formulations recently developed in topological data analysis that utilize alternative homology theories, extending the standard homology for analyzing point-cloud data.

2.4.1 Persistent sheaf theory

The aforementioned methods, such as persistent homology and persistent combinatorial Laplacians, have proven to be powerful for analyzing the topological structure of data [253]. However, they remain limited in their ability to describe the local features along with the relationships among them, which often play a crucial role in practical applications.

Sheaf theory provides a more powerful and general framework for understanding topological spaces. In particular, it enables the study of local-to-global properties of structures defined on spaces. In recent years, it has found various applications in topological data analysis [78, 79, 133, 147, 148, 321], particularly through the notion of cellular sheaves [271], which is more accessible to researchers and easier to implement for practical computations. For a simplicial complex, a cellular sheaf can be thought of as an assignment of a vector space to each simplex, together with linear transformations, called restriction maps, defined between the vector spaces by face relations. These restriction maps must satisfy certain consistency conditions, allowing global information to be determined from local features. This makes cellular sheaves valuable for analyzing data with localized properties, from which global features can be inferred.

Due to the direction of the restriction maps, sheaves are naturally associated with cohomology. The consistency conditions lead to a well-defined coboundary operator δ on sheaf cochain complexes satisfying $\delta\delta = 0$, which defines the sheaf cohomology, with the standard cohomology being a special case. A sheaf-theoretical perspective on persistent homology was first investigated in [83] and later extended to the persistent setting in [261, 319]. A notion of persistent sheaves was also proposed in [146].

By introducing an inner product on the cochain groups, the sheaf Laplacians [148] can be defined. Their kernels correspond to the sheaf cohomology groups with dimensions given by the Betti numbers. By incorporating persistence, Wei and Wei [309] proposed persistent sheaf Laplacians to address the limitation of persistent combinatorial Laplacians in handling data points with labels or weights. The framework generalizes persistent combinatorial Laplacians, as the latter can be seen as a special case of the former when all data points are assigned the same quantity. Given a pair of sheaves in a filtration, the persistent boundary operator is defined analogously to the case of persistent combinatorial Laplacians. The main difference lies in the underlying structure of the data and the use of cohomology instead of homology, along with a coboundary operator defined by scalar multiplication subject to consistency rules. The persistent sheaf Laplacians follow with the same formulas, with kernels isomorphic to the persistent cohomology groups and dimensions given by the persistent Betti numbers. In addition, the matrix representations of persistent sheaf Laplacians can be calculated in the same way as for the persistent combinatorial Laplacians [309].

2.4.2 Persistent Mayer topology

Note that all aforementioned methods in topological data analysis rely on the construction of the standard chain complexes, where the boundary operator satisfies $\partial^2 = 0$. These methods might be limited to revealing more intricate information about the underlying structure of data as the boundary operator captures only the relationship between simplices of consecutive dimensions.

To address this, Shen et al. [269] introduced the persistent Mayer homology and persistent Mayer Laplacians, which generalize the Mayer homology [215] to the persistent setting, allowing the study of connections between simplices across multiple dimensions rather than just adjacent ones. The approach is built on an N -chain complex with the boundary operator satisfying a more general condition $\partial^N = 0$ with $N \geq 2$. This extends the current existing frameworks based on standard homology where $\partial^2 = 0$. For a discussion about the relationship between Mayer homology and the standard homology, see [278].

Analogous to the standard case, the kernels of the persistent Mayer Laplacians correspond to the persistent Mayer homology groups, with ranks given by the persistent Mayer Betti numbers. However, in contrast to the standard persistent homology governed solely by the filtration parameter, the persistent

Mayer homology and persistent Mayer Laplacians for each degree k are controlled by an additional parameter. The framework thus enables the encoding of richer topological and geometric or combinatorial information about the underlying structure, in particular, the relationship between simplices across different dimensions. A further generalization of persistent Mayer Laplacians, called persistent Mayer Dirac, has been proposed in [287], which is more sensitive to local features.

Persistent Mayer Laplacians and the Mayer Dirac have been applied to the study of molecular structure analysis [269, 287]. Persistent Mayer homology has been used as multichannel features for predicting protein-ligand binding affinities [122].

2.4.3 Persistent interaction topology

In algebraic geometry, intersection theory studies the intersection of subvarieties within a given variety. In topology, intersection (co)homology studies singular spaces. In the discrete setting, interaction involves understanding how components within a complex system influence each other. This concept is particularly important in fields such as complex networks, molecular interactions, and organism interactions in biological systems. Traditional tools in topological data analysis often lack the ability to capture such intersections, as they typically focus on global topological features instead of focusing on localized interactions between elements in the system.

Recently, motivated by the notion of interaction cohomology [177], Liu et al. [193] developed the interaction topology to gain insight into the interactions between topological spaces and characterize such interactions using topological invariants. The theory is built on the category of interaction spaces. In the paper, they introduced the concept of interaction spaces, studied their homotopy and singular homology, and proved that interaction singular homology is a homotopy invariant. In addition, they proposed the interaction simplicial homology and cohomology to facilitate applications in point-cloud data analysis.

Based on interaction homology, Liu et al. [194] proposed the persistent interaction homology and persistent interaction Laplacians, and demonstrate their stability. The paper also provides an algorithm for computing the interaction homology and interaction Laplacians for finite point sets. The strength of the framework is demonstrated through the analysis of interactions among different types of atoms within molecules.

3 Differential topology approaches

In this section, we shift our attention to differential topology analysis of data on differential manifolds. The approach uses tools from differential topology, such as vector fields, differential forms, Laplace-Beltrami operators, and Hodge Laplacians, along with their extensions to multiscale analysis, to understand

the underlying structure of differential manifolds. These tools are deeply connected to the topology of the underlying manifolds through de Rham-Hodge theory, a landmark theory of the 20th century mathematics, and thus are well-suited for analyzing data naturally represented as differential manifolds. Examples include 2D surfaces, 3D solids, and image data such as 2D slices or 3D volumes from magnetic resonance imaging (MRI) or computed tomography (CT) scans. These data are commonly found in fields such as computer graphics, biological modeling, and medical imaging.

Although in many scenarios, techniques in algebraic topology analysis for point cloud data can be applied to data using point cloud representation, it is more appropriate to model the data as continuous or differential manifolds. These data are naturally continuous in observation, even though they are practically represented as discrete points. Relying solely on sampled points can lead to inconsistencies due to their varying combinatorial representations, resulting in different outcomes using algebraic and combinatorial techniques. A more natural solution is to model the data theoretically as continuous or differential manifolds, and perform discretization only in the implementation. This way ensures that the analysis remains consistent and stable regardless of different discretizations, and the results are convergent.

In addition, the study of differential manifolds naturally incorporates the boundary conditions, such as the normal (Dirichlet) and tangential (Neumann) boundary conditions, which are significantly important in applications where the boundary carries physical or geometric meaning, such as electromagnetism and fluid dynamics. These boundary conditions provide constraints that govern the behavior of a system near the boundary, and ensure the existence and uniqueness of solutions when solving boundary value problems. It also ensures the well-posedness of differential operators and Hodge Laplacians' capability to perform differential calculus on manifolds with boundary. However, there is no intrinsic notion of domain boundary for combinatorial Laplacians.

The implementation of de Rham Hodge theory has been challenging for years due to the complicated boundary conditions. Earlier methods, such as the standard finite element method [71], are insufficient to capture the manifold topology and can lead to entirely wrong eigenvalues for the vector Laplacians under normal or tangential boundary conditions [11]. With recent advances, such as the finite element exterior calculus (FEEC) [12] and the discrete exterior calculus (DEC) [94], the practical realization of the theory has become possible. These techniques ensure that the discretizations are consistent with the continuous theory and thus allow for accurate computation of the cohomologies of the underlying manifolds. Based on these techniques, computational frameworks have been developed for computing Hodge decompositions, Hodge Laplacians and their multiscale extension. These developments have enabled the application of de Rham Hodge theory in topological data analysis, opening new directions for analyzing the topological and geometric features of data on manifolds.

Below, we start with a review of de Rham-Hodge theory for manifolds with boundary, including the de Rham cohomology, Hodge Laplacians, and Hodge decomposition, along with their recent implementations. We then provide a short discussion on persistent de Rham cohomology [64]. We also focus on the recently developed persistent Hodge Laplacians [64, 282] in capturing the topological and geometric features of 2D and 3D compact domains. We include both the Lagrangian formulation [64] and the Eulerian formulation [282] of the de Rham-Hodge theory, depending on whether the underlying manifolds are discretized using simplicial meshes or represented as regions in a Cartesian grid defined by level-set functions, respectively.

3.1 De Rham-Hodge theory for manifolds with boundary

The de Rham-Hodge theory offers a foundational framework for understanding the topological and geometrical structures of manifolds through differential forms. Its formulation is particularly subtle in the case of manifolds with boundary, where appropriate boundary conditions, such as normal and tangential boundary conditions [272], must be imposed. These boundary conditions are essential for the core results of the theory, i.e., the correspondence between harmonic forms and the cohomology.

The notion of differential forms might be abstract at first. However, they are closely related to familiar concepts in vector calculus. For example, in \mathbb{R}^3 , a 0- or 3-form is just a scalar field, while a 1- or 2-form corresponds to a vector field. In general, differential forms can be viewed as a generalization of scalar and vector fields on differential manifolds. These correspondences between differential forms and scalar or vector fields provide an intuitive understanding of the boundary conditions in the case of 1-forms: a 1-form is normal if its corresponding vector field is normal to the boundary, or tangential if the field is tangential to the boundary. The formal definitions of normal and tangential boundary conditions, and also a complete correspondence between differential forms and their counterparts in vector calculus in \mathbb{R}^3 can be found in [324]. The two types of boundary conditions are dual to each other under the Hodge star operator.

The differential d^k (i.e., the exterior derivative) and the codifferential operator δ^k , are two fundamental operators in building the de Rham-Hodge theory. We briefly recall them here with their correspondences with classical operators in vector calculus for an intuitive understanding. The differential d^k maps k -forms to $(k+1)$ -forms and satisfies $d^k d^{k-1} = 0$. It generalizes and unifies the notions of gradient ∇ , curl $\nabla \times$, and divergence $\nabla \cdot$ in vector analysis, which correspond to d applied to 0-, 1-, and 2-forms in \mathbb{R}^3 , respectively. The codifferential δ^k can be defined as an adjoint operator of d^{k-1} with respect to the Hodge L^2 inner product induced from the Hodge star operator on the space of differential forms. It maps k -forms to $(k-1)$ -forms and satisfies $\delta^{k-1} \delta^k = 0$. In \mathbb{R}^3 , it corresponds to $-\nabla$, $\nabla \times$, and $-\nabla \cdot$ when applied to 2-, 1-, and 0-forms, respectively.

3.1.1 De Rham cohomology

Let Ω_n^k and Ω_t^k be the space of normal k -forms and the space of tangential k -forms over a smooth manifold M of dimension m , respectively. For their formal definitions, see [324]. These two spaces are dual to each other under the Hodge star operator, i.e., $\Omega_n^k \cong \Omega_t^{m-k}$. The de Rham cohomologies of differential forms are defined on two sequences connected by the differential d^k and the codifferential δ^k :

$$\begin{aligned} \dots &\xrightarrow{d^{k-2}} \Omega_n^{k-1} \xrightarrow{d^{k-1}} \Omega_n^k \xrightarrow{d^k} \Omega_n^{k+1} \xrightarrow{d^{k+1}} \dots, \\ \dots &\xleftarrow{\delta^{k-1}} \Omega_t^{k-1} \xleftarrow{\delta^k} \Omega_t^k \xleftarrow{\delta^{k+1}} \Omega_t^{k+1} \xleftarrow{\delta^{k+2}} \dots. \end{aligned}$$

These two complexes are well-defined as d preserves the normal boundary conditions and δ preserves the tangential boundary conditions. The resulting cohomologies correspond to the k -th relative de Rham cohomology $H_{dR}^k(M, \partial M)$ and the k -th absolute de Rham cohomology $H_{dR}^k(M)$ as follows:

$$\begin{aligned} H_{dR}^k(M, \partial M) &= \ker d^k / \text{im } d^{k-1}, \\ H_{dR}^k(M) &= \ker \delta^k / \text{im } \delta^{k+1}, \end{aligned}$$

where $\ker d^k$ (resp. $\ker \delta^k$) is called the space of closed (resp. coclosed) k -forms and $\text{im } d^{k-1}$ (resp. $\text{im } \delta^{k+1}$) is the space of exact (resp. coexact) k -forms. The ranks of $H_{dR}^k(M, \partial M)$ and $H_{dR}^k(M)$ are given by the $(m-k)$ -th Betti number β_{m-k} and the k -th Betti number β_k , respectively. Here the domains of these operators are restricted under the corresponding boundary conditions. For a detailed discussion of the absolute and relative de Rham cohomology and the corresponding isomorphisms, refer to [266].

3.1.2 Hodge Laplacians

The k -th Hodge Laplacian Δ^k on the space of k -forms is defined as follows:

$$\Delta^k = d^{k-1} \delta^k + \delta^{k+1} d^k.$$

When $k = 0$, the Hodge Laplacian reduces to the Laplace–Beltrami operator $\delta^1 d^0$ acting on the space of functions defined on the manifold. In the presence of the boundary, the Hodge Laplacian Δ^k has no direct correspondence with the manifold topology, as its kernel is infinite-dimensional. However, one can define the Hodge Laplacians Δ_n^k and Δ_t^k by restricting the Hodge Laplacian Δ^k to the space of normal forms and to the space of tangential forms, respectively. These two types of Hodge Laplacians lead to finite-dimensional kernels. The space $\ker \Delta_n^k$ corresponds to the relative de Rham cohomology, and $\ker \Delta_t^k$ to the absolute de Rham cohomology [126]. In addition, these two kernels are related by the Poincaré duality, leading to the following identifications:

$$\ker \Delta_n^k \cong H_{dR}^k(M, \partial M) \cong H_{dR}^{m-k}(M) \cong \ker \Delta_t^{m-k}$$

The elements in $\ker \Delta_n^k$ are called the normal harmonic k -fields, and those in $\ker \Delta_t^k$ are called the tangential harmonic k -fields. These identifications enable the use of either type of boundary conditions for the study of the topology of the underlying manifold and its boundary.

Note that similar to the combinatorial case, the harmonic spectra of Hodge Laplacians capture the topological information about the underlying manifolds with the number of zero-eigenvalues given by the Betti numbers. However, the non-harmonic spectra of the Hodge Laplacians encode rich geometric information such as the metric and the curvature of the underlying manifolds. This is fundamentally different from the combinatorial case, where the non-harmonic spectra reflect only the combinatorial or the discrete structural information rather than the smooth geometric features. For a thorough discussion about the similarity and difference between the combinatorial Laplacians and Hodge Laplacians, see [260].

The implementation of Hodge Laplacians has been realized through FEEC in [12] and through DEC [260, 324], with both approaches providing accurate eigenvalue results and preserving the cohomology of the underlying domain. In addition, a notion of boundary-induced graph (BIG) Laplacians has been introduced [260], defined on Cartesian domains with appropriate boundary conditions. It is a special case of the discretized Hodge Laplacians and is simpler to compute while delivering correct topological dimensions of the data. The Hodge Laplacian of degree 0, i.e., the Laplace-Beltrami operator, has been widely applied to the study of the shapes of manifolds. For example, ShapeDNA [259] provides a simple representation of a shape using the spectra of the operator. Global point signature (GPS) [262] encodes a shape using a vector of scaled eigenfunctions of the operator. ShapeDNA utilizes standard finite element approaches, while GPS employs tools in DEC.

3.1.3 Hodge decomposition

The Hodge decomposition is a central result in Hodge theory, which decomposes the space of differential k -forms into a direct sum of mutually orthogonal subspaces. A 3-component Hodge decomposition of differential forms, called the Hodge Morrey decomposition [230], is given as follows:

$$\Omega^k = d\Omega_n^{k-1} \oplus \delta\Omega_t^{k+1} \oplus \mathcal{H}^k$$

Here $\mathcal{H}^k = \ker d^k \cap \ker \delta^k \subset \ker \Delta^k$, known as the space of harmonic k -fields. When applied to vector fields on a domain, this decomposition states that any vector field can be orthogonally decomposed into a curl-free, a divergence-free, and a harmonic component, leading to the classical Helmholtz-Hodge decomposition [181]. The space \mathcal{H}^k is equal to the kernel of the Hodge Laplacian $\ker \Delta^k$ and isomorphic to the k -th de Rham cohomology group. However, in the presence of a boundary, the space \mathcal{H}^k is infinite-dimensional [266], providing no connections to the manifold topology. It has been shown that this space can be further split by enforcing boundary conditions, revealing the relative

and absolute cohomology [126]. In particular, there is a 5-component topology-preserving Hodge decomposition for compact domains in the Euclidean spaces [272], given as follows:

$$\Omega^k = d\Omega_n^{k-1} \oplus \delta\Omega_t^{k+1} \oplus \mathcal{H}_n^k \oplus \mathcal{H}_t^k \oplus (d\Omega^{k-1} \cap \delta\Omega^{k+1}),$$

where $\mathcal{H}_n^k = \ker \Delta_n^k$ and $\mathcal{H}_t^k = \ker \Delta_t^k$ are the restrictions of \mathcal{H}^k to the space of normal forms and the space of tangential forms, respectively, encoding the relative and absolute cohomology of the underlying manifold as discussed previously.

For the computational frameworks of Hodge decomposition, we focus here only on the complete topology-preserving 5-component Hodge decomposition. A review of the classical 3-component Helmholtz-Hodge decomposition is available in [26]. The discretization of the 5-component Hodge decomposition of vector fields has been proposed using their differential form representations for tetrahedral meshes [324] and for 2D and 3D compact domains defined by level set functions in Cartesian grids [281, 283]. Both methods are based on DEC. Other approaches, assuming piecewise-constant vector fields, have been developed for surface triangle meshes [249] and also tetrahedral meshes [248]. A discussion of this 5-component Hodge decomposition for compact domains in \mathbb{R}^3 in terms of vector and scalar fields can be found in [43].

