

PERSISTENT TOPOLOGICAL LAPLACIANS AND THEIR APPLICATIONS

By

Xiaoqi Wei

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Mathematics—Doctor of Philosophy

2024

ABSTRACT

Topological and geometrical methods are known for their capability to reduce noise and have achieved significant success in analyzing complex biological data. A key method in topological data analysis is persistent homology, which leverages a filtration of simplicial complexes to extract multiscale spatial information. To integrate non-spatial information, specially tailored persistent homology approaches, such as element-specific persistent homology, have been proposed and have shown significant success in predictive modeling of molecular structures.

Recently, it was discovered that persistent Laplacians can be defined for a filtration, and the nullity of a persistent Laplacian is equal to the corresponding persistent Betti number, suggesting that the spectra of persistent Laplacians offer additional information beyond traditional persistent homology. Spectra of persistent Laplacians can be used in combination with persistent homology to enhance the featurization of raw biological data. Inspired by the theory of cellular sheaves, the theory of persistent sheaf Laplacians was proposed; spectra of persistent sheaf Laplacians encode both spatial and non-spatial information of a labeled point cloud. The theory of persistent sheaf Laplacians provides an elegant method for fusing different types of data and holds significant potential for future development.

The construction of persistent Laplacians can also be easily generalized to other settings, such as digraphs and hypergraphs. These generalizations are important, as they offer various ways to integrate different types of biological information. In this thesis, we introduce persistent Laplacians and some generalizations, such as persistent sheaf Laplacians, and discuss their applications in biology.

Copyright by
XIAOQI WEI
2024

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my PhD supervisor, Guo-Wei Wei, for his invaluable guidance, continuous support, and for leading me onto the path of applied mathematics. His mentorship has been fundamental to my growth as a researcher. I would also like to extend my thanks to my committee members, Ekaterina Rapinchuk, Moxun Tang, and Yiyong Tong, for their support and guidance throughout my PhD study.

I am also grateful to my research group members, in particular Dong Chen, Jiahui Chen, Kaifu Gao, Shen Li, Rui Wang, and Menglun Wang, for their valuable discussions, insights, and assistance throughout my PhD journey. Their collaboration has been an essential part of my research experience.

I wish to acknowledge the professors who have greatly influenced my academic journey. During my master's studies, I would like to thank Qingtao Chen for his guidance and support in my research. I am also grateful to Dong Tian, Bingzhe Hou, and Wenjie Gao for their guidance during my undergraduate years.

My heartfelt thanks go to my friends from Changchun, Zurich, and East Lansing—particularly Wenxian, Zheng, Hantian, Haozhi, Honghao, and Lin-Da—who have made my life brighter and more colorful throughout my studies.

A special acknowledgment goes to Mr. Wei Wang, who sparked my interest in mathematics and nurtured my passion for the subject from a young age.

Lastly, I would like to extend my deepest appreciation to my family. In particular, I am profoundly grateful to my parents and my grandmother for their endless encouragement and love.

TABLE OF CONTENTS

CHAPTER 1	INTRODUCTION	1
CHAPTER 2	PERSISTENT LAPLACIANS	4
CHAPTER 3	(CO)SHEAVES AND PERSISTENT SHEAF LAPLACIANS	27
CHAPTER 4	OTHER PERSISTENT TOPOLOGICAL LAPLACIANS	44
CHAPTER 5	UNDERSTANDING DOMINANT VARIANTS OF SARS-COV-2 USING LAPLACIANS	52
CHAPTER 6	THESIS CONTRIBUTION AND FUTURE WORK	71
BIBLIOGRAPHY	73
APPENDIX A	STATISTICS OF NONZERO EIGENVALUES OF PERSISTENT LAPLACIANS	84
APPENDIX B	METHOD OF HOMOTOPY CONTINUATION	88
APPENDIX C	APPENDIX OF THE SHEAF CHAPTER	91
APPENDIX D	APPENDIX OF THE SARS-COV-2 CHAPTER	96

CHAPTER 1

INTRODUCTION

In recent years, advancements in experimental techniques in biological research have generated vast amounts of data requiring extensive analysis. A biological object can often be viewed as a geometrical object in various ways; for instance, a protein's atoms form a point cloud, its molecular surface can be seen as a manifold, and its polypeptides may be treated as knots or links. Since certain types of biological data possess spatial information, topological and geometrical methods have shown great potential in analyzing them [17, 40, 48, 78, 137, 146].