3.2 Persistent de Rham cohomology

Persistent de Rham cohomology was first introduced by Chen et al. in [64]. The work provides a detailed discussion about the persistence of harmonic forms under normal and tangential boundary conditions. The persistent de Rham cohomology holds the same importance both in theory and practical applications as persistent homology does for point-cloud data. It shows promise as a foundation tool and opens new directions for analyzing the topological structures of data that can be represented as differential manifolds.

To define the persistence de Rham cohomology, we consider a manifold M , defined by a sublevel set of a level set function f , i.e., $M = f^{-1}([-\infty, c])$ for $c \in \mathbb{R}$. By varying the isovalue of f , a filtration of differential manifolds with inclusion maps can be obtained as follows:

$$M_0 \hookrightarrow M_1 \cdots \hookrightarrow M_{s-1} \hookrightarrow M_s,$$

where each M_i is the sublevel set of f corresponding to a different isovalue. The inclusion map for any pair of manifolds $M_i \hookrightarrow M_j$ in the sequence induces a unique extension on the normal differential forms $\Omega_n^k(M_i) \rightarrow \Omega_n^k(M_j)$ satisfying certain conditions. This map further induces a map on the cohomology groups $\psi_{i,j}^k : H_{dR,i}^k(M, \partial M) \rightarrow H_{dR,j}^k(M, \partial M)$ that tracks the evolution of topological features, namely, normal harmonic fields as $H_{dR,i}^k(M, \partial M) \cong \mathcal{H}_{n,i}^k$. Here $\mathcal{H}_{n,i}^k$ denotes the space of normal harmonic k -fields on M_i . One can then

define the persistent relative de Rham cohomology as the image of the map:

$$H_{dR,i,j}^k(M, \partial M) = \text{im } \psi_{i,j}^k = \ker d_i^k / (\text{im } d_j^{k-1} \cap \ker d_i^k).$$

Its rank provides the $(m - k)$ -th persistent Betti number $\beta_{m-k}^{i,j}$. The birth and death, as well as the persistence of normal harmonic fields, can then be defined correspondingly by examining the elements in the persistent cohomology groups, analogous to the combinatorial case. Note that only the spaces of normal differential forms are considered here. The dual formulation under the tangential boundary conditions can be constructed in a similar manner [64].

3.3 Persistent Hodge Laplacians

In the same paper [64], Chen et al. first theoretically proposed the persistent Hodge Laplacians. The paper also presents an implementation of multiscaled Hodge Laplacians using DEC for manifolds in the Lagrangian formulation, i.e., manifolds discretized using simplicial meshes. Subsequently, a computational framework of persistent Hodge Laplacians was proposed by Su et al. in [282] in the Eulerian setting with the adapted DEC from simplicial meshes to regular grids. The manifolds in this context are modeled as domains in a Cartesian grid bounded by isocurves or isosurfaces of level-set functions. These developments are based on combining de Rham cohomology with the evolution of differential manifolds in a filtration, and can be viewed as extensions of earlier work on persistent homology in the cubic setting [298, 299].

Given any pair of manifolds $M_i \hookrightarrow M_j$ in the filtration aforementioned in the previous subsection, the k -th persistent Hodge Laplacian on the space of normal k -forms $\Omega^k(M_i)$ is given as follows:

$$\Delta_{i,j}^k = d_{i,j}^{k-1} \delta_{i,j}^k + \delta_i^{k+1} d_i^k,$$

where d_i^k and δ^k is the k -th differential and codifferential on normal forms on M_i , respectively, and $d_{i,j}^{k-1}$ is the persistent differential operator from normal $(k-1)$ -forms on M_j to normal k -forms on M_i , with $\delta_{i,j}^k$ its adjoint operator. The persistent Hodge Laplacians can similarly be defined on the space of tangential forms. However, they are equivalent due to the duality of the normal and tangential boundary conditions. When $i = j$, the persistent Hodge Laplacian reduces to the usual Hodge Laplacian restricted to normal forms. The kernel of $\Delta_{i,j}^k$ defines the persistent harmonic fields. It is isomorphic to the persistent relative de Rham cohomology, and its rank leads to the persistent Betti number.

We now turn to the discretization frameworks and briefly discuss the Lagrangian and Eulerian formulations using DEC, depending on how the underlying manifolds are discretized.

3.3.1 Lagrangian formulation

In the Lagrangian formulation, the underlying manifolds are discretized as simplicial meshes, specifically triangular meshes in 2D or tetrahedral meshes in 3D. A discrete k -form is just a vector with each entry assigned to a k -simplex. For example, a discrete 0-form is a vector with values assigned to the vertices, while a 1-form has values corresponding to edges. The differential is a sparse matrix given as the transpose of the signed incidence matrix between simplices of consecutive dimensions. The codifferential is the product of the incident matrix and the discrete Hodge star operators, which takes the form of a diagonal matrix. For further details of these discrete operators, see [64, 324].

The Lagrangian formulation requires explicit tessellation of the domains, which relies on high-quality meshing tools such as CGAL [119], Gmsh [131], Pymesh [327], etc. This makes the implementation of persistent Hodge Laplacians challenging in this context, as generating consistent meshes throughout the filtration can be problematic, and the resulting discrete operators may not be sparse matrices. The multiscale version of Hodge Laplacians [64] in the Lagrangian formulation, in addition to its dependence on these meshing tools, has several limitations. It is sensitive and numerically inconsistent because of variations in tetrahedral mesh representations of a compact domain. The method is also computationally expensive, as it requires repeated tessellation of the domains to create the evolution of manifolds in a filtration. These limitations make the framework unsuitable for machine learning tasks, which require feature representation computed on the same scale for comparability, and also efficiency, which is essential for computing features for large datasets.

3.3.2 Eulerian formulation

The Eulerian formulation of persistent Hodge Laplacians [282] has been introduced with the adapted DEC to address the limitations in the Lagrangian case. A computational algorithm for persistent Hodge Laplacians is presented in the paper. The underlying manifold in this setting is modeled as a sub-level set of a level-set function on a standard Cartesian grid, i.e., a domain bounded by an isocurve or an isosurface of a level set function. The adapted DEC from simplicial meshes to regular grids follows a similar structure as in the Lagrangian case. Note that in the Eulerian formulation, adjusting the isovalues of a level-set function naturally leads to a nested sequence of grid complexes in a filtration, making the construction of discrete persistent Hodge Laplacians feasible.

The Eulerian representation offers several advantages: First, it eliminates the need for explicit tessellation, thereby removing the reliance on high-quality meshing tools. Second, the computation of Laplacians for all objects can be performed on a pre-designed Cartesian grid, the extracted features from different objects are thus consistent and comparable for machine learning tasks. In addition, as the vertices, edges, faces, and cells are fixed in a Cartesian grid, the data structure and the discrete operators are significantly simplified. In

particular, the evolution of manifolds can be easily obtained by adjusting the isovalue of the level-set functions. These make the computation of Laplacians more efficient in the Eulerian formulation. The framework was demonstrated in the same paper by integrating it into machine learning models for predicting the protein-ligand binding affinities.

4 Geometric topology approaches

This section focuses on methods grounded in geometric topology, in particular knot theory, for analyzing curves embedded in 3-space, such as knots, links, tangles, and linkoids. Such data structures are often found in chemical compounds [241], DNA [13], and protein structures [284]. Knot theory studies these objects up to ambient isotopy, informally referring to a continuous deformation that does not involve cutting and rejoining. The key tools in Knot theory are knot invariants, which are quantities that remain unchanged under these deformations and play an essential role in distinguishing knots and links. Examples include the crossing number [1], the bridge number [265], the knot group [82], the Jones polynomial [164], the Alexander polynomial [7], the Khovanov homology [172], and the knot Floer homology [238].

Knot theory has been applied in various fields, such as physics [237], chemistry [189], and biology [223, 264, 285], but its success in practical applications has been somewhat limited. One of the challenges lies in the fact that classical tools in knot theory are inherently global and topological, offering little insight into local geometric and topological features. This makes these tools insufficient for analyzing biological structures, such as DNA and proteins, which often carry intricate local structures and/or local entanglement between segments. In addition, many datasets encountered in real applications consist of open curves rather than closed loops, which makes classical knot invariants inapplicable. These challenges highlight the need to develop more effective tools for analyzing curves in practical settings.

Motivated by the success of topological data analysis for various domains, the concept of knot data analysis (KDA) was introduced in [270]. The paper proposes a framework called multiscale Gauss linking integral (mGLI), which is the first to combine multiscale analysis with tools in knot theory. Following the work, several other geometric topological frameworks have been proposed to address the aforementioned limitations of classical methods in analyzing curves in 3-space. These include a localized Jones polynomial [277], persistent Khovanov homology [268], and Khovanov Laplacians, and Khovanov Dirac operator [162].

In the following, we explore these recently developed geometric topology approaches. The multiscale Gauss linking integral (mGLI) [270] and the localized Jones polynomials [277] build upon the Gauss linking integral [129] and the Jones polynomial of open-ended knot diagrams [239], respectively, which are currently the only two known continuous measures of single or pairwise topological complexity for open curves in 3-space. These approaches enable

the analysis of local and global entanglement properties across multiple spatial scales. The persistent Khovanov homology [268] extends the Khovanov homology by introducing filtrations of link diagrams based on smoothing transformations, revealing finer topological and geometric features that are often overlooked by traditional invariants. The Khovanov Laplacians and Dirac operator [162] extend the Khovanov homology by incorporating spectral information, offering new descriptors that recover known invariants from homology while encoding additional non-harmonic information.

4.1 Multiscale Gauss linking integral

Shen et al. [270] proposed the multiscale Gauss linking integral (mGLI), offering a multiscale approach for analyzing the local structure, connectivity, and entanglement of segments in both open and closed curves. These local features are crucial for understanding the physical properties and functions of data, such as chemical compounds and biological molecules. Although local, the method can also recover the global topological features when sufficiently large scales are applied. The framework builds on the Gauss linking integral, which is a double line integral that provides the linking number of two curves and has been successfully applied in understanding polymers [241] and proteins [18, 19, 240].

The method requires curve segmentation to partition curves into small segments for a given pair of curves. The pairwise Gauss linking integrals between these curve segments then result in a segmentation matrix whose entries encode the pairwise linking information. For multiscale analysis, weights on these entries, defined by indicator functions in a specific distance range from each segment, are then incorporated. By varying the distance range, this process generates a collection of segmentation matrices that capture the local entanglement across various spatial scales. The cumulative integrals across scales also provide additional insights into local structures, and reveal the global relationships at sufficiently large scales. For each segment in one curve, the local linking information with all segments of the other curve, calculated within each distance range, can be used to generate the multiscale feature of that segment through statistical methods. The construction can also be applied to a single curve by considering the local linking information between segments of the same curve.

The segmentation of curves and the distance range are crucial in determining the effectiveness of Gauss linking integrals for feature extraction. Therefore, the segmentation strategy and the distance range should be chosen carefully based on specific applications. The framework has been integrated with machine learning models to address a variety of complex biological problems, including protein *B*-factor prediction, protein-ligand binding affinity prediction, and others. The model consistently outperforms the traditional TDA models and demonstrates superior performance in certain tasks [270].

By combining the classical Gauss linking integral with multiscale analysis, mGLI offers a powerful framework for analyzing the local and global structures

of curve-type data, opening a new area in data analysis and knot learning. The framework addresses the limitations of traditional TDA techniques by examining the local entanglement of curve segments, and it can be applied to open curves. In addition, it is computationally efficient, requiring only a few minutes on a personal computer to generate the features, which makes it practical for analyzing large datasets. The method shows promise and holds strong potential for practical applications across various fields.

4.2 Localized Jones polynomials

The localized Jones polynomials were introduced by Song et al. [277] for analyzing the local and global entanglement properties of collections of disjoint open or closed curves. The paper proposes two models, termed the multiscale Jones polynomial and the persistent Jones polynomial, which extend the Jones polynomial proposed in [20] by incorporating spatial multiscale analysis. The approach in [20] generalizes the traditional Jones polynomial [164] for closed curves and the one in [239] for an open curve, offering a continuous measure of linking complexity for collections of open curves in 3-space.

The stability of both proposed localized Jones polynomials has been analyzed in the same paper, ensuring their robustness for practical use. These models have also been applied to protein structure analysis, providing insight into the entanglement that goes beyond atomic positions alone. The frameworks demonstrate potential to benefit a range of practical applications.

Multiscale Jones polynomial

The multiscale Jones polynomial [277] follows a similar strategy used for the multiscale Gauss linking integral [270]. Given a collection of disjoint open or closed curves with a segmentation, the method computes, for each curve segment, the Jones polynomial of the set consisting of that curve segment together with all curve segments that lie within a specified distance range from that curve segment. By varying the distance interval, the process generates a vector of Jones polynomials for each segment, which captures the topological entanglement at multiple spacial scales. By considering all curve segments, the method finally leads to a matrix of Jones polynomials, encoding both the local and global entanglement properties for the collection of curves. These Jones polynomials are then evaluated at a fixed parameter for real applications, resulting in a real characteristic matrix suitable for downstream analysis. Analogous to the multiscale Gauss linking integral [270], the featurization depends on the choice of segmentation of the collections of curves, and the selection of distance ranges, both of which can be tuned appropriately for specific applications.

Persistent Jones polynomial

The persistent Jones polynomial was proposed by considering the pairwise distance matrix of curve segments for a collection of disjoint open or closed curves with a segmentation, which then induces a filtration of Vietoris-Rips complexes

(other types of simplicial complexes may also be used), with the vertices corresponding to the curve segments [277]. Each facet in the filtration is assigned a weight given by the Jones polynomial of the set of curve segments forming that facet, leading to a weighted filtration of polynomials of the segmentation. This filtration gives rise to weighted barcodes, which encode the birth and death of facets along with their associated polynomial weights. For practical applications, the Jones polynomials are evaluated at a fixed parameter value. By introducing a filtration, the persistent Jones polynomial encodes the local and global interactions among curve segments within a segmentation. In contrast to the persistent homology for point cloud data, which focuses on the behavior of holes, the persistent Jones polynomial captures the entanglement of curves in 3-space.

4.3 Persistent Khovanov homology

Khovanov homology [172] is one of the most significant advances in knot theory, providing a powerful link invariant that categorifies the Jones polynomial [164]. It reveals richer information such as torsion. It can also detect the unknot, which is a very hard problem in general, while it remains an open question whether the Jones polynomial can do so. To incorporate the persistence into the analysis of curve-type data, Shen et al. [268] proposed the persistent Khovanov homology (or evolutionary Khovanov homology) of knots and links by defining filtrations of links obtained through the smoothing transformations of crossings in a link diagram. This work addresses the limitations that classical knot theory primarily focuses on global topological properties and may overlook the important local features of curve-type data. It enables a more refined characterization of the topological structures and geometric shapes of knots and links across various scales in a filtration.

Conceptually, Khovanov homology for knots and links shares similarities with the classical homology for simplicial complexes, although they are developed in different contexts. Both are constructed from chain complexes, which consist of sequences of abelian groups connected by differentials or boundary maps that vanish when applied twice. A key difference lies in the grading, where Khovanov homology is bigraded, with both a homological grading and a quantum grading, whereas the simplicial homology has only a single homological grading. The graded Betti numbers in Khovanov homology reflect the graded dimensions, compared to the classical Betti numbers in simplicial homology, which count the number of holes in each dimension.

The paper presents two approaches for constructing a filtration of link diagrams: distance-based filtration and unzipping filtration. The idea is to generate an order of crossings, which are gradually resolved by a specific smoothing, finally leading to a simple link diagram consisting of disjoint circles. This process provides a filtration. The distance-based filtration considers disks centered at all crossings determined by one single radius parameter, and a crossing is isolated if its disk does not intersect with any other disks. Starting with a large initial radius, the order is then obtained by labeling isolated

crossings as the radius gradually decreases. For the unzipping filtration, the order is chosen by starting with an initial crossing and labeling the crossings of the link by moving in a specific direction along the curve.

The persistent Khovanov can capture nontrivial evolutionary information even for trivial knots and unknotted links. The effectiveness of the methods has been demonstrated in the analysis of the SARS-CoV-2 frameshifting pseudo-knot in the same paper. Later, Liu et al. [197] proposed theoretically the persistent Khovanov homology of tangles, focusing on characterizing local topological characteristics in curve-type data. By enabling a persistence analysis of topological invariants across multiple scales, these frameworks provide powerful tools capable of capturing finer topological details or local structures of curve-type data, offering promising geometric topology approaches in data science.

4.4 Khovanov Laplacians and Khovanov Dirac

Motivated by the success of combinatorial and Hodge Laplacians and Dirac in data science, and their connections to the (co)homology, Jones and Wei [162] introduced the Khovanov Laplacians and Khovanov Dirac for knots and links, extending the Khovanov homology by incorporating the spectral analysis in the study of knots and links. The harmonic spectra of these operators can recover the topological invariants from Khovanov homology, while the non-harmonic spectra reveal additional information not accessible through homology alone. The framework can also be seen as an extension of the Laplacians on directed graphs associated with link diagrams [273], which establishes a connection between graph Laplacians and key properties of knots. These graph Laplacians naturally lead to the Seifert matrix, the Alexander matrix, and the Goeritz matrix of a link, along with the invariants derived from them, including the Alexander polynomial and ω -signature.

The construction of the Khovanov Laplacians and Khovanov Dirac parallels that of the combinatorial Laplacians and the Dirac operator on simplicial complexes. As the differential in Khovanov homology increases only the homological grading by 1 while preserving the quantum grading, one can just restrict to chain complexes with fixed quantum grading. The paper defines an inner product on a chain group by assuming that its basis elements are orthonormal, which then induces a unique adjoint of the differential. The Khovanov Laplacians are subsequently defined using the same formula as for the combinatorial Laplacians, and the Dirac operator follows the same matrix structure. Analogous to the combinatorial case, the kernels of Khovanov Laplacians correspond to the Khovanov homology groups.

As noted in the paper [162], the construction of Khovanov Laplacians and the Dirac operator may also be extended to other knot homology theories, such as the knot Floer homology [238], which categorifies the Alexander polynomial. This raises possible connections between the Laplacians derived from the Floer knot homology and the Laplacians mentioned above on directed graphs [273]. In addition, in analogy with the combinatorial and differential

manifold settings, one may define the persistent Khovanov Laplacians and the persistent Dirac, further generalizing the framework of persistent Khovanov homology [268].

5 Machine learning featurization

Simplicial complexes are the most widely used tools for modeling point cloud data in topological data analysis, owing to their conceptual simplicity and computational efficiency. However, they are not universally optimal for all data types. For example, in network data, a coauthorship network is more appropriately modeled as a hypergraph rather than a simplicial complex. This is because the presence of a joint publication among three authors does not necessarily imply that each pair of authors has collaborated independently, a condition inherent in the simplicial complex construction. Similarly, data on manifolds such as images often arise from smooth structures and therefore possess intrinsic differential properties and boundary information. These characteristics cannot be fully captured by the purely combinatorial structure of simplicial complexes. Instead, modeling the data as differential manifolds provides a more suitable framework for incorporating smoothness and boundary information. Thus, the selection of a mathematical structure for modeling data in TDA should be guided by the intrinsic nature of the data and the objectives of the specific analysis task. Moreover, different modeling choices naturally lead to different classes of topological tools and representations.

In this section, we begin by categorizing various data types commonly encountered in TDA, including point-cloud data, distance matrix data, network data, data on manifolds, sequential data, knots and links data, and data with additional non-geometric information. For each category, we discuss appropriate mathematical models and their implications. We then present methods for extracting topological representations from these data structures. Finally, we describe approaches for transforming these topological representations into features suitable for machine learning applications. We mainly focus on vectorization techniques, which convert topological representations into fixed-size feature vectors. Other strategies, such as kernel-based methods, are also briefly discussed.

5.1 Input data types and the choice of topological modeling

5.1.1 Point-cloud data

Point-cloud data is one of the most commonly encountered data formats in a wide range of fields. Examples include atoms in biomolecules such as proteins, DNA, and RNA, as well as nodes in various networks such as social networks, transportation networks, and molecular interaction networks. Modeling point cloud data using mathematical objects such as digraphs, simplicial complexes, and hypergraphs enables the application of advanced topological

tools, including persistent homology and persistent combinatorial Laplacians, for comprehensive data analysis. The selection of an appropriate mathematical model for point cloud data depends on the nature and available additional information of the data.

The Vietoris-Rips complex and Alpha complex are suitable when data points are given without any additional information or when they need to be treated equally, while the weighted simplicial complex or sheaf is better suited for data points that carry additional label information, such as type, mass, or color. The digraphs or hyperdigraphs are appropriate for representing data with asymmetry or directed relations, such as gene expression datasets, where directional information encodes gene regulation. Hypergraphs and super-hypergraphs can be effective for handling data points with complex relationships or incomplete structures.

5.1.2 Distance matrix data

For distance matrix data, it is important to emphasize the flexibility and diversity in the construction of distance metrics. While the Euclidean distance is the most widely used, it is not universally optimal for all applications. In many cases, modifying the distance metric or adopting alternative definitions can lead to a more accurate extraction of topological features relevant to the underlying problem.

The interactive distance [38] is designed to capture the interactions between two distinct types of objects. Taking protein-ligand interactions as an example, directly applying Euclidean distance to construct simplicial complexes from the entire protein-ligand structure can obscure the biologically meaningful interactions due to overwhelming contributions from intra-protein or intra-ligand interactions. To address this, the interactive distance assigns the original Euclidean distance to point pairs that originate from different structures (i.e., protein and ligand), while assigning an infinite distance to point pairs within the same structure. This strategy ensures that the resulting simplicial complex focuses specifically on cross-structure interactions, thereby offering a more targeted characterization of the protein-ligand interactions.

The multi-level distance [40] is proposed to highlight noncovalent atomic interactions, such as hydrogen bonds and van der Waals forces. In this approach, bonded atom pairs are assigned infinite distances since bonded atoms stay closer than non-bonded ones in most cases, allowing us to focus on non-bonded, noncovalent interactions. Although this metric may violate the triangle inequality, it remains valid for constructing Vietoris-Rips complexes, where only pairwise distances are required. This strategy leads to not only zero-dimensional topological features encoding noncovalent interactions but also higher-dimensional features that capture small structural fluctuations among different conformations of the same molecule.

The kernel function-based distance [40] was introduced to model atomic interaction strengths, as they often do not align linearly with the Euclidean distance. The frameworks uses the negative of a radial basis function to define

a correlation function-based filtration matrix. Stronger interactions correspond to smaller distances and appear earlier in the filtration process. Similarly, in the flexibility rigidity index-based distance, a Lorentz-type kernel with a tunable scale parameter is employed. By adjusting this scale parameter, one can capture structural patterns at different spatial resolutions.

The KNN-based distance strategy [182] assigns distances between pairs of points based on their mutual nearest-neighbor relationships. Specifically, for a pair of points (a, b) , if a is the s -th nearest neighbor of b , and b is the t -th nearest neighbor of a , then the distance between a and b is defined as the smaller of the two values, i.e., $\min(s, t)$. The standard Vietoris-Rips or Alpha filtration can then be constructed from the resulting distance matrix of the point cloud. This strategy has been used to analyze single-cell RNA sequencing data [80].

5.1.3 Network data

Network data consists of nodes and edges that represent interactions or relationships among entities. Such data arises in a wide range of domains. For example, in molecular networks, nodes correspond to atoms and edges to chemical bonds, in social networks, nodes represent individuals and edges denote social connections, and in trade networks, nodes are countries while edges indicate trading relationships. Mathematically, these networks can be modeled as graphs, enabling the application of tools from graph theory for pairwise interaction analysis. The choice of topological tools depends on the nature of the network and the specific application context.

The (directed) undirected network can be naturally modeled as a one-dimensional (ordered) simplicial complex, where nodes represent 0-simplices and edges represent 1-simplices. To analyze high-order interactions beyond pairwise connections, one can extend the graph to a higher-dimensional simplicial complex. Several constructions are available: the flag complex (or clique complex) [163], which includes a simplex for each complete subgraph, the neighborhood complex [200], which encodes shared neighborhood structures, and the Hom complex [204], which can be interpreted analogously to convolutional kernels in convolutional neural networks. In addition, hypergraphs and other generalized mathematical objects can be employed to model high-order interactions directly [136, 206].

5.1.4 Data on manifolds

Common examples for data on manifolds include images, such as 3D volumes or 2D slices from MRI and CT scans, as well as isosurfaces, such as isothermal surfaces representing a uniform heat distribution in fluid dynamics and electron density isosurfaces modeling molecular orbitals in molecular visualization. Although they are discrete in their raw form, it is natural to treat them as data on manifolds, as they are often sampled from continuous or smooth structures. In previous studies, topological analysis of such data, particularly image data, has often been conducted using point cloud-based

techniques. For example, a common approach in medical imaging is to select points from the region of interest in the image and then construct a simplicial complex from these selected points to perform topological analysis [275]. Additionally, cubical complex structures have also been employed to model images for topological analysis [299]. Mathematically, both simplicial complexes and cubical complexes are combinatorial objects, so they are capable of capturing the topological structures arising from the data on manifolds. However, due to their inherently combinatorial nature, these representations neglect the differential structure and boundary conditions of the underlying manifolds, thereby limiting their ability to preserve the differential properties intrinsic to such data.

Therefore, modeling these data as differential manifolds offers a more suitable framework for capturing their intrinsic differential properties. In this setting, tools from differential topology or differential geometry, such as differential forms, vector fields, and differential operators, can be employed to facilitate quantitative analysis of the data. In addition, these data, such as images, inherently have boundaries that separate objects from the background, making the boundary conditions essential. These boundary conditions not only impose constraints that govern the behavior of the system near the boundary but also ensure the existence and uniqueness of solutions in boundary value problems. Moreover, the boundary conditions modify the differential operators, such as the Laplacian operators and the Dirac operator, making them different from their counterparts defined on closed manifolds (without boundary). It is worth noting that the Hodge Laplacians on differential manifolds differ fundamentally from the combinatorial Laplacians defined on simplicial complexes [260]. The effectiveness of differential topology-based methods has been demonstrated in various applications, such as Hodge decomposition in single-cell RNA velocity analysis [281], PHL in drug design and discovery [282], and MTDL [207] in medical image analysis.

Building on this perspective, recent work has integrated topological and geometric priors into deep learning architectures. The Topological Convolutional Neural Network (TCNN) introduced by Love et al. [209] replaces standard convolutional layers with manifold-parameterized filters that reflect the structure of image and video data. This approach preserves topological relationships and leverages geometric localization across layers, leading to faster learning, better generalization, and improved interpretability. TCNNs also extend naturally to 3D data, highlighting the broader potential of topology in deep learning.

5.1.5 Sequential data

Sequential data arise in a wide range of fields, such as protein and DNA sequences in biology, word sequences in natural language processing, temporal stock price fluctuations in time series analysis, and frame sequences in video processing. Many approaches have been developed for topological sequence

analysis (TSA). One strategy for TSA is to directly compute a pairwise distance matrix [54], which can then be used to construct Vietoris-Rips filtration by varying the distance parameter. For time series data, an approach involves dividing curves into short segments and building a Vietoris-Rips filtration for each segment [326]. There are also TSA approaches to transform the sequence into a point cloud, thereby enabling the direct use of established techniques for point cloud data. These strategies include Takens' delay embedding, the sliding window method, and the k -mer topology approach.

In Takens' delay embedding [288], the sequence x_1, x_2, \dots is transformed into points using two parameters: the embedding dimension d and the delay parameter τ . For instance, with $d = 2$, each point in the resulting point cloud takes the form $(x_i, x_{i+\tau})$. The sliding window method similarly employs two parameters: the point dimension parameter M and the window parameter τ . For example, if $M = 3$, a point in the embedded space is constructed as $(x_i, x_{i+\tau}, x_{i+2\tau})$. There are also other variations for the sliding window method, such as SW1PerS [246] and toroidal sliding window embeddings [245].

In the k -mer topology method [157], the parameter k specifies the window size used to slide along the sequence, generating a collection of k -mers (subsequences of length k), which are treated as a point cloud. The element-specific strategy is then applied to emphasize particular types of points. The pairwise distances between these points are computed to construct a distance matrix, which serves as the foundation for subsequent topological analysis using methods such as persistent homology or persistent Laplacians.

Delta complex approaches have been proposed for TSA [198]. The essential idea is to consider the segments of a sequence as simplices, which serve as the building blocks of a Δ -complex. Since Δ -complexes allow repeated elements, they align with the nature of biological sequences. Unlike k -mer topology, Δ -complex-based TSA methods do not systematically analyze k -mers in a sequence. These approaches eventually utilize persistent homology or persistent Laplacians and are faster than the early k -mer topology method for large genome sequences.

Recent developments in TDL introduce novel methods for integrating temporal and topological structures in sequential data, particularly in the context of neural spike decoding [226]. These approaches enable end-to-end learning where topology is a learnable model component [170, 243]. In many cases, the data consists of temporally ordered events, such as neural spikes, where each observation contributes to the construction of a simplicial complex that captures multi-way relationships among entities (e.g., neurons spiking together within a short time window). These complexes encode higher-order interactions beyond simple pairwise connections, providing a richer structural representation.

Rather than focusing on individual time points or fixed-size windows, some TDL methods track the temporal evolution of these simplicial complexes, producing a sequence of topological states. In event-based settings like spike decoding, this evolution is typically modeled through discrete-time

or state-space processes. Other approaches, [117, 228], model the evolution of topological features over complexes using continuous-time dynamics and partial differential equations. In these cases, each simplicial complex can be embedded as a learnable vector using neural architectures designed for topological spaces, such as simplicial message passing networks [30] or cell complex encoders.

The result is a unified, data-driven framework that captures both temporal dynamics and topological structure. In applications like neural decoding, this enables the model to learn how patterns of co-activation change over time and relate to behavior, stimulus, or latent neural states. By representing sequences as evolving topological objects, TDL provides a flexible and expressive way to analyze structured time-series data, enabling the discovery of complex patterns embedded in both time and topology.

5.1.6 Knots, links, tangles, and braids

Knots, links, tangles, and braids are fundamental topological objects that can be naturally modeled as curves embedded in 3-space [268, 270, 277]. In practical applications, such data structures are commonly found in a wide range of fields. Examples include ropes, shoelaces, polymers, DNA, RNA, nucleosomes, and chromosomes.

The representation of these objects as curves in 3-space has advantages, as it allows the use of tools from knot theory for analyzing complex structure relationships. The approach has proven effective in various studies. For example, the entanglement of polymer chains can be studied through their knotting and linking properties, with the Gauss linking integral correlating with the physical properties of polymeric material [241]. In addition, studies have shown that the Gauss linking integral is related to protein entanglements, particularly in understanding protein folding kinetics [240]. It has also been applied to the study of SARS-CoV-2 spike protein [19]. The effectiveness of the knot-based topological frameworks has also been demonstrated in protein flexibility analysis and protein-ligand interactions [270], protein structure and functions [277], and SARS-CoV-2 frame-shifting pseudo-knot [268].

Databases such as KnotProt [160] and LinkProt [85] have been developed to support the study of knots and links of proteins. The Python package Topoly [86] provides tools for calculating topological polynomial invariants to distinguish knots, slipknots, links, and spatial graphs.

5.1.7 Data with additional non-geometric information

For data with additional non-geometric information, we refer to all the aforementioned data types that are further enriched with background-related attributes beyond geometric structures. We set this as a distinct data type because such data is common in real-world applications, and the supplementary information often plays a crucial role in addressing domain-specific problems. For example, the protein structures not only contain the 3D coordinates of atoms (geometric information) but also include the residue types,

partial charges, and solvent accessibility (non-geometric biological information), which are critical for protein stability and protein-ligand interactions. Similarly, in network data, beyond the topological structure of nodes and edges, one may have node-associated attributes such as the user interests in social networks, which are critical for the clustering and link prediction tasks.

The element-specific method [37, 38, 40] is a powerful framework for systematically modeling different atomic and molecular interactions into a unified topological system. For example, given a point cloud of molecular atoms, one can construct element-specific simplicial complexes by grouping atoms based on their element types. This enables us to focus on particular atomic interactions, such as covalent bonds, hydrogen bonds, van der Waals forces, and other noncovalent interactions. The integration of element-specific strategy with standard persistent homology has demonstrated great promise in various biological data analysis tasks, such as protein-ligand binding affinity prediction, protein-protein interaction analysis, and protein stability analysis.

Another approach for incorporating non-geometric information is to encode it as weights and integrate these weights into the underlying mathematical frameworks. For example, in the enriched barcode method [39], the persistent cohomology is computed from the underlying simplicial complex modeling of the protein atoms, and the additional biochemical information is embedded as weights into the cohomology representatives. This results in an enriched, or colored, persistence barcode that captures both topological and chemical properties. In the persistent sheaf Laplacian method [310], the atomic properties are encoded into the sheaf construction, particularly, the construction of the boundary operators. Thus the derived persistent sheaf cohomology and sheaf Laplacians will automatically carry this additional information.

5.2 Approaches for topological representations

5.2.1 Representations of persistent homology/cohomology

Persistent homology and persistent cohomology produce identical barcodes when computed with coefficients in a field [90], which is the most common choice in practical applications. Therefore, here we focus only on the case of persistent homology for simplicity. Once an appropriate topological model is established for a given data set, the corresponding persistent homology can be computed to extract meaningful topological representations, such as persistent simplicial homology for simplicial complexes, persistent hypergraph homology for hypergraphs, and persistent sheaf homology for sheaves. Note that persistent homology can be visualized and encoded in various forms, each offering a distinct perspective on the topological features. Next, we summarize several common representation formats of persistent homology.

Persistence barcode/diagram

The two standard representations of persistent homology, persistence barcodes [48] and persistence diagrams [112], both offer intuitive ways to represent the

birth and death of topological features during the filtration process. Persistence barcodes summarize the lifespans of topological features, such as connected components, holes, and cavities, across different dimensions. Each barcode consists of horizontal line segments, where the left endpoint of each line segment represents the birth time and the right endpoint indicates the death time of a topological feature. A persistence diagram is equivalent to a barcode but encodes the same topological information as a 2D scatter plot, where the x -coordinate represents the birth time and the y -coordinate represents the death time. As the death time of a feature always occurs after its birth time, all points in a persistence diagram are positioned above the diagonal $y = x$. Long segments in a barcode typically correspond to significant topological features, while short segments are often considered as noise. Similarly, in a persistence diagram, features far from the diagonal are usually considered significant, and those closer to the diagonal are often treated as noise. However, this assumption might not hold when considering physical, chemical and biological data, where short segments or features near the diagonal might carry important physical information such as bond lengths, benzene ring size, and other properties [313]. There is also a Grassmannian persistence diagram construction [141, 142] for multiparameter persistence. Persistence barcodes/diagrams can effectively capture the topological features of the data, and have been used in the study of image patches [2, 49], hand-drawn letters [75], and unusual events in viral evolution [54].

Persistent Betti numbers/ Betti curves

Persistent Betti numbers [112] and Betti curves are multiscale analogs of the Betti numbers that count the number of topological features that persist in the filtration. Persistent Betti numbers are defined as the ranks of persistent homology groups and are often visualized as Betti curves, which offer a step function-based representation of persistent homology. The representation transforms persistence barcodes/diagrams into functional representations. The vector space structure of functions then enables basic operations like addition and scalar multiplication, making Betti curves suitable for statistical tasks such as computing the means and variances. These operators are not naturally defined for persistence barcodes/diagrams.

Betti curves are simple and easy to compute. In addition, as small perturbations often introduce short lifetimes that have minimal impact on the overall Betti number distribution, they are also robust to noise. However, one drawback of Betti curves is their limited expression power, as they only contain counts of topological features. In addition, the mapping from a diagram to a curve is not injective [68], which means that different diagrams can lead to the same curve.