One prime example of applying topological methods in biology is the use of persistent homology in predictive modeling of molecular structures [16, 18, 19, 20, 144, 145]. A fundamental principle in molecular biology is that structure dictates function. Since critical structural information often varies depending on the problem and dataset, using machine learning to capture this information is more effective. When employing supervised learning methods, such as gradient boosting trees, the challenge lies in featurization, i.e., mapping a high-dimensional raw molecular structure to a low-dimensional space while preserving sufficient structural information. As molecules can be naturally viewed as point cloud data, persistent homology provides a solution to this challenge. The basic idea of persistent homology is to create a multiscale family of simplicial complexes (referred to as a filtration) from the molecule and describe the evolution of these simplicial complexes. For any two simplicial complexes $X \subset Y$ in the filtration, the persistent homology group $H_i(X, Y)$ can be calculated, and the collection of all persistent homology groups describes the shape evolution of the simplicial complexes and provides a multiscale and low-dimensional topological characterization of the point cloud.

It is important to remember that biological data are not purely geometrical, and successful applications of topological or geometrical methods must incorporate non-spatial information. A good mathematical representation of biological data should capture the spatial information

and the non-spatial information at the same time. For example, when studying molecular structures, to extract finer structural information, we can use only atoms of certain element types or of certain roles when performing persistent homology analysis (this approach is referred to as the element specific persistent homology [20]). If Vietoris-Rips filtration is employed, we can also modify the distance matrix to emphasize specific types of interactions between atoms.

The major theme of this thesis is the persistent Laplacian approach, proposed to complement persistent homology and its generalizations that integrate various non-spatial information. Laplacians are ubiquitous in science and engineering and are deeply connected to homology theories. In graph theory, it is known that the nullity of the graph Laplacian is equal to the number of connected components, and the smallest nonzero eigenvalue, called the Fiedler value, reflects the graph's connectivity. For simplicial complexes, the combinatorial Laplacian [43] is defined for each dimension on a simplicial complex, and we can prove that the kernel of a combinatorial Laplacian is isomorphic to the corresponding simplicial homology group. On a differentiable manifold, de Rham-Hodge theory states that the kernel of a Hodge Laplacian is isomorphic to the corresponding de Rham cohomology group. Thus, combinatorial Laplacians are discretizations of Hodge Laplacians. Another discretization of Hodge Laplacians can be achieved through discrete exterior calculus [42], and a multiscale formulation of Hodge Laplacians on manifolds was introduced by Chen et al. [32] to study manifold-type data. For knot-type data, multiscale Laplacians also exist in the context of Khovanov homology [120].

Persistent Laplacians [84, 133] are the counterparts of combinatorial Laplacians in the context of persistent homology. It has been suggested that the spectra of persistent Laplacians not only retain information from persistent homology but also provide additional spatial information¹. In the most general sense, any method that utilizes multiscale Laplacians to quantitatively analyze data can be referred to as a persistent Laplacian approach. The the-

¹However, beyond simple cases, making explicit statements about the relationship between shape and spectrum is often challenging.

ory of persistent Laplacians has been studied extensively in [63, 88, 96]. Algorithms have been developed for computing persistent Laplacians [96, 135], and these Laplacians have been applied to protein-ligand binding prediction [97], SARS-CoV-2 research [28, 139], and protein engineering [105]. To integrate non-spatial information, persistent Laplacians have been extended to various mathematical settings, such as cellular sheaves [141], flag complexes [75], digraphs [134], hypergraphs [90], and hyperdigraphs [24].