Persistence landscape

A stable representation of persistent homology, named persistence landscapes [34, 35], offers a more structured and expressive representation of persistence

diagrams while maintaining compatibility with statistical and machine learning frameworks. They were introduced to convert persistence diagrams into sequences of piecewise linear functions, which can be viewed as a horizontal version of persistence diagrams. Compared to Betti curves, which track only the count of topological features, persistence landscapes encode both the birth-and-death pairs and their persistence and maintain a bijective correspondence with persistence diagrams, making the latter preferable when detailed topological analysis is needed. Recently, it has been extended to the multiparameter persistence landscape (MPPL) [296] for the multiparameter case.

A key feature of persistence landscapes is their hierarchical structure, where different levels of functions capture topological features with varying significance. The function with the lowest lifespan subscript captures the most persistent topological features, while those with higher lifespan-subscript capture less persistent features. Persistence landscapes have been used to study conformational changes in protein binding[178], Microstructure Analysis [109], time-dependent functional networks constructed from time series [279], and image classification tasks and orbit classifications [173].

Persistence image

Persistence images [4] were introduced as an alternative stable representation that transforms persistence diagrams into fixed-size, grid-based functional representations, making them particularly well suited for convolutional neural networks (CNN). Persistence images are constructed by computing the integral of a persistence surface on a grid obtained by discretizing a subdomain of \mathbb{R}^2 . Here, a persistence surface is defined as a weighted sum of Gaussian functions centered at each point in the persistence diagram, with the weighting function controlling the significance of each point. Persistence images have been proven to be stable with respect to the 1-Wasserstein distance between persistence diagrams in the same paper.

A key advantage of persistence images is that they are easy to compute and offer flexibility in three choices: the resolution, the distribution (and its associated parameters), and the weighting function. However, these choices are noncanonical, meaning that there are no guidelines for setting these parameters to achieve the best performance in machine learning tasks. PI-Net [276], a simple one-step differentiable architecture, is the first to propose the use of deep learning for computing persistence images directly from data. The persistence image has also been extended to a multiparameter setting, leading to the Multiparameter Persistence Image (MPPI) [52].

Persistent Betti number image

The notion of multidimensional persistence was proposed in [314] for transforming dynamical biomolecular data into persistent Betti number images by encoding the topological evolution of molecular systems over time. Specifically, given a sequence of data frames representing a dynamical process, persistent homology is applied to each frame, and the resulting persistent Betti numbers

are encoded as one-dimensional vectors. By stacking these vectors across all time steps, a two-dimensional image is constructed that reflects the temporal topological changes of the system. The multiscale multidimensional persistence proposed in the same paper further encodes molecular structures at various spatial scales within a single image. In particular, the flexibility-rigidity function with tunable scale parameters is used to generate a volumetric density representation of the molecule. For specific chosen scale parameters, the persistent Betti numbers are computed and recorded as one-dimensional vectors, then a 2D, 3D, or high-dimensional image is formed by assembling these vectors across all scales.

Colored persistence barcode

Several methods have been proposed to enrich persistence barcodes with additional information, resulting in the colored persistence barcode. For example, in the persistent enriched barcode approach [39], persistent cohomology is first computed, and supplementary weight information associated with simplices is incorporated into the cohomology representatives through specific functions to assign colors to the corresponding bars. In the persistent Jones polynomial method [277], each simplex is assigned a Jones polynomial, and the polynomial associated with the positive simplex corresponding to a bar is evaluated at a selected value to determine its color. These enhanced barcodes effectively encode background-related information, thereby providing a more informative topological characterization.

Graded persistence barcode

The most commonly used persistence barcode is derived from single-graded persistent homology. However, there are also methods that provide multi-graded persistent homology representations. For example, in the persistent Mayer homology method [269], the additional grading in the N -chain complex determines a specific type of boundary operator. By varying the grading index, a series of chain complexes can be constructed, resulting in a sequence of persistent homologies, i.e., a bigraded persistent homology. Similarly, in the persistent Khovanov homology method [197], an additional quantum grading that is determined by the smoothing type of the crossings, is introduced into the cochain complex, resulting in a bigraded persistent homology that captures richer structural information. These graded persistent homologies naturally give rise to graded persistence barcode representations.

5.2.2 Representations of persistent Laplacians

The representation of Laplacians primarily involves their spectral properties, that is, their eigenvalues and eigenvectors, which encode essential information about the topological, and the combinatorial or geometric structures of the underlying datasets. Another important aspect is that the Hodge Laplacian enables the decomposition of vector fields into three or five orthogonal components through the Hodge decomposition, providing a powerful tool for

analyzing flows, gradients, and cycles in data. In the following, we describe these representations in detail.

Eigenvalue representation

A fundamental property of topological Laplacians is the isomorphism between their kernel and the corresponding cohomology groups. In practice, computations are typically performed with a field coefficient, under which cohomology is isomorphic to homology. This implies that the multiplicity of zero eigenvalues of the topological Laplacian is equal to the Betti number in the corresponding dimension. This result extends to the persistent setting: the persistent Laplacian exhibits a similar isomorphism, where the multiplicity of its zero eigenvalues corresponds to the persistent Betti numbers [217]. Consequently, the zero eigenvalues of the persistent Laplacian fully encode the persistent homology information. In contrast, the nonzero eigenvalues provide additional geometric and combinatorial insights of the underlying data structure. Laplacian matrices are symmetric and semi-positive definite, which guarantees that all nonzero eigenvalues are positive. For Hodge Laplacians on manifolds, the larger eigenvalues are associated with higher-frequency components of the corresponding eigenfunctions. These high-frequency modes reflect fine-scale structural variations or local irregularities and may represent noise, especially when the manifold exhibits non-smooth features such as sharp corners. The integration of the eigenvalues of Laplacians with machine learning techniques has been successfully applied in a variety of domains, including drug design and discovery, computational biology and chemistry, and material science.

Eigenvector representation

It is known that the eigenvectors associated with topological Laplacians capture rich topological and geometric information about the underlying objects. For example, in shape analysis, the global point signature (GPS) [262], a representation of surfaces defined using scaled eigenfunctions of the Laplace-Beltrami operator, has been shown to carry sufficient information of surfaces for various shape processing tasks, including shape classification, segmentation, and correspondence. The eigenvectors of Hodge Laplacians, including the harmonic eigenfields, the gradient eigenfields, and the curl eigenfields under certain boundary conditions, have been applied to the analysis of biological macromolecules [325]. These eigenvectors can be used either individually as inputs in machine learning models, or as a supplement to eigenvalue features.

For persistent topological Laplacians, while the eigenvalues are known to capture the topological and geometric changes during the filtration process, the interpretation of the eigenvectors, however, remains unclear and requires further investigation.

Hodge decomposition representation

Hodge decomposition orthogonally splits vector fields into gradient, curl, and harmonic components. This decomposition enables a more detailed analysis of

a vector field by extracting different dynamic features, making it particularly useful in understanding the underlying mechanisms of a field. It has numerous applications in electromagnetism [222], fluids and dynamic systems [318, 320], geometric modeling [304], and spectral data analysis [171].

The most well-known Hodge decomposition is the 3-component Helmholtz-Hodge decomposition [181]. For compact domains in Euclidean spaces, there is a refined 5-component Hodge decomposition of vector fields [272], which preserves the topology of the underlying domains. The discrete 5-component Hodge decomposition has been proposed for surface triangle meshes [249], for tetrahedral meshes [248, 324], and for domains defined by level set functions in Cartesian grids [281, 283]. Hodge decomposition has been successfully applied to the study of single-cell RNA velocities for extracting different dynamic information of cells by utilizing their gene expressions [280]. In addition, manifold topological deep learning (MTDL) [207] has been recently proposed based on the Hodge decomposition technique for medical image analysis.

5.2.3 Representations of persistent Dirac operator

In the combinatorial case, the persistent Dirac operator can be seen as a generalization of the persistent combinatorial Laplacians [9, 308]. It captures the same topological information as the persistent combinatorial Laplacians, with its eigenvalues corresponding to the square root of the eigenvalues of the persistent combinatorial Laplacians in all dimensions. However, as a first-order operator, it is more sensitive to local features compared to the second-order Laplacians, and its matrix couples different topological dimensions.

One significant advantage of the persistent Dirac operator is that it enables efficient quantum computation, providing exponential speed-up over classical TDA algorithms. This operator not only recovers the topological outputs from persistent homology but also leads to spectral features of the point cloud data. The persistent Dirac operator can also be formulated similarly for differential manifolds, extending the Hodge-Dirac operator in the differential topology setting, thus facilitating the analysis of data on manifolds.

5.2.4 Representations of multiscale Gauss linking integral

The multiscale Gauss linking integral (mGLI) method [270] is developed for the analysis of one-dimensional curve data, including knots, links, and tangles. The method begins by decomposing the curve into a collection of short segments, after which the pairwise Gauss linking integrals are computed between all segment pairs. These integrals form a segmentation matrix whose entries quantify the topological linking information between curve segments. Then pairwise distances between segments are incorporated as weights into the entries, yielding a refined segmentation matrix that encodes linking structures across multiple spatial scales. By systematically varying the scale parameter, a sequence of such matrices can be constructed, each capturing localized topological linking information at a distinct resolution. Collectively, these matrices provide

a comprehensive multiscale topological representation of the underlying curve structure.

5.2.5 Representations of multiscale/persistent Jones polynomial

The multiscale and persistent Jones polynomials [277] extend the classical Jones polynomial [20] to a multiscale setting, enabling the analysis of local and global entanglements for collections of disjoint open and closed curves. Given a specific curve segmentation, the multiscale Jones polynomial framework offers a matrix of Jones polynomials, where each row corresponds to a feature vector of a curve segment determined by the Jones polynomials computed across various distance ranges. Each entry of the feature vector is the Jones polynomial of the set of curves consisting of that curve segment and all other segments that fall within a specific distance range from that segment. The persistent Jones polynomial builds on a Vietoris-Rips complex constructed from the pairwise distance matrix of curve segments, which leads to the birth, death, and persistence information of facets in the complex. By incorporating the Jones polynomials of facets, this framework leads to a weighted barcode, with weights given by the Jones polynomials of the associated facets.

5.2.6 Software and code resources

Table 1 lists several widely used software packages for computing persistent homology, persistent Laplacians, and Hodge decompositions. In addition to these, another popular tool is Mapper [274], which provides a simplified representation of high-dimensional datasets in the form of simplicial complexes. Taking point cloud data as an example, the data is first divided into clusters using a clustering algorithm. A simplicial complex is then constructed from these clusters: each cluster corresponds to a vertex, and a set of n clusters forms an $(n - 1)$ -simplex if their intersection is nonempty. The one-skeleton of the resulting simplicial complex is commonly referred to as the Mapper graph, which provides a simple description of the data and captures important information about its topological structures. Several software tools have implemented the Mapper algorithm, including KeplerMapper [294], giotto-tda [289], GUDHI [211], and Mapper Interactive [328].

5.3 Featurization methods

Persistent homology and persistent Jones polynomial can be encoded as persistence barcodes, which consist of a collection of (birth, death) pairs representing the lifespan of topological features across the filtration. In contrast, the multiscale Gauss linking integral and multiscale Jones polynomial are typically represented by segmentation matrices, where the entries qualify the local linking information between segments of the underlying data. For persistent Laplacians and persistent Dirac operator, the spectral information (eigenvalue and eigenvector) is usually used for feature extraction. These representations

Table 1 Software packages for computing persistent homology and persistent Laplacians.

Software	Simplicial Complex	Cubical Complex	Persistent Homology	Persistent Cohomology	Persistent Laplacian	Hodge Decom.
JavaPlex [3]	Yes		Yes			
Perseus [224]	Yes	Yes	Yes			
Dinoysus [229]	Yes		Yes			
GUDHI [211]	Yes	Yes	Yes	Yes		
Ripser [22]	Yes		Yes		Yes	
PHAT [24]	Yes	Yes	Yes			
DIPHA [23]	Yes	Yes	Yes			
R-TDA [121]	Yes	Yes	Yes	Yes		
Giotto-tda [289]	Yes	Yes	Yes			
OpenPH [218]	Yes		Yes			
CTL	Yes	Yes	Yes			
Eirene [152]	Yes	Yes	Yes			
Cubicle [297]		Yes	Yes			
CubicalRips [166]		Yes	Yes			
TTK [290]	Yes	Yes	Yes			
libstick [114]	Yes		Yes			
HERMES [303]	Yes				Yes	
PersistLap [217]	Yes				Yes	
HHD [280]		Yes				Yes
5ComponentHD [281]		Yes				Yes
3DHodgeDecom [324]	Yes					Yes

give rise to three types of topological descriptors: persistence barcode, persistence spectral, and segmentation matrix. In addition to these, the Hodge Laplacians enable three- and five-component orthogonal decompositions of vector fields, which naturally provide a multi-channel image representation of the original vector field in the Eulerian setting. In the following, we describe methods for transforming these topological descriptors into feature vectors suitable for machine learning algorithms.

5.3.1 Functional formulations

Persistence barcodes/diagrams can not be directly applied to machine learning tasks, as they are not naturally structured as fixed-size feature vectors. Persistence barcodes consist of varying numbers of intervals with different lengths, while persistent diagrams contain varying numbers of unordered points. One straightforward approach for vectorizing persistent barcodes/diagrams is to use their functional formulations, which can be directly discretized as machine learning inputs by sampling the resulting functions at selected points. Commonly used functional formulations include Betti curves, persistence landscape [34], persistence surfaces and persistence images [4], etc.

5.3.2 Counting in bins

Another method for vectorizing persistence barcode is the binning-based approach [40]. In this method, the uniform-length filtration range is divided into a collection of bins, and the numbers of births, deaths, and persistences within each bin are counted. This process yields three corresponding feature vectors: birth count, death count, and persistence count. This approach effectively captures structural properties such as bond lengths, including those

of noncovalent interactions in biomolecules, and has been widely adopted in various topological data analysis studies [37, 38, 233].

5.3.3 Statistical property

For persistence barcode, a basic and widely used approach for vectorization is to extract their statistical properties [8, 14, 40, 42]. For example, one can compute descriptive statistics of the bar lengths, birth times, death times, and midpoints, including the maximum, minimum, sum, median, mean, standard deviation, variance, and entropy. This yields a feature vector of dimension $4k$, where k denotes the number of statistical measures for each of the four quantities. A more fine-grained representation can be obtained by partitioning the filtration range into s fixed bins and computing statistical properties within each bin. For instance, counting the number of bars whose birth-death intervals cover each bin produces an s -dimensional vector, while recording the maximum and minimum birth values in each bin results in a $2s$ -dimensional vector.

For persistence spectra, a common practice is to divide the filtration range into uniform intervals and compute the corresponding Laplacian matrices at each interval, the eigenvalues and eigenvectors of these matrices are then extracted to form feature representations [202, 219]. Basic statistical measures, such as maximum, minimum, sum, median, mean, standard deviation, variance, and entropy, can be applied to the eigenvalues of each Laplacian matrix, yielding feature vectors that capture topological, geometric, and combinatorial characteristics across scales. However, when the Laplacian matrices are large, computing the full spectrum becomes computationally expensive. In such cases, it is common to retain only the smallest k nonzero eigenvalues, as well as the Betti numbers, which capture the most global and stable spectral components. This truncated spectral representation is not only computationally efficient but also robust to noise.

The segmentation matrix encodes the pairwise linking information between segments of a given pair of curves. A straightforward vectorization strategy is to compute global statistical properties, such as the maximum, minimum, sum, median, mean, standard deviation, variance, and entropy, over all entries in the matrix to construct feature vectors. Alternatively, localized linking patterns can be captured by summing the entries of each row or column, thereby quantifying the overall linking contribution of individual segments. Statistical measures can then be applied to these localized values across all segments to generate feature vectors that preserve local structural information.

5.3.4 Algebraic formulations

For persistence barcode, algebraic vectorization is primarily achieved through polynomial maps constructed from the barcode. For example, in the Adcock-Carlsson coordinate method [6], given a barcode with n intervals, a ring of algebraic functions, polynomials in $2n$ variables that satisfy certain properties, is constructed. A k -dimensional feature vector is then obtained by selecting k algebraic functions from this ring and evaluating them on the n (birth, death)

pairs. In the tropical coordinate method [167], for a barcode with n intervals, a function involving $2n$ variables is constructed using only the operations of maximum, minimum, addition, and subtraction. This function is invariant under permutation of the variables. The feature vector is obtained by evaluating the function on all n (birth, death) pairs or (birth, persistence) pairs. There are also methods that utilize complex polynomials defined on barcode intervals to generate feature vectors [105, 125].

For persistence spectra, vectorization primarily involves applying functions to the eigenvalues of the associated Laplacian matrices. For example, in the persistent attribute method [219], the filtration range is divided into uniform bins, and spectral indices of the eigenvalues are computed within each bin. These spectral indices [252] include the Laplacian graph energy, generalized graph energy, generalized average graph energy, spectral radius, spectral diameter, Ivanciu matrix spectrum operators, spectral moments, Lovász–Pelikán index, and quasi-Wiener index. In addition to these spectral descriptors, other spectral properties, such as algebraic connectivity, modularity, Cheeger constant, vertex/edge expansion, and quantities derived from flow, random walks, and heat kernels, can also be incorporated into the vectorization process. Beyond these classical spectral properties, more advanced functions such as the Riemann zeta function [203] and symmetry functions [25] have also been applied to eigenvalues for feature extraction.

5.3.5 Multidimensional formulations

Several formulations incorporate multiple parameters or indices into the feature generation process. For example, Mayer homology includes an additional index in the filtration [269], resulting in 3D persistence barcodes. In addition, multidimensional persistent homology provides an effective framework for transforming dynamic molecular data into image representations, and the multiscale multidimensional persistence further encodes molecular structures at various spatial scales within a single image [314].

For persistence spectra, one approach to generate a multidimensional feature vector is to use the filtration values as the x -axis and plot either the first k eigenvalues or the statistical properties of the entire eigenvalue spectrum on the y -axis. An alternative strategy is to extend the idea of the persistence image to spectral features. Specifically, for each filtration value, one computes the first k eigenvalues or k statistical measures of the full spectrum, resulting in k (filtration, eigen-property) pairs. These pairs can be treated analogously to (birth, death) pairs in persistence barcodes, allowing for the construction of a persistence image. If all such pairs are collected across the entire filtration process, a 2D image representation is obtained. Alternatively, if a separate image is constructed at each filtration value and these are concatenated along the filtration axis, a 3D image can be derived.