This thesis is organized as follows. In Chapter 2, we introduce the theory of persistent homology and persistent Laplacians. In Chapter 3, we develop the theory of persistent sheaf Laplacians, which can be applied to analyze labeled point clouds. In Chapter 4, we review some recent developments concerning persistent Laplacians. In Chapter 5, we demonstrate the application of persistent Laplacians in COVID-19 research.

CHAPTER 2

PERSISTENT LAPLACIANS

In this chapter we first present a very brief introduction to persistent homology theory and then give the definition of a persistent Laplacian. [99] is a very good reference on simplicial complexes. [44] contains an introduction to basic concepts of topology from a computational perspective.

Simplicial complexes and combinatorial Laplacians

Definition 2.0.1. A q -simplex, denoted as $\sigma_q = \{u_0, \dots, u_q\}$, is the convex hull of $q + 1$ affinely independent points u_0, \dots, u_q in \mathbb{R}^n . The orientation of σ_q is determined by the ordering of the vertices and two orderings define the same orientation if and only if they differ by an even permutation. The dimension of $\sigma_q = \{u_0, \dots, u_q\}$ is defined as q . For $0 \leq i \leq n$, $\{u_0, \dots, \hat{u}_i, \dots, u_q\}$ is said to be a face of σ_q , where the hat indicates the omission of the vertex u_i .

Definition 2.0.2. A finite set of simplices, X , is a simplicial complex if the following conditions are satisfied: (1) all faces of any simplex in X are also in X ; (2) the non-empty intersection of any two simplices in X is a common face of the two simplices. The dimension of a simplicial complex X is defined as the maximal dimension of its simplices.

Definition 2.0.3. Given a finite set V , an *abstract simplicial complex* X is a collection of subsets of V , such that if a set σ is in X , then any subset of σ is also in X . A set σ that consists of $q + 1$ elements is referred to as a q -simplex. If σ is a subset of τ , then we say that σ is a face of τ and denote the face relation by $\sigma \leq \tau$. If X and Y are abstract simplicial complexes and $X \subset Y$, then X is referred to as a subcomplex of Y .



Figure 2.1 Illustrations of simplices.

The concepts of abstract simplicial complexes and simplicial complexes are closely related. One may build a simplicial complex from an abstract simplicial complex or vice versa [44]. They contain exactly the same combinatorial information. From now on we will not distinguish abstract simplicial complexes from simplicial complexes.

Example 2.0.1. A simple graph $G = (V, E)$ can be seen as a simplicial complex, since each edge $\{v_i, v_j\} \in E$ is a subset of the vertex set V .

Example 2.0.2. Suppose V is a finite set of points in \mathbb{R}^n . Given a real number d , a *Rips complex* X_d can be defined as follows. A set $\sigma = \{v_{a_0}, \dots, v_{a_q}\} \in X_d$ if and only if the Euclidean distance $\|v_{a_i}, v_{a_j}\| \leq d$ for any pair of points v_{a_i} and v_{a_j} . We can see that $X_s \subset X_t$ if $s \leq t$.

Example 2.0.3. Given a finite set of points V in \mathbb{R}^n , we can also build an *Alpha complex*. First we define the Voronoi cell. The Voronoi cell of a point u in V is

$$V_u = \{x \in \mathbb{R}^n \mid \|x - u\| \leq \|x - v\|, v \in V\}.$$

Let $B_u(r)$ be the closed ball with center u and radius r . Denote the intersection $B_u(r) \cap V_u$ by $R_u(r)$. Then the Alpha complex $\text{Alpha}(r)$ is defined by

$$\{\sigma \subset V \mid \bigcap_{u \in \sigma} R_u(r) \neq \emptyset\}.$$

In other words, $\text{Alpha}(r)$ is the nerve of cover $\{R_u(r), u \in V\}$. It is also true that $\text{Alpha}(r_1) \subset \text{Alpha}(r_2)$ if $r_1 \leq r_2$.