Additionally, when applying Hodge Laplacians to Cartesian grids, image representations can be obtained directly via Hodge decomposition. For example, applying the three-component decomposition on a grid yields curl-free,

divergence-free, and harmonic components, each of which has the same spatial dimensions as the original one-form. A multi-channel image can then be constructed by concatenating these three components along the channel axis.

5.3.6 Kernel-based models

Kernel-based machine learning models rely on kernel functions to implicitly map data into high-dimensional feature spaces, enabling the modeling of complex nonlinear relationships without explicitly computing the coordinates in those spaces. The core component of these methods is the kernel function, which acts as a similarity measure between all pairs of data points.

To integrate topological representations such as persistence diagrams into kernel-based learning frameworks, it is essential to define a notion of comparability between such diagrams. A natural way is to define metric structures on the space of persistence diagrams. These metric structures induce distances that can quantify the differences between persistence barcodes/diagrams. Several metrics, such as Hausdorff distance, bottleneck distance [73], and Wasserstein distance [40, 221], have been studied and shown to be robust to small perturbations. However, the computation of these distances, especially for large datasets, can be computationally expensive [105]. It is also challenging to define an appropriate averaging process, as the local mean may not be unique [221, 232]. In particular, these metrics are primarily used for distance-based learning methods such as k-nearest neighbors (KNN) and some clustering algorithms. Therefore, they are unsuitable for machine learning techniques that require a Hilbert space structure, or fixed-size feature inputs, such as support vector machines (SVM), Principal Component Analysis (PCA), decision tree classification, and deep neural networks.

A kernel-based model addresses the issue by defining a kernel that maps persistence diagrams into a Hilbert space, allowing them to be directly integrated with kernel-based machine learning techniques, such as SVM and PCA. The kernel encodes the similarity between persistence diagrams by computing an inner product in the Hilbert space. Based on this construction, various persistent homology-based kernels, such as persistence scale space kernel [255], persistence weighted Gaussian kernel [180], sliced Wasserstein kernel [53], and persistence Fisher kernel [183], have been proposed and demonstrated success in applications as reported in their papers, including shape classification, shape segmentation, texture recognition, etc.

5.3.7 Statistical Frameworks for Persistent Homology

The study of persistence diagrams through statistical methods has advanced significantly in the past decade, addressing the challenges posed by their unordered nature and variable cardinality, especially in the presence of small data perturbations. The metric structure of diagram space, particularly under Wasserstein or bottleneck distances, naturally leads to statistical notions of central tendency, such as Fréchet means and medians [73, 221, 293].

To facilitate the application of classical statistical principles within topological data analysis, persistence diagrams have been increasingly modeled as random variables. The development of functional representations that embed persistence diagrams into more tractable mathematical spaces has played a key role in driving this progress. For example, persistence landscapes [34] embed persistence diagrams into Banach spaces, allowing for the direct computation of means, variances, and confidence bands, thus integrating topological summaries more directly into established statistical methodologies. Similarly, persistence images [4] offer a complementary approach, transforming diagrams into stable, finite-dimensional vector representations that readily support standard statistical analyses. These embeddings have facilitated the establishment of Central Limit Theorems (CLT) for persistence diagrams under Wasserstein distances and analogues of the Law of Large Numbers (LLN) for various functional representations [34, 108], strengthening TDA as a more robust and statistically grounded field, and enhancing its applicability across a wider range of scientific and engineering domains. Furthermore, bootstrap methods have been developed to assess population-level differences in topological summaries [55, 57], enabling formal comparisons between groups by constructing confidence sets and analyzing the asymptotic behavior of topological features [120].

Sufficient statistics encapsulate all the information in a sample that is relevant to the underlying probability distribution. A tropical geometry-based framework constructs sufficient statistics for persistent homology [227]. The developed model maps barcodes into tropical coordinates that are injective and stable. This embedding into a Euclidean space via tropical geometry retains the complete topological information required for parametric modeling while enabling a seamless integration with classical statistical techniques.

Kernel Density Estimation (KDE)-based frameworks work directly on the space of persistence diagrams. Some approaches define expected diagrams as mean objects in a space of Radon measures, drawing on point process theory [107], while others utilize Wasserstein and bottleneck distances to capture the intrinsic geometry of topological features [213]. These methods estimate the distribution of topological features without converting diagrams into vector representations, offering a more intrinsic and geometrically faithful characterization of their variability.

Maroulas et al. [213] introduced nonparametric density estimation techniques tailored for persistence diagrams using the framework of random finite sets. Using Probability Hypothesis Density (PHD), they estimate the distribution of topological features without restrictive assumptions, enabling a robust separation of meaningful topological features from noise, which is essential to analyze complex or noisy data. They provide theoretical convergence guarantees, including bounds on the mean absolute bottleneck deviation, ensuring that density estimates converge to the true distribution.

In a complementary direction, a Bayesian framework for persistent homology has been developed in which persistence diagrams are modeled as realizations from Poisson point processes with prior intensities, enabling posterior inference over topological feature distributions and supporting a probabilistic interpretation of topological summaries [214]. In a related development focused on generative modeling, Papamarkou et al. [242] introduced a method to generate synthetic persistence diagrams based on pairwise interacting point processes and RJ-MCMC sampling, enabling realistic simulation of topological feature distributions while capturing dependencies among features. This helps perform statistical analysis in topological data analysis, especially when data is limited, such as in materials science. A related model extends this probabilistic framework to the classification of biological networks by modeling persistence diagrams as independent and identically distributed cluster point processes, using Gaussian mixtures to capture spatial distribution and binomial distributions to model variability in the number of topological features [212]. Additionally, Bayesian topological signal processing further advances this line of work to handle time-series data, supporting uncertainty-aware classification of dynamic topological features in noisy or evolving signals [236].

6 Concluding remarks

Aided by artificial intelligence (AI), topological data analysis (TDA) has become a rapidly growing field in applied mathematics and data science. A key technique of TDA is persistent homology, a multiscale algebraic topology tool that has demonstrated impressive success across diverse disciplines in science, engineering, medicine, industry, and defense. However, persistent homology with simplicial complex faces several limitations, including: inability to handle non-geometric information, reliance on qualitative descriptions of high-order cycles, inability to track non-topological changes, incapability in handling directed networks, and insensitivity to structured data such as hypergraphs. Additionally, while algebraic topology techniques typically analyze point-cloud data, they are not directly applicable to data on differentiable manifolds or one-dimensional curves embedded in 3-space. These challenges in TDA have motivated significant advancements in the past decade.

This paper provides a comprehensive review of TDA techniques developed to address persistent homology's limitations, emphasizing their mathematical foundations and strengths. Some techniques extend persistent homology from simplicial complexes to other topological domains, such as cell complexes, path complexes, directed flag complexes, cellular sheaves, and hyper(di)graphs. Others introduce alternative topological formulations, including Laplacians [302], Dirac operators [9], sheaf theory, Mayer topology [269], and interaction topology [193]. For data on differentiable manifolds, methods rooted in differential topology, such as persistent de Rham cohomology, persistent Hodge Laplacians, and Hodge decomposition, have been developed [64, 282]. For data on one-dimensional curves in 3-space, tools from geometric topology, such as the

Gauss linking integral [270], Jones polynomial [277], and Khovanov homology [197, 268], have been proposed.

The rapid growth of TDA is driven by machine learning, such as topological deep learning (TDL) [37] and mathematical AI. However, topological representations from TDA cannot be directly implemented in machine learning without appropriate vectorization. This paper also reviews the selection of topological algorithms for different input data formats, their associated topological representations, and methods for topological vectorization in machine learning contexts.

TDA and TDL are dynamic and rapidly evolving fields. The future development of TDA and TDL will be driven by both theoretical curiosity and practical needs, including the demand for interpretable deep neural networks. We envision future advancements in the following directions.

Currently, the local geometry and/or topology surrounding individual data points can be described by persistent cohomology [39], evolutionary homology [41], interaction topology [193], and persistent sheaf Laplacian [309]. Further efforts are needed to refine local topology methods for practical applications.

Most current TDA approaches were developed for point-cloud data. Sequential data, however, are widely used in applications such as DNA sequencing, natural language processing, and electroencephalogram (EEG) analysis. These areas require novel TDA methodologies.

In differential geometry, the Atiyah–Singer index theory establishes a topological index. A persistent index theory for families of elliptic differential operators could be developed for data on compact manifolds.

The advancement of low-dimensional topology methods is critical for analyzing 1D curves embedded in 3-space (e.g., entangled polymers and neural networks). Persistent Floer homology may offer an effective framework for modeling protein-protein binding interactions. Additionally, robust computational algorithms for persistent Khovanov homology and persistent Khovanov Laplacians remain an urgent priority.

Finally, large language models (LLMs) herald a new era in AI, creating opportunities across disciplines. Chatbots like DeepSeek and ChatGPT can transform abstract mathematical theories—including computational topology—into practical tools [196]. Rapid growth is anticipated in AI-assisted topology and topology-enhanced AI in the coming years.

TDA is a vast interdisciplinary field, spanning algebraic, differential, and geometric topology; computational algorithms; AI-driven methodologies; and real-world applications. While this paper aims to comprehensively review TDA and TDL beyond persistent homology, we inevitably omit many significant contributions due to the field’s breadth. We note that tremendous efforts have been made to develop topological neural networks since the introduction of TDL in 2017 [37]. To focus on TDA, we have restricted ourselves from elaborating on this development. The interested reader is referred to a recent review on various aspects of TDL[243]. We hope this work serves as a guide

for researchers exploring topological techniques and helps them select optimal approaches for analyzing specific problems across diverse fields.

Acknowledgments

This work was supported in part by NIH grants R01AI164266, and R35GM148196, NSF grant DMS-2052983, MSU Research Foundation, and Bristol-Myers Squibb 65109. VM's has been partially funded by the National Science Foundation under grants DMS-2012609, DMR-2309083, DRL-2314155, DMS-1821241, MCB-1715794, and DGE-2152168; the U.S. Army Research Laboratory (ARL) under contracts W911NF-22-2-0143, W911NF-21-2-0186, and W911NF-19-2-0328; the U.S. Army Research Office (ARO) under contracts W911NF-21-1-0094 and W911NF-17-1-0313; and Thor Industries/ARL under contract W911NF-17-2-0141.

References

- [1] Adams, C.C.: *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*. American Mathematical Society, (2004)
- [2] Adams, H., Carlsson, G.: On the nonlinear statistics of range image patches. *SIAM Journal on Imaging Sciences* **2**(1), 110–117 (2009)
- [3] Adams, H., Tausz, A., Vejdemo-Johansson, M.: Javaplex: a research software package for persistent (co)homology. In: *Mathematical Software—ICMS 2014: 4th International Congress, Seoul, South Korea, August 5–9, 2014. Proceedings* 4, pp. 129–136 (2014). Springer
- [4] Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., Ziegelmeier, L.: Persistence images: a stable vector representation of persistent homology. *Journal of Machine Learning Research* **18**(8), 1–35 (2017)
- [5] Adams, H., Lagoda, E., Moy, M., Sadovek, N., De Saha, A.: Persistent equivariant cohomology. arXiv preprint arXiv:2408.17331 (2024)
- [6] Adcock, A., Carlsson, E., Carlsson, G.: The ring of algebraic functions on persistence bar codes. *Homology, Homotopy and Applications* **18**(1), 381–402 (2016)
- [7] Alexander, J.W.: Topological invariants of knots and links. *Transactions of the American Mathematical Society* **30**(2), 275–306 (1928)

- [8] Ali, D., Asaad, A., Jimenez, M.-J., Nanda, V., Paluzo-Hidalgo, E., Soriano-Trigueros, M.: A survey of vectorization methods in topological data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(12), 14069–14080 (2023)
- [9] Ameneyro, B., Maroulas, V., Siopsis, G.: Quantum persistent homology. *Journal of Applied and Computational Topology* **8**(7), 1961–1980 (2024)
- [10] Ameneyro, B., Siopsis, G., Maroulas, V.: Quantum persistent homology for time series. In: APS March Meeting Abstracts, vol. 2023, pp. 73–003 (2023)
- [11] Arnold, D.N.: Finite Element Exterior Calculus. SIAM, (2018)
- [12] Arnold, D.N., Falk, R.S., Winther, R.: Finite element exterior calculus, homological techniques, and applications. *Acta numerica* **15**, 1–155 (2006)
- [13] Arsuaga, J., Vazquez, M., McGuirk, P., Trigueros, S., Sumners, D.W., Roca, J.: DNA knots reveal a chiral organization of DNA in phage capsids. *Proceedings of the National Academy of Sciences* **102**(26), 9165–9169 (2005)
- [14] Asaad, A., Ali, D., Majeed, T., Rashid, R.: Persistent homology for breast tumor classification using mammogram scans. *Mathematics* **10**(21), 4039 (2022)
- [15] Atienza, N., Gonzalez-Diaz, R., Rucco, M.: Persistent entropy for separating topological features from noise in Vietoris-Rips complexes. *Journal of Intelligent Information Systems* **52**, 637–655 (2019)
- [16] Baccini, F., Geraci, F., Bianconi, G.: Weighted simplicial complexes and their representation power of higher-order network data and topology. *Physical Review E* **106**(3), 034319 (2022)
- [17] Bae, W., Yoo, J., Chul Ye, J.: Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 145–153 (2017)
- [18] Baldwin, Q., Panagiotou, E.: The local topological free energy of proteins. *Journal of Theoretical Biology* **529**, 110854 (2021)
- [19] Baldwin, Q., Sumpter, B., Panagiotou, E.: The local topological free energy of the SARS-CoV-2 spike protein. *Polymers* **14**(15), 3014 (2022)
- [20] Barkataki, K., Panagiotou, E.: The Jones polynomial of collections of

- open curves in 3-space. *Proceedings of the Royal Society A* **478**(2267), 20220302 (2022)
- [21] Barnes, D., Polanco, L., Perea, J.A.: A comparative study of machine learning methods for persistence diagrams. *Frontiers in Artificial Intelligence* **4**, 681174 (2021)
 - [22] Bauer, U.: Ripser: efficient computation of Vietoris–Rips persistence barcodes. *Journal of Applied and Computational Topology* **5**(3), 391–423 (2021)
 - [23] Bauer, U., Kerber, M., Reininghaus, J.: Diph'a (a distributed persistent homology algorithm). Software available at <https://github.com/DIPHA/diph'a> (2014)
 - [24] Bauer, U., Kerber, M., Reininghaus, J., Wagner, H.: Phat—persistent homology algorithms toolbox. *Journal of symbolic computation* **78**, 76–90 (2017)
 - [25] Behler, J., Parrinello, M.: Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical review letters* **98**(14), 146401 (2007)
 - [26] Bhatia, H., Norgard, G., Pascucci, V., Bremer, P.-T.: The Helmholtz–Hodge decomposition - a survey. *IEEE Transactions on visualization and computer graphics* **19**(8), 1386–1404 (2012)
 - [27] Bi, W., Li, J., Liu, J., Wu, J.: On the Cayley-persistence algebra. arXiv preprint arXiv:2205.10796 (2022)
 - [28] Bianconi, G.: The topological Dirac equation of networks and simplicial complexes. *Journal of Physics: Complexity* **2**(3), 035022 (2021)
 - [29] Biasotti, S., Giorgi, D., Spagnuolo, M., Falcidieno, B.: Reeb graphs for shape analysis and applications. *Theoretical computer science* **392**(1-3), 5–22 (2008)
 - [30] Bodnar, C., Frasca, F., Wang, Y., Otter, N., Montúfar, G.F., Liò, P., Bronstein, M.: Weisfeiler and lehman go topological: Message passing simplicial networks. In: *Proceedings of the 38th International Conference on Machine Learning*, pp. 1026–1037 (2021)
 - [31] Botnan, M., Lesnick, M.: An introduction to multiparameter persistence. In: *Representations of Algebras and Related Structures*, pp. 77–150 (2023)

- [32] Botnan, M.B., Lebovici, V., Oudot, S.: On rectangle-decomposable 2-parameter persistence modules. In: Leibniz International Proceedings in Informatics, vol. 164, pp. 22–1 (2020). Leibniz-Zentrum für Informatik
- [33] Bressan, S., Li, J., Ren, S., Wu, J.: The embedded homology of hypergraphs and applications. *Asian Journal of Mathematics* **23**(3), 479–500 (2019)
- [34] Bubenik, P.: Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research* **16**(1), 77–102 (2015)
- [35] Bubenik, P., Dłotko, P.: A persistence landscapes toolbox for topological statistics. *Journal of Symbolic Computation* **78**, 91–114 (2017)
- [36] Calmon, L., Schaub, M.T., Bianconi, G.: Dirac signal processing of higher-order topological signals. *New Journal of Physics* **25**(9), 093013 (2023)
- [37] Cang, Z., Wei, G.-W.: Topologynet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions. *PLoS computational biology* **13**(7), 1005690 (2017)
- [38] Cang, Z., Wei, G.-W.: Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction. *International journal for numerical methods in biomedical engineering* **34**(2), 2914 (2018)
- [39] Cang, Z., Wei, G.-W.: Persistent cohomology for data with multicomponent heterogeneous information. *SIAM journal on mathematics of data science* **2**(2), 396–418 (2020)
- [40] Cang, Z., Mu, L., Wei, G.-W.: Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening. *PLoS computational biology* **14**(1), 1005929 (2018)
- [41] Cang, Z., Munch, E., Wei, G.-W.: Evolutionary homology on coupled dynamical systems with applications to protein flexibility analysis. *Journal of applied and computational topology* **4**(4), 481–507 (2020)
- [42] Cang, Z., Mu, L., Wu, K., Opron, K., Xia, K., Wei, G.-W.: A topological approach for protein classification. *Computational and Mathematical Biophysics* **3**(1) (2015)
- [43] Cantarella, J., DeTurck, D., Gluck, H.: Vector calculus and the topology of domains in 3-space. *The American Mathematical Monthly* **109**(5), 409–442 (2002)

- [44] Carlsson, G.: Topology and data. *Bulletin of the American Mathematical Society* **46**(2), 255–308 (2009)
- [45] Carlsson, G., De Silva, V.: Zigzag persistence. *Foundations of computational mathematics* **10**, 367–405 (2010)
- [46] Carlsson, G., Zomorodian, A.: The theory of multidimensional persistence. In: *Proceedings of the Twenty-third Annual Symposium on Computational Geometry*, pp. 184–193 (2007)
- [47] Carlsson, G., De Silva, V., Morozov, D.: Zigzag persistent homology and real-valued functions. In: *Proceedings of the Twenty-fifth Annual Symposium on Computational Geometry*, pp. 247–256 (2009)
- [48] Carlsson, G., Zomorodian, A., Collins, A., Guibas, L.: Persistence barcodes for shapes. In: *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, pp. 124–135 (2004)
- [49] Carlsson, G., Ishkhanov, T., De Silva, V., Zomorodian, A.: On the local behavior of spaces of natural images. *International journal of computer vision* **76**, 1–12 (2008)
- [50] Carlsson, G., De Silva, V., Kališnik, S., Morozov, D.: Parametrized homology via zigzag persistence. *Algebraic & Geometric Topology* **19**(2), 657–700 (2019)
- [51] Carrière, M.: Cover complex. In: *GUDHI User and Reference Manual*, (2025)
- [52] Carriere, M., Blumberg, A.: Multiparameter persistence image for topological machine learning. *Advances in Neural Information Processing Systems* **33**, 22432–22444 (2020)
- [53] Carriere, M., Cuturi, M., Oudot, S.: Sliced Wasserstein kernel for persistence diagrams. In: *International Conference on Machine Learning*, pp. 664–673 (2017). PMLR
- [54] Chan, J.M., Carlsson, G., Rabadian, R.: Topology of viral evolution. *Proceedings of the National Academy of Sciences* **110**(46), 18566–18571 (2013)
- [55] Chazal, F., Fasy, B.T., Lecci, F., Rinaldo, A., Wasserman, L.: Stochastic convergence of persistence landscapes and silhouettes. In: *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, pp. 474–483 (2014)
- [56] Chazal, F., De Silva, V., Glisse, M., Oudot, S.: The Structure and

- Stability of Persistence Modules vol. 10. Springer, (2016)
- [57] Chazal, F., Fasy, B.T., Lecci, F., Michel, B., Rinaldo, A., Wasserman, L.: Robust topological inference: Distance to a measure and kernel distance. *Journal of Machine Learning Research* **18**(159), 1–40 (2018)
 - [58] Chen, D., Liu, J., Wei, G.-W.: Multiscale topology-enabled structure-to-sequence transformer for protein–ligand interaction predictions. *Nature Machine Intelligence* **6**(7), 799–810 (2024)
 - [59] Chen, D., Liu, J., Wu, J., Wei, G.-W., Pan, F., Yau, S.-T.: Path topology in molecular and materials sciences. *The journal of physical chemistry letters* **14**(4), 954–964 (2023)
 - [60] Chen, D., Liu, J., Wu, J., Wei, G.-W.: Persistent hyperdigraph homology and persistent hyperdigraph Laplacians. *Foundations of Data Science* **5**(4), 558–588 (2023)
 - [61] Chen, G., Mischaikow, K., Laramee, R.S., Zhang, E.: Efficient Morse decompositions of vector fields. *IEEE Transactions on Visualization and Computer Graphics* **14**(4), 848–862 (2008)
 - [62] Chen, J., Wei, G.-W.: Omicron BA. 2 (B. 1.1. 529.2): high potential for becoming the next dominant variant. *The journal of physical chemistry letters* **13**(17), 3840–3849 (2022)
 - [63] Chen, J., Wang, R., Wang, M., Wei, G.-W.: Mutations strengthened SARS-CoV-2 infectivity. *Journal of molecular biology* **432**(19), 5212–5226 (2020)
 - [64] Chen, J., Zhao, R., Tong, Y., Wei, G.-W.: Evolutionary de Rham-Hodge method. *Discrete and continuous dynamical systems. Series B* **26**(7), 3785 (2021)
 - [65] Chen, J., Qiu, Y., Wang, R., Wei, G.-W.: Persistent Laplacian projected Omicron BA. 4 and BA. 5 to become new dominating variants. *Computers in Biology and Medicine* **151**, 106262 (2022)
 - [66] Chen, J., Woldring, D.R., Huang, F., Huang, X., Wei, G.-W.: Topological deep learning based deep mutational scanning. *Computers in biology and medicine* **164**, 107258 (2023)
 - [67] Chen, Y., Segovia, I., Gel, Y.R.: Z-GCNETs: time zigzags at graph convolutional networks for time series forecasting. In: International Conference on Machine Learning, pp. 1684–1694 (2021). PMLR

- [68] Chevyrev, I., Nanda, V., Oberhauser, H.: Persistence paths and signature features in topological data analysis. *IEEE transactions on pattern analysis and machine intelligence* **42**(1), 192–202 (2018)
- [69] Chowdhury, S., Mémoli, F.: Persistent path homology of directed networks. In: *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1152–1169 (2018). SIAM
- [70] Chung, Y.-M., Lawson, A.: Persistence curves: A canonical framework for summarizing persistence diagrams. *Advances in Computational Mathematics* **48**(1), 6 (2022)
- [71] Ciarlet, P.G.: *The Finite Element Method for Elliptic Problems*. SIAM, (2002)
- [72] Clough, J.R., Byrne, N., Oksuz, I., Zimmer, V.A., Schnabel, J.A., King, A.P.: A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE transactions on pattern analysis and machine intelligence* **44**(12), 8766–8778 (2020)
- [73] Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. In: *Proceedings of the Twenty-first Annual Symposium on Computational Geometry*, pp. 263–271 (2005)
- [74] Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Extending persistence using Poincaré and Lefschetz duality. *Foundations of Computational Mathematics* **9**(1), 79–103 (2009)
- [75] Collins, A., Zomorodian, A., Carlsson, G., Guibas, L.J.: A barcode shape descriptor for curve point cloud data. *Computers & Graphics* **28**(6), 881–894 (2004)
- [76] Conley, C.C.: *Isolated Invariant Sets and the Morse Index* vol. 38. American Mathematical Soc., (1978)
- [77] Conti, F., Moroni, D., Pascali, M.A.: A topological machine learning pipeline for classification. *Mathematics* **10**(17), 3086 (2022)
- [78] Cooperband, Z., Ghrist, R.: Unified origami kinematics via cosheaf homology. *arXiv preprint arXiv:2501.02581* (2025)
- [79] Cooperband, Z., Ghrist, R., Hansen, J.: A cosheaf theory of reciprocal figures: Planar and higher genus graphic statics. *arXiv preprint arXiv:2311.12946* (2023)
- [80] Cottrell, S., Hozumi, Y., Wei, G.-W.: K-nearest-neighbors induced topological PCA for single cell RNA-sequence data analysis. *Computers in*

biology and medicine **175**, 108497 (2024)

- [81] Cottrell, S., Wang, R., Wei, G.-W.: PLPCA: persistent Laplacian-enhanced PCA for microarray data analysis. *Journal of chemical information and modeling* **64**(7), 2405–2420 (2023)
- [82] Crowell, R.H., Fox, R.H.: *Introduction to Knot Theory* vol. 57. Springer, (2012)
- [83] Curry, J.M.: *Sheaves, Cosheaves and Applications*. University of Pennsylvania, (2014)
- [84] Dabaghian, Y., Mémoli, F., Frank, L., Carlsson, G.: A topological paradigm for hippocampal spatial map formation using persistent homology (2012)
- [85] Dabrowski-Tumanski, P., Jarmolinska, A.I., Niemyska, W., Rawdon, E.J., Millett, K.C., Sulkowska, J.I.: Linkprot: a database collecting information about biological links. *Nucleic acids research*, 976 (2016)
- [86] Dabrowski-Tumanski, P., Rubach, P., Niemyska, W., Gren, B.A., Sulkowska, J.I.: Topoly: Python package to analyze topology of polymers. *Briefings in Bioinformatics* **22**(3), 196 (2021)
- [87] Dawson, R.J.M.: Homology of weighted simplicial complexes. *Cahiers de Topologie et Géométrie Différentielle Catégoriques* **31**(3), 229–243 (1990)
- [88] De Silva, V., Carlsson, G.E.: Topological estimation using witness complexes. In: PBG, pp. 157–166 (2004)
- [89] De Silva, V., Vejdemo-Johansson, M.: Persistent cohomology and circular coordinates. In: Proceedings of the Twenty-fifth Annual Symposium on Computational Geometry, pp. 227–236 (2009)
- [90] De Silva, V., Morozov, D., Vejdemo-Johansson, M.: Dualities in persistent (co)homology. *Inverse Problems* **27**(12), 124003 (2011)
- [91] Delaunay, B.: Sur la sphère vide. *Izvestia Akademii Nauk SSSR* **7**, 793–800 (1934)
- [92] Deligne, P.: La conjecture de weil. i. *Publications Mathématiques de l’Institut des Hautes Études Scientifiques* **43**, 273–307 (1974)
- [93] Deligne, P.: La conjecture de weil: II. *Publications Mathématiques de l’IHES* **52**, 137–252 (1980)

- [94] Desbrun, M., Kanso, E., Tong, Y.: Discrete differential forms for computational modeling. In: ACM SIGGRAPH 2006 Courses, pp. 39–54 (2006)
- [95] Dey, T.K., Hou, T.: Computing zigzag persistence on graphs in near-linear time. In: 37th International Symposium on Computational Geometry (2021)
- [96] Dey, T.K., Hou, T.: Updating zigzag persistence and maintaining representatives over changing filtrations. arXiv preprint arXiv:2112.02352 (2021)
- [97] Dey, T.K., Hou, T.: Fast computation of zigzag persistence. In: 30th Annual European Symposium on Algorithms (ESA 2022). Leibniz International Proceedings in Informatics (LIPIcs), vol. 244 (2022). pp. 43:1–43:15
- [98] Dey, T.K., Kim, W., Mémoli, F.: Computing generalized rank invariant for 2-parameter persistence modules via zigzag persistence and its applications. *Discrete & Computational Geometry* **71**(1), 67–94 (2024)
- [99] Dey, T.K., Li, T., Wang, Y.: An efficient algorithm for 1-dimensional (persistent) path homology. *Discrete & Computational Geometry* **68**(4), 1102–1132 (2022)
- [100] Dey, T.K., Mrozek, M., Słoboda, R.: Persistence of the Conley index in combinatorial dynamical systems. In: 36th International Symposium on Computational Geometry (SoCG 2020). Leibniz International Proceedings in Informatics (LIPIcs), vol. 164 (2020). pp. 37:1–37:17
- [101] Dey, T.K., Mrozek, M., Słoboda, R.: Persistence of Conley-Morse graphs in combinatorial dynamical systems. *SIAM Journal on Applied Dynamical Systems* **21**(2), 817–839 (2022)
- [102] Dey, T.K., Juda, M., Kapela, T., Kubica, J., Lipiński, M., Mrozek, M.: Persistent homology of Morse decompositions in combinatorial dynamics. *SIAM Journal on Applied Dynamical Systems* **18**(1), 510–530 (2019)
- [103] Dey, T.K., Wang, Y.: Computational Topology for Data Analysis. Cambridge University Press, (2022)
- [104] Di, S., Ivanov, S.O., Mukoseev, L., Zhang, M.: On the path homology of cayley digraphs and covering digraphs. *Journal of Algebra* **653**, 156–199 (2024)

- [105] Di Fabio, B., Ferri, M.: Comparing persistence diagrams through complex vectors. In: *Image Analysis and Processing-ICIAP 2015: 18th International Conference, Genoa, Italy, September 7-11, 2015, Proceedings, Part I* 18, pp. 294–305 (2015). Springer
- [106] Dirac, P.A.M.: The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **117**(778), 610–624 (1928)
- [107] Divol, V., Chazal, F.: The density of expected persistence diagrams and its kernel based estimation. *Journal of Computational Geometry* **10**(2), 1–31 (2019). <https://doi.org/10.20382/jocg.v10i2a7>
- [108] Divol, V., Lacombe, T.: Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *Journal of Applied and Computational Topology* **5**(1), 1–53 (2021). <https://doi.org/10.1007/s41468-020-00061-z>
- [109] Dłotko, P., Wanner, T.: Topological microstructure analysis using persistence landscapes. *Physica D: Nonlinear Phenomena* **334**, 60–81 (2016)
- [110] Du, H., Wei, G.-W., Hou, T.: Multiscale topology in interactomic network: from transcriptome to antiaddiction drug repurposing. *Briefings in Bioinformatics* **25**(2), 054 (2024)
- [111] Eckmann, B.: Harmonische funktionen und randwertaufgaben in einem komplex. *Commentarii Mathematici Helvetici* **17**(1), 240–255 (1944)
- [112] Edelsbrunner, Letscher, Zomorodian: Topological persistence and simplification. *Discrete & computational geometry* **28**, 511–533 (2002)
- [113] Edelsbrunner, H.: Alpha shapes - a survey. In: *Tessellations in the Sciences: Virtues, Techniques and Applications of Geometric Tilings*, (2011)
- [114] Edelsbrunner, H., Harer, J.: *Computational Topology: An Introduction*. American Mathematical Soc., (2010)
- [115] Edelsbrunner, H., Morozov, D.: *Persistent Homology: Theory and Practice*. eScholarship, University of California, (2013)
- [116] Edelsbrunner, H., Harer, J., *et al.*: Persistent homology - a survey. *Contemporary mathematics* **453**(26), 257–282 (2008)
- [117] Einzade, A., Thanou, D., Malliaros, F.D., Giraldo, J.H.: Cosmos: Continuous simplicial neural networks. arXiv preprint arXiv:2503.12919 (2025)

- [118] Estrada, E.: Path Laplacian matrices: introduction and application to the analysis of consensus in networks. *Linear algebra and its applications* **436**(9), 3373–3391 (2012)
- [119] Fabri, A., Pion, S.: CGAL: the computational geometry algorithms library. In: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 538–539 (2009)
- [120] Fasy, B.T., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S., Singh, A.: Confidence sets for persistence diagrams. *The Annals of Statistics* **42**(6), 2301–2339 (2014). <https://doi.org/10.1214/14-AOS1252>
- [121] Fasy, B.T., Kim, J., Lecci, F., Maria, C.: Introduction to the R package TDA. arXiv preprint arXiv:1411.1830 (2014)
- [122] Feng, H., Shen, L., Liu, J., Wei, G.-W.: Mayer-homology learning prediction of protein-ligand binding affinities. *Journal of Computational Biophysics and Chemistry*, 1–14 (2024)
- [123] Feng, L., Gong, H., Zhang, S., Liu, X., Wang, Y., Che, J., Dong, A., Griffin, C.H., Gragnoli, C., Wu, J., Yau, S.-T., Wu, R.: Hypernetwork modeling and topology of high-order interactions for complex systems. *Proceedings of the National Academy of Sciences* **121**(40), 2412220121 (2024)
- [124] Feng, L., Yang, D., Wu, S., Xue, C., Sang, M., Liu, X., Che, J., Wu, J., Gragnoli, C., Griffin, C., Wang, C., Yau, S.-T., Wu, R.: Network modeling and topology of aging. *Physics reports* **1101**, 1–65 (2025)
- [125] Ferri, M., Landi, C.: Representing size functions by complex polynomials. *Proc. Math. Met. in Pattern Recognition* **9**, 16–19 (1999)
- [126] Friedrichs, K.O.: Differential forms on Riemannian manifolds. *Communications on Pure and Applied Mathematics* **8**(4), 551–590 (1955)
- [127] Frosini, P., Mulazzani, M.: Size homotopy groups for computation of natural size distances. *Bulletin of the Belgian Mathematical Society-Simon Stevin* **6**(3), 455–464 (1999)
- [128] Gameiro, M., Hiraoka, Y., Izumi, S., Kramar, M., Mischaikow, K., Nanda, V.: A topological measurement of protein compressibility. *Japan Journal of Industrial and Applied Mathematics* **32**, 1–17 (2015)
- [129] Gauss, C.F.: Zur Mathematischen Theorie der Electrodynamicischen Wirkungen, pp. 601–630. Springer, Berlin, Heidelberg (1877)

- [130] Ge, X., Safa, I., Belkin, M., Wang, Y.: Data skeletonization via Reeb graphs. *Advances in neural information processing systems* **24** (2011)
- [131] Geuzaine, C., Remacle, J.-F.: Gmsh: a 3-D finite element mesh generator with built-in pre- and post-processing facilities. *International journal for numerical methods in engineering* **79**(11), 1309–1331 (2009)
- [132] Ghrist, R.: Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society* **45**(1), 61–75 (2008)
- [133] Ghrist, R.W.: Elementary Applied Topology vol. 1. Createspace Seattle, (2014)
- [134] Goldberg, T.E.: Combinatorial Laplacians of simplicial complexes. PhD thesis, Citeseer (2002)
- [135] Gong, H., Wang, H., Wang, Y., Zhang, S., Liu, X., Che, J., Wu, S., Wu, J., Sun, X., Zhang, S., Yau, S.-T., Wu, R.: Topological change of soil microbiota networks for forest resilience under global warming. *Physics of Life Reviews* (2024)
- [136] Grbić, J., Wu, J., Xia, K., Wei, G.-W.: Aspects of topological approaches for data science. *Foundations of data science (Springfield, Mo.)* **4**(2), 165 (2022)
- [137] Grigor'yan, A., Lin, Y., Muranov, Y., Yau, S.-T.: Homologies of path complexes and digraphs. *arXiv preprint arXiv:1207.2834* (2012)
- [138] Grigor'yan, A.: Advances in path homology theory of digraphs. *Notices of the International Consortium of Chinese Mathematicians* **10**(2), 61–124 (2022)
- [139] Grigor'yan, A., Muranov, Y., Yau, S.-T.: On a cohomology of digraphs and hochschild cohomology. *Journal of Homotopy and Related Structures* **11**(2), 209–230 (2016)
- [140] Grigor'yan, A., Lin, Y., Muranov, Y., Yau, S.-T.: Homotopy theory for digraphs. *Pure and Applied Mathematics Quarterly* **10**(4), 619–674 (2023)
- [141] Gülen, A.B., Mémoli, F., Wan, Z.: Orthogonal möbius inversion and Grassmannian persistence diagrams. *arXiv preprint arXiv:2311.06870* (2023)
- [142] Gülen, A.B., Mémoli, F., Wan, Z.: Grassmannian persistence diagrams: Special properties in the 1-parameter setting. *arXiv preprint arXiv:2504.06077* (2025)

- [143] Gülen, A.B., Mémoli, F., Wan, Z., Wang, Y.: A Generalization of the Persistent Laplacian to Simplicial Maps. In: 39th International Symposium on Computational Geometry (SoCG 2023). Leibniz International Proceedings in Informatics (LIPIcs), vol. 258 (2023). pp. 37:1–37:17
- [144] Gundert, A., Szedlák, M.: Higher dimensional Cheeger inequalities. In: Proceedings of the Thirtieth Annual Symposium on Computational Geometry, pp. 181–188 (2014)
- [145] Günther, D., Reininghaus, J., Wagner, H., Hotz, I.: Efficient computation of 3D Morse–Smale complexes and persistent homology using discrete Morse theory. *The Visual Computer* **28**, 959–969 (2012)
- [146] Hang, H., Mio, W.: Correspondence modules and persistence sheaves: a unifying perspective on one-parameter persistent homology. *Japan Journal of Industrial and Applied Mathematics* **40**(1), 41–93 (2023)
- [147] Hansen, J.: Laplacians of cellular sheaves: Theory and applications. PhD thesis, University of Pennsylvania (2020)
- [148] Hansen, J., Ghrist, R.: Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology* **3**(4), 315–358 (2019)
- [149] Hayakawa, R.: Quantum algorithm for persistent Betti numbers and topological data analysis. *Quantum* **6**, 873 (2022)
- [150] He, Y., Liu, J.: Multi-scale hochschild spectral analysis on graph data. *AIMS Mathematics* **10**(1), 1384–1406 (2025)
- [151] Hensel, F., Moor, M., Rieck, B.: A survey of topological machine learning methods. *Frontiers in Artificial Intelligence* **4**, 681108 (2021)
- [152] Henselman, G., Ghrist, R.: Matroid filtrations and computational persistent homology. arXiv preprint arXiv:1606.00199 (2016)
- [153] Hernández-García, P., Serrano, D.H., Gómez, D.S.: From persistence to resilience: New betti numbers for analyzing robustness in simplicial complex networks. arXiv preprint arXiv:2505.10467 (2025)
- [154] Holme, P., Saramäki, J.: Temporal networks. *Physics reports* **519**(3), 97–125 (2012)
- [155] Horak, D., Jost, J.: Spectra of combinatorial Laplace operators on simplicial complexes. *Advances in Mathematics* **244**, 303–336 (2013)
- [156] Horak, D., Maletić, S., Rajković, M.: Persistent homology of complex networks. *Journal of Statistical Mechanics: Theory and Experiment* **2009**(03), 03034 (2009)

- [157] Hozumi, Y., Wei, G.-W.: Revealing the shape of genome space via k-mer topology. arXiv preprint arXiv:2412.20202 (2024)
- [158] Ivanov, S.O., Pavutnitskiy, F.: Simplicial approach to path homology of quivers, marked categories, groups and algebras. Journal of the London Mathematical Society **109**(1), 12812 (2024)
- [159] Jamin, C.: Tangential complex. In: GUDHI User and Reference Manual, (2025)
- [160] Jamroz, M., Niemyska, W., Rawdon, E.J., Stasiak, A., Millett, K.C., Sulkowski, P., Sulkowska, J.I.: Knotprot: a database of proteins with knots and slipknots. Nucleic acids research **43**(D1), 306–314 (2015)
- [161] Jiang, Y., Chen, D., Chen, X., Li, T., Wei, G.-W., Pan, F.: Topological representations of crystalline compounds for the machine-learning prediction of materials properties. npj computational materials **7**(1), 28 (2021)
- [162] Jones, B., Wei, G.-W.: Khovanov Laplacian and Khovanov Dirac for knots and links. Journal of Physics: Complexity **6**(2), 025014 (2025)
- [163] Jones, B., Wei, G.-W.: Persistent directed flag Laplacian. Foundations of Data Science **7**(3), 737–758 (2025)
- [164] Jones, V.F.: A polynomial invariant for knots via von Neumann algebras. In: Fields Medallists' Lectures, pp. 448–458. World Scientific, (1997)
- [165] Jonsson, J.: Simplicial Complexes of Graphs vol. 1928. Springer, (2008)
- [166] Kaji, S., Sudo, T., Ahara, K.: Cubical Ripser: Software for computing persistent homology of image and volume data. arXiv preprint arXiv:2005.12692 (2020)
- [167] Kališnik, S.: Tropical coordinates on the space of persistence barcodes. Foundations of Computational Mathematics **19**(1), 101–129 (2019)
- [168] Kannan, H., Saucan, E., Roy, I., Samal, A.: Persistent homology of unweighted complex networks via discrete Morse theory. Scientific reports **9**(1), 13817 (2019)
- [169] Karagüler, D.: A survey on multidimensional persistence theory. Master's thesis, Middle East Technical University (Turkey) (2021)
- [170] Karan, A., Kaygun, A.: Time series classification via topological data analysis. Expert Systems with Applications **183**, 115326 (2021). <https://doi.org/10.1016/j.eswa.2021.115326>

- [171] Keros, A., Subr, K.: Spectral coarsening with Hodge Laplacians. In: ACM SIGGRAPH 2023 Conference Proceedings, pp. 1–11 (2023)
- [172] Khovanov, M.: A categorification of the Jones polynomial (2000)
- [173] Kim, K., Kim, J., Zaheer, M., Kim, J., Chazal, F., Wasserman, L.: Plplay: Efficient topological layer based on persistent landscapes. Advances in Neural Information Processing Systems **33**, 15965–15977 (2020)
- [174] Kim, W., Mémoli, F.: Generalized persistence diagrams for persistence modules over posets. Journal of Applied and Computational Topology **5**(4), 533–581 (2021)
- [175] Kim, W., Mémoli, F.: Persistence over posets. Notices of the American Mathematical Society **70**(08) (2023)
- [176] Kirchhoff, G.: Ueber die auflösung der gleichungen, auf welche man bei der untersuchung der linearen vertheilung galvanischer ströme geführt wird. Annalen der Physik **148**(12), 497–508 (1847)
- [177] Knill, O.: The cohomology for Wu characteristics. arXiv preprint arXiv:1803.06788 (2018)
- [178] Kovacev-Nikolic, V., Bubenik, P., Nikolić, D., Heo, G.: Using persistent homology and dynamical distances to analyze protein binding. Statistical applications in genetics and molecular biology **15**(1), 19–38 (2016)
- [179] Krishnagopal, S., Bianconi, G.: Topology and dynamics of higher-order multiplex networks. Chaos, Solitons & Fractals **177**, 114296 (2023)
- [180] Kusano, G., Hiraoka, Y., Fukumizu, K.: Persistence weighted Gaussian kernel for topological data analysis. In: International Conference on Machine Learning, pp. 2004–2013 (2016). PMLR
- [181] Ladyzhenskaya, O.A.: The Mathematical Theory of Viscous Incompressible Flow. Gordon & Breach, (1969)
- [182] Le, M.Q., Taylor, D.: Persistent homology with k-nearest-neighbor filtrations reveals topological convergence of pagerank. Foundations of Data Science **7**(2), 536–567 (2025)
- [183] Le, T., Yamada, M.: Persistence Fisher kernel: A Riemannian manifold kernel for persistence diagrams. Advances in neural information processing systems **31** (2018)
- [184] Lee, Y., Barthel, S.D., Dłotko, P., Moosavi, S.M., Hess, K., Smit, B.: Quantifying similarity of pore-geometry in nanoporous materials. Nature communications **8**(1), 1–8 (2017)

- [185] Leray, J.: L’anneau d’homologie d’une représentation. *C.R. Acad. Sci. Paris* (222), 1366–1368 (1946)
- [186] Li, J., Muranov, Y., Wu, J., Yau, S.-T.: On singular homology theories of digraphs and quivers (2024)
- [187] Li, J., Muranov, Y., Wu, J., Yau, S.-T.: Primitive path homology. arXiv preprint arXiv:2411.18955 (2024)
- [188] LI, J., VERSHININ, V., WU, J.: Twisted simplicial groups and twisted homology of categories. *Homology, Homotopy & Applications* **19**(2) (2017)
- [189] Liang, C., Mislow, K.: Knots in proteins. *Journal of the American Chemical Society* **116**(24), 11189–11190 (1994)
- [190] Lieutier, A.: Talk: Persistent harmonic forms (2014)
- [191] Lim, L.-H.: Hodge Laplacians on graphs. *Siam Review* **62**(3), 685–715 (2020)
- [192] Lin, Y., Ren, S., Wang, C., Wu, J.: Weighted path homology of weighted digraphs and persistence. arXiv preprint arXiv:1910.09891 (2019)
- [193] Liu, J., Chen, D., Wei, G.-W.: Interaction homotopy and interaction homology. arXiv preprint arXiv:2311.16322 (2023)
- [194] Liu, J., Chen, D., Wei, G.-W.: Persistent interaction topology in data analysis. arXiv preprint arXiv:2404.11799 (2024)
- [195] Liu, J., Li, J., Wu, J.: The algebraic stability for persistent Laplacians. *Homology, Homotopy and Applications* **26**(2), 297–323 (2024)
- [196] Liu, J., Shen, L., Wei, G.-W.: ChatGPT for computational topology. *Foundations of data science* **6**(2), 221–250 (2024)
- [197] Liu, J., Shen, L., Wei, G.-W.: Persistent Khovanov homology of tangles. arXiv preprint arXiv:2409.18312 (2024)
- [198] Liu, J., Shen, L., Chen, D., Wei, G.-W.: Topological sequence analysis of genomes: Delta complex approaches. arXiv preprint arXiv:2507.05452 (2025)
- [199] Liu, R., Liu, X., Wu, J.: Persistent path-spectral (pps) based machine learning for protein–ligand binding affinity prediction. *Journal of chemical information and modeling* **63**(3), 1066–1075 (2023)
- [200] Liu, X., Xia, K.: Neighborhood complex based machine learning (NCML)

- models for drug design. In: International Workshop on Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data, pp. 87–97 (2021)
- [201] Liu, X., Wang, X., Wu, J., Xia, K.: Hypergraph-based persistent cohomology (HPC) for molecular representations in drug design. *Briefings in Bioinformatics* **22**(5), 411 (2021)
 - [202] Liu, X., Feng, H., Wu, J., Xia, K.: Persistent spectral hypergraph based machine learning (PSH-ML) for protein-ligand binding affinity prediction. *Briefings in Bioinformatics* **22**(5), 127 (2021)
 - [203] Liu, X., Feng, H., Wu, J., Xia, K.: Dowker complex based machine learning (DCML) models for protein-ligand binding affinity prediction. *PLoS computational biology* **18**(4), 1009943 (2022)
 - [204] Liu, X., Feng, H., Wu, J., Xia, K.: Hom-complex-based machine learning (HCML) for the prediction of protein-protein binding affinity changes upon mutation. *Journal of chemical information and modeling* **62**(17), 3961–3969 (2022)
 - [205] Liu, X., Feng, H., Wu, J., Xia, K.: Computing hypergraph homology. *Foundations of Data Science* **6**(2), 172–194 (2024)
 - [206] Liu, X., Liu, R., Li, J., Wu, R., Wu, J.: Intcomplex for high-order interactions. *arXiv preprint arXiv:2412.02806* (2024)
 - [207] Liu, X., Su, Z., Shi, Y., Tong, Y., Wang, G., Wei, G.-W.: Manifold topological deep learning for biomedical data. *arXiv preprint arXiv:2503.00175* (2025)
 - [208] Lloyd, S., Garnerone, S., Zanardi, P.: Quantum algorithms for topological and geometric analysis of data. *Nature communications* **7**(1), 10138 (2016)
 - [209] Love, E.R., Filippenko, B., Maroulas, V., Carlsson, G.: Topological convolutional layers for deep learning. *Journal of Machine Learning Research* **24**(59), 1–35 (2023)
 - [210] Lütgehetmann, D., Govc, D., Smith, J.P., Levi, R.: Computing persistent homology of directed flag complexes. *Algorithms* **13**(1), 19 (2020)
 - [211] Maria, C., Boissonnat, J.-D., Glisse, M., Yvinec, M.: The GUDHI library: simplicial complexes and persistent homology. In: Mathematical Software—ICMS 2014: 4th International Congress, Seoul, South Korea, August 5–9, 2014. Proceedings 4, pp. 167–174 (2014). Springer

- [212] Maroulas, V., Micucci, C.P., Nasrin, F.: Bayesian topological learning for classifying the structure of biological networks. *Bayesian Analysis* **17**(3), 711–736 (2022). <https://doi.org/10.1214/21-BA1270>
- [213] Maroulas, V., Mike, J.L., Oballe, C.: Nonparametric estimation of probability density functions of random persistence diagrams. *Journal of Machine Learning Research* **20**(196), 1–49 (2019)
- [214] Maroulas, V., Nasrin, F., Oballe, C.: A Bayesian framework for persistent homology. *SIAM Journal on Mathematics of Data Science* **2**(1), 48–74 (2020). <https://doi.org/10.1137/19M1268719>
- [215] Mayer, W.: A new homology theory. *Annals of Mathematics* **43**(2), 370–380 (1942)
- [216] McCleary, J.: A User’s Guide to Spectral Sequences vol. 58. Cambridge University Press, (2001)
- [217] Mémoli, F., Wan, Z., Wang, Y.: Persistent Laplacians: Properties, algorithms and implications. *SIAM Journal on Mathematics of Data Science* **4**(2), 858–884 (2022)
- [218] Mendoza-Smith, R., Tanner, J.: Parallel multi-scale reduction of persistent homology filtrations. arXiv preprint arXiv:1708.04710 (2017)
- [219] Meng, Z., Xia, K.: Persistent spectral-based machine learning (PerSpect ML) for protein-ligand binding affinity prediction. *Science advances* **7**(19), 5329 (2021)
- [220] Meng, Z., Anand, D.V., Lu, Y., Wu, J., Xia, K.: Weighted persistent homology for biomolecular data analysis. *Scientific reports* **10**(1), 2079 (2020)
- [221] Mileyko, Y., Mukherjee, S., Harer, J.: Probability measures on the space of persistence diagrams. *Inverse Problems* **27**(12), 124007 (2011)
- [222] Miller, B.P.: Interpretations from Helmholtz’ theorem in classical electromagnetism. *American Journal of Physics* **52**(10), 948–950 (1984)
- [223] Millett, K.C., Rawdon, E.J., Stasiak, A., Sulkowska, J.I.: Identifying knots in proteins. *Biochemical Society Transactions* **41**(2), 533–537 (2013)
- [224] Mischaikow, K., Nanda, V.: Morse theory for filtrations and efficient computation of persistent homology. *Discrete & Computational Geometry* **50**, 330–353 (2013)
- [225] Mischaikow, K., Mrozek, M., Zgliczyński, P.: Conley Index Theory vol.

47. Springer, (1999)
- [226] Mitchell, E.C., Story, B., Boothe, D., Franaszczuk, P.J., Maroulas, V.: A topological deep learning framework for neural spike decoding. *Bio-physical Journal* **123**(17), 2781–2789 (2024). <https://doi.org/10.1016/j.bpj.2024.01.025>
- [227] Monod, A., Kališnik, S., Patiño-Galindo, J.Á., Crawford, L.: Tropical sufficient statistics for persistent homology. *SIAM Journal on Applied Algebra and Geometry* **3**(2), 337–371 (2019). <https://doi.org/10.1137/17M1148037>
- [228] Montagna, M., Scardapane, S., Telyatnikov, L.: Topological deep learning with state-space models: A mamba approach for simplicial complexes. arXiv preprint arXiv:2409.12033 (2024)
- [229] Morozov, D.: Dionysus, a C++ library for computing persistent homology (2007)
- [230] Morrey, C.B.: A variational method in the theory of harmonic integrals, ii. *American Journal of Mathematics* **78**(1), 137–170 (1956)
- [231] Morse, M.: Relations between the critical points of a real function of n independent variables. *Transactions of the American Mathematical Society* **27**(3), 345–396 (1925)
- [232] Munch, E., Turner, K., Bendich, P., Mukherjee, S., Mattingly, J., Harer, J.: Probabilistic Fréchet means for time varying persistence diagrams. *Electronic Journal of Statistics* **9**, 1173–1204 (2015)
- [233] Nguyen, D.D., Cang, Z., Wu, K., Wang, M., Cao, Y., Wei, G.-W.: Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges. *Journal of computer-aided molecular design* **33**, 71–82 (2019)
- [234] Nguyen, D.D., Gao, K., Wang, M., Wei, G.-W.: Mathdl: mathematical deep learning for D3R Grand Challenge 4. *Journal of computer-aided molecular design* **34**, 131–147 (2020)
- [235] Nielsen, M.A., Chuang, I.L.: *Quantum Computation and Quantum Information*. Cambridge university press, (2010)
- [236] Oballe, C., Cherne, A., Boothe, D., Kerick, S., Franaszczuk, P.J., Maroulas, V.: Bayesian topological signal processing. *Discrete and Continuous Dynamical Systems - Series S* **15**(4), 797–817 (2022). <https://doi.org/10.3934/dcdss.2021084>

- [237] Ohtsuki, T.: Quantum Invariants: a Study of Knots, 3-manifolds, and Their Sets vol. 29. World Scientific, (2001)
- [238] Ozsváth, P., Szabó, Z.: Holomorphic disks and knot invariants. *Advances in Mathematics* **186**(1), 58–116 (2004)
- [239] Panagiotou, E., Kauffman, L.H.: Knot polynomials of open and closed curves. *Proceedings of the Royal Society A* **476**(2240), 20200124 (2020)
- [240] Panagiotou, E., Plaxco, K.W.: A topological study of protein folding kinetics. *Topol. Geom. Biopolym. AMS Contemp. Math. Ser* **746**, 223–233 (2020)
- [241] Panagiotou, E., Millett, K.C., Atzberger, P.J.: Topological methods for polymeric materials: characterizing the relationship between polymer entanglement and viscoelasticity. *Polymers* **11**(3), 437 (2019)
- [242] Papamarkou, T., Nasrin, F., Lawson, A., Gong, N., Rios, O., Maroulas, V.: A random persistence diagram generator. *Statistics and Computing* **32**(88) (2022). <https://doi.org/10.1007/s11222-022-10141-y>
- [243] Papamarkou, T., Birdal, T., Bronstein, M.M., Carlsson, G., Curry, J., Gao, Y., Hajij, M., Kwitt, R., Liò, P., Di Lorenzo, P., Maroulas, V., Miolane, N., Nasrin, F., Natesan Ramamurthy, K., Rieck, B., Scardapane, S., Schaub, M.T., Veličković, P., Wang, B., Wang, Y., Wei, G.-W., Zamzmi, G.: Position: Topological deep learning is the new frontier for relational learning. *Proceedings of Machine Learning Research* **235**, 39529–39555 (2024). 41st International Conference on Machine Learning (ICML 2024)
- [244] Patel, A.: Generalized persistence diagrams. *Journal of Applied and Computational Topology* **1**(3), 397–419 (2018)
- [245] Perea, J.A.: Persistent homology of toroidal sliding window embeddings. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (icassp), pp. 6435–6439 (2016). IEEE
- [246] Perea, J.A., Harer, J.: Sliding windows and persistence: an application of topological methods to signal analysis. *Foundations of computational mathematics* **15**, 799–838 (2015)
- [247] Petri, G., Scolamiero, M., Donato, I., Vaccarino, F.: Topological strata of weighted complex networks. *PloS one* **8**(6), 66506 (2013)
- [248] Poelke, K.: Hodge-type decompositions for piecewise constant vector fields on simplicial surfaces and solids with boundary. PhD thesis (2017)

- [249] Poelke, K., Polthier, K.: Boundary-aware Hodge decompositions for piecewise constant vector fields. *Computer-Aided Design* **78**, 126–136 (2016)
- [250] Pun, C.S., Lee, S.X., Xia, K.: Persistent-homology-based machine learning: a survey and a comparative study. *Artificial Intelligence Review* **55**(7), 5169–5213 (2022)
- [251] Pun, C.S., Yong, B.Y.S., Xia, K.: Weighted-persistent-homology-based machine learning for RNA flexibility analysis. *PloS one* **15**(8), 0237747 (2020)
- [252] Puzyn, T., Leszczynski, J., Cronin, M.T.: Recent advances in QSAR studies: methods and applications (2010)
- [253] Qiu, Y., Wei, G.-W.: Persistent spectral theory-guided protein engineering. *Nature computational science* **3**(2), 149–163 (2023)
- [254] Reeb, G.: Sur les points singuliers d'une forme de pfaff completement integrable ou d'une fonction numerique [on the singular points of a completely integrable pfaff form or of a numerical function]. *Comptes Rendus Acad. Sciences Paris* **222**, 847–849 (1946)
- [255] Reininghaus, J., Huber, S., Bauer, U., Kwitt, R.: A stable multi-scale kernel for topological machine learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4741–4748 (2015)
- [256] Ren, S., Wu, J.: The stability of persistent homology of hypergraphs. *arXiv preprint arXiv:2002.02237* (2020)
- [257] Ren, S., Wu, C., Wu, J.: Weighted persistent homology. *The Rocky Mountain Journal of Mathematics* **48**(8), 2661–2687 (2018)
- [258] Ren, S., Wang, C., Wu, C., Wu, J.: On the discrete Morse functions for hypergraphs. *arXiv preprint arXiv:2108.02384* (2021)
- [259] Reuter, M., Wolter, F.-E., Peinecke, N.: Laplace-Beltrami spectra as ‘Shape-DNA’ of surfaces and solids. *Computer-Aided Design* **38**(4), 342–366 (2006)
- [260] Ribando-Gros, E., Wang, R., Chen, J., Tong, Y., Wei, G.-W.: Combinatorial and Hodge Laplacians: Similarity and difference. *SIAM Review* **66**(3), 575–601 (2024)
- [261] Russold, F.: Persistent sheaf cohomology. *arXiv preprint arXiv:2204.13446* (2022)

- [262] Rustamov, R.M., *et al.*: Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In: Symposium on Geometry Processing, vol. 257, pp. 225–233 (2007)
- [263] Schenck, H.: Algebraic Foundations for Applied Topology and Data Analysis. Springer, (2022)
- [264] Schlick, T., Zhu, Q., Dey, A., Jain, S., Yan, S., Laederach, A.: To knot or not to knot: multiple conformations of the SARS-CoV-2 frameshifting RNA element. *Journal of the American Chemical Society* **143**(30), 11404–11422 (2021)
- [265] Schubert, H.: Über eine numerische knoteninvariante. *Mathematische Zeitschrift* **61**(1), 245–288 (1954)
- [266] Schwarz, G.: Hodge Decomposition - A Method for Solving Boundary Value Problems. Springer, (2006)
- [267] Segal, G.: Classifying spaces and spectral sequences. *Publications Mathématiques de l'IHÉS* **34**, 105–112 (1968)
- [268] Shen, L., Liu, J., Wei, G.-W.: Evolutionary Khovanov homology. *AIMS Mathematics* **9**(9), 26139–26165 (2024)
- [269] Shen, L., Liu, J., Wei, G.-W.: Persistent Mayer homology and persistent Mayer Laplacian. *Foundations of Data Science* **6**(4), 584–612 (2024)
- [270] Shen, L., Feng, H., Li, F., Lei, F., Wu, J., Wei, G.-W.: Knot data analysis using multiscale Gauss link integral. *Proceedings of the National Academy of Sciences* **121**(42), 2408431121 (2024)
- [271] Shepard, A.D.: A Cellular Description of the Derived Category of a Stratified Space. Brown University, (1985)
- [272] Shonkwiler, C.: Poincaré duality angles on Riemannian manifolds with boundary. PhD thesis, University of Pennsylvania (2009)
- [273] Silver, D.S., Williams, S.G.: Knot invariants from Laplacian matrices. *Journal of Knot Theory and Its Ramifications* **28**(09), 1950058 (2019)
- [274] Singh, G., Mémoli, F., Carlsson, G.E., *et al.*: Topological methods for the analysis of high dimensional data sets and 3D object recognition. *PBG@ Eurographics* **2**, 091–100 (2007)
- [275] Singh, Y., Farrelly, C.M., Hathaway, Q.A., Leiner, T., Jagtap, J., Carlsson, G.E., Erickson, B.J.: Topological data analysis in medical imaging: current state of the art. *Insights into Imaging* **14**(1), 58 (2023)

- [276] Som, A., Choi, H., Ramamurthy, K.N., Buman, M.P., Turaga, P.: Pi-net: A deep learning approach to extract topological persistence images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 834–835 (2020)
- [277] Song, R., Li, F., Wu, J., Lei, F., Wei, G.-W.: Multi-scale jones polynomial and persistent jones polynomial for knot data analysis. *AIMS Mathematics* **10**(1), 1463–1487 (2025)
- [278] Spanier, E.H.: The Mayer homology theory. *Bull. Amer. Math. Soc.* **55**(12), 102–112 (1949)
- [279] Stolz, B.J., Harrington, H.A., Porter, M.A.: Persistent homology of time-dependent functional networks constructed from coupled time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27**(4) (2017)
- [280] Su, Z., Tong, Y., Wei, G.-W.: Hodge decomposition of single-cell RNA velocity. *Journal of chemical information and modeling* **64**(8), 3558–3568 (2024)
- [281] Su, Z., Tong, Y., Wei, G.-W.: Hodge decomposition of vector fields in Cartesian grids. In: SIGGRAPH Asia 2024 Conference Papers, pp. 1–10 (2024)
- [282] Su, Z., Tong, Y., Wei, G.-W.: Persistent de Rham-Hodge Laplacians in Eulerian representation for manifold topological learning. *AIMS Mathematics* **9**(10), 27438–27470 (2024)
- [283] Su, Z., Tong, Y., Wei, G.-W.: Topology-preserving Hodge decomposition in the Eulerian representation. arXiv preprint arXiv:2408.14356 (2024)
- [284] Sulkowska, J.I., Rawdon, E.J., Millet, K.C., Onuchic, J.N., Stasiak, A.: Conservation of complex knotting and slipknotting patterns in proteins. *Biophysical Journal* **102**(3), 253 (2012)
- [285] Sumners, D.: The role of knot theory in DNA research. In: *Geometry and Topology*, pp. 297–318. CRC Press, (2020)
- [286] Suwayyid, F., Wei, G.-W.: Persistent Dirac of paths on digraphs and hypergraphs. *Foundations of data science (Springfield, Mo.)* **6**(2), 124 (2024)
- [287] Suwayyid, F., Wei, G.-W.: Persistent Mayer Dirac. *Journal of Physics: Complexity* **5**(4), 045005 (2024)
- [288] Takens, F.: Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium*

- Held at the University of Warwick 1979/80, pp. 366–381 (2006). Springer
- [289] Tauzin, G., Lupo, U., Tunstall, L., Pérez, J.B., Caorsi, M., Medina-Mardones, A.M., Dassatti, A., Hess, K.: giotto-tda:: A topological data analysis toolkit for machine learning and data exploration. *Journal of Machine Learning Research* **22**(39), 1–6 (2021)
 - [290] Tierny, J., Favelier, G., Levine, J.A., Gueunet, C., Michaux, M.: The topology toolkit. *IEEE transactions on visualization and computer graphics* **24**(1), 832–842 (2017)
 - [291] Townsend, J., Micucci, C.P., Hymel, J.H., Maroulas, V., Vogiatzis, K.D.: Representation of molecular structures with persistent homology for machine learning applications in chemistry. *Nature communications* **11**(1), 3230 (2020)
 - [292] Turner, K., Robins, V., Morgan, J.: The extended persistent homology transform of manifolds with boundary. *Journal of Applied and Computational Topology* **8**(7), 2111–2154 (2024)
 - [293] Turner, K., Mileyko, Y., Mukherjee, S., Harer, J.: Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry* **52**(1), 44–70 (2014)
 - [294] Van Veen, H.J., Saul, N., Eargle, D., Mangham, S.W.: Kepler Mapper: a flexible Python implementation of the Mapper algorithm. *Journal of Open Source Software* **4**(42), 1315 (2019)
 - [295] Vietoris, L.: Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen. *Mathematische Annalen* **97**(1), 454–472 (1927)
 - [296] Vipond, O.: Multiparameter persistence landscapes. *Journal of Machine Learning Research* **21**(61), 1–38 (2020)
 - [297] Wagner, H.: Slice, simplify and stitch: Topology-preserving simplification scheme for massive voxel data. In: 39th International Symposium on Computational Geometry (SoCG 2023) (2023). Schloss Dagstuhl-Leibniz-Zentrum für Informatik
 - [298] Wagner, H., Chen, C., Vuçini, E.: Efficient computation of persistent homology for cubical data. In: *Topological Methods in Data Analysis and Visualization II: Theory, Algorithms, and Applications*, pp. 91–106. Springer, (2011)
 - [299] Wang, B., Wei, G.-W.: Object-oriented persistent homology. *Journal of computational physics* **305**, 276–299 (2016)

- [300] Wang, R., Wei, G.-W.: Persistent path Laplacian. Foundations of data science (Springfield, Mo.) **5**(1), 26 (2023)
- [301] Wang, R., Chen, J., Wei, G.-W.: Mechanisms of SARS-CoV-2 evolution revealing vaccine-resistant mutations in Europe and America. The journal of physical chemistry letters **12**(49), 11850–11857 (2021)
- [302] Wang, R., Nguyen, D.D., Wei, G.-W.: Persistent spectral graph. International Journal for Numerical Methods in Biomedical Engineering **36**(9), 3376 (2020)
- [303] Wang, R., Zhao, R., Ribando-Gros, E., Chen, J., Tong, Y., Wei, G.-W.: HERMES: Persistent spectral graph software. Foundations of Data Science **3**(1), 67–97 (2021)
- [304] Wang, S., Chern, A.: Computing minimal surfaces with differential forms. ACM Transactions on Graphics (TOG) **40**(4), 1–14 (2021)
- [305] Wang, Y., Liu, X., Zhang, Y., Wang, X., Xia, K.: Join persistent homology (JPH)-based machine learning for metalloprotein–ligand binding affinity prediction. Journal of Chemical Information and Modeling (2025)
- [306] Warner, F.W.: Foundations of Differentiable Manifolds and Lie Groups vol. 94. Springer, (1983)
- [307] Wee, J., Xia, K.: Persistent spectral based ensemble learning (PerSpect-EL) for protein-protein binding affinity prediction. Briefings in Bioinformatics **23**(2), 024 (2022)
- [308] Wee, J., Bianconi, G., Xia, K.: Persistent Dirac for molecular representation. Scientific Reports **13**(1), 11183 (2023)
- [309] Wei, X., Wei, G.-W.: Persistent sheaf Laplacians. Foundations of Data Science **7**(2), 446–463 (2025)
- [310] Wei, X., Wei, G.-W.: Persistent topological Laplacians – a survey. Mathematics **13**(2), 208 (2025)
- [311] Weil, A.: Numbers of solutions of equations in finite fields. Bulletin of the American Mathematical Society **55**(5), 497–508 (1949)
- [312] Wu, S., Liu, X., Dong, A., Gragnoli, C., Griffin, C., Wu, J., Yau, S.-T., Wu, R.: The metabolomic physics of complex diseases. Proceedings of the National Academy of Sciences **120**(42), 2308496120 (2023)
- [313] Xia, K., Wei, G.-W.: Persistent homology analysis of protein structure, flexibility, and folding. International journal for numerical methods in

- biomedical engineering **30**(8), 814–844 (2014)
- [314] Xia, K., Wei, G.-W.: Multidimensional persistence in biomolecular data. *Journal of computational chemistry* **36**(20), 1502–1520 (2015)
 - [315] Xia, K., Zhao, Z., Wei, G.-W.: Multiresolution persistent homology for excessively large biomolecular datasets. *The Journal of chemical physics* **143**(13) (2015)
 - [316] Xia, K., Zhao, Z., Wei, G.-W.: Multiresolution topological simplification. *Journal of Computational Biology* **22**(9), 887–891 (2015)
 - [317] Xia, K., Feng, X., Tong, Y., Wei, G.W.: Persistent homology for the quantitative prediction of fullerene stability. *Journal of computational chemistry* **36**(6), 408–422 (2015)
 - [318] Yang, S., Xiong, S., Zhang, Y., Feng, F., Liu, J., Zhu, B.: Clebsch gauge fluid. *ACM Transactions on Graphics (TOG)* **40**(4), 1–11 (2021)
 - [319] Yegnesh, K.: Persistence and sheaves. arXiv preprint arXiv:1612.03522 (2016)
 - [320] Yin, H., Nabizadeh, M.S., Wu, B., Wang, S., Chern, A.: Fluid cohomology. *ACM Trans. Graph.* **42**(4) (2023)
 - [321] Yoon, H.R.: Cellular sheaves and cosheaves for distributed topological data analysis. PhD thesis, University of Pennsylvania (2018)
 - [322] Zhang, M.M., Li, J.Y., Wu, J.: The twisted homology of simplicial set. *Acta Mathematica Sinica, English Series* **38**(10), 1781–1802 (2022)
 - [323] Zhao, R., Cang, Z., Tong, Y., Wei, G.-W.: Protein pocket detection via convex hull surface evolution and associated reeb graph. *Bioinformatics* **34**(17), 830–837 (2018)
 - [324] Zhao, R., Desbrun, M., Wei, G.-W., Tong, Y.: 3D Hodge decompositions of edge- and face-based vector fields. *ACM Transactions on Graphics (TOG)* **38**(6), 1–13 (2019)
 - [325] Zhao, R., Wang, M., Chen, J., Tong, Y., Wei, G.-W.: The de Rham–Hodge analysis and modeling of biomolecules. *Bulletin of mathematical biology* **82**, 1–38 (2020)
 - [326] Zheng, J., Feng, Z., Ekstrom, A.D.: Towards analysis of multivariate time series using topological data analysis. *Mathematics* **12**(11), 1727 (2024)
 - [327] Zhou, Q.: Pymesh-geometry processing library for Python. Software available for download at <https://github.com/PyMesh/PyMesh>

(2019)

- [328] Zhou, Y., Chalapathi, N., Rathore, A., Zhao, Y., Wang, B.: Mapper interactive: a scalable, extendable, and interactive toolbox for the visual exploration of high-dimensional data. In: 2021 IEEE 14th Pacific Visualization Symposium (PacificVis), pp. 101–110 (2021). IEEE
- [329] Zia, M., Jones, B., Feng, H., Wei, G.-W.: Persistent directed flag Laplacian (PDLF)-based machine learning for protein–ligand binding affinity prediction. *Journal of Chemical Theory and Computation* **21**(8), 4276–4285 (2025)
- [330] Zieliński, B., Lipiński, M., Juda, M., Zeppelzauer, M., Dłotko, P.: Persistence bag-of-words for topological data analysis. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence, pp. 4489–4495 (2019)
- [331] Zieliński, B., Lipiński, M., Juda, M., Zeppelzauer, M., Dłotko, P.: Persistence codebooks for topological data analysis. *Artificial Intelligence Review* **54**, 1969–2009 (2021)
- [332] Zomorodian, A.: Fast construction of the Vietoris-Rips complex. *Computers & Graphics* **34**(3), 263–271 (2010)
- [333] Zomorodian, A., Carlsson, G.: Computing persistent homology. *Discrete & Computational Geometry* **33**(2), 249–274 (2005)
- [334] Zomorodian, A., Carlsson, G.: Localized homology. *Computational Geometry* **41**(3), 126–148 (2008)