Each simplicial complex has an associated algebraic structure called the *simplicial chain complex*. For the sake of simplicity, we will designate a fixed global ordering of vertices in a simplicial complex, and require that vertices of any simplices should be ordered according to the fixed ordering¹. For example, suppose we use the natural ordering $0 < 1 < 2$ for the simplicial complex $\{\{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}\}$ (Figure 2.2b), then we must not

¹A fixed ordering is not necessary. The reader can find more information in [99].

write the simplex $\{0, 1\}$ as $\{1, 0\}$. To emphasize that a simplex $\{v_0, \dots, v_q\}$ is ordered, we will use notation $[v_0, \dots, v_q]$ or $v_0 \dots v_q$.

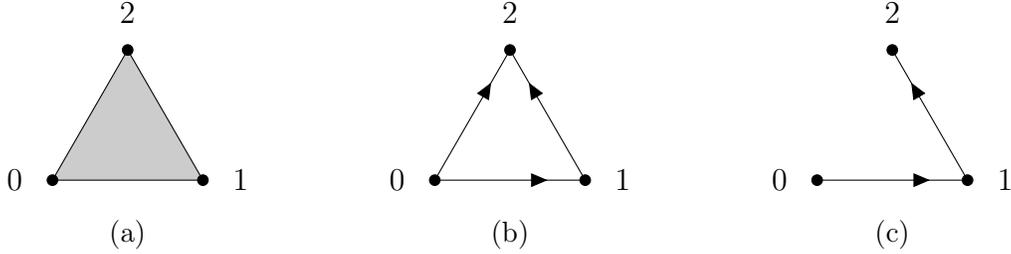


Figure 2.2 (a) The simplicial complex $\{\{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}$. (b) The oriented simplicial complex $\{0, 1, 2, 01, 02, 12\}$. Arrows emphasize that vertices are ordered. (c) The oriented simplicial complex $\{0, 1, 2, 01, 12\}$.

Definition 2.0.4. A simplicial chain complex consists of a sequence of real vector space $C_q(X)$ and a sequence of linear homomorphisms ∂_q between them. It is usually written out as follows:

$$\dots \xrightarrow{\partial_3} C_2(X) \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \longrightarrow 0$$

where the real vector space $C_q(X)$ is generated by q -simplices. An element of $C_q(X)$ is called a q -chain, and we can actually represent a q -chain by a function f_q whose domain is the set of q -simplices. The boundary operator ∂_q is a linear map such that

$$\partial_q[v_{a_0}, \dots, v_{a_q}] = \sum_i (-1)^i [v_{a_0}, \dots, \hat{v}_{a_i}, \dots, v_{a_q}].$$

Here the symbol \hat{v}_{a_i} means that \hat{v}_{a_i} is deleted. The ordering of vertices ensures that the boundary operator is well-defined.

It is well known that $\partial_q \partial_{q+1} = 0$, so the q -th homology group $H_q = \ker \partial_q / \text{im } \partial_{q+1}$ is well-defined. The dimension of the homology group H_q is referred to as the q -th Betti number.

Example 2.0.4. At least for some simple simplicial complexes, the q -th Betti number counts the number of q -dimensional holes. The simplicial complex $X = \{0, 1, 2, 01, 02, 12\}$ (Figure

2.2b) has only two chain groups C_0 and C_1 , and one boundary map ∂_1 which can be represented by the matrix

$$\begin{array}{c} 01 \quad 12 \quad 02 \\ 0 \\ 1 \\ 2 \end{array} \begin{pmatrix} -1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

Recall that we can identify a q -chain $\sum_{\sigma} c_{\sigma} \sigma$ with a function f_q such that $f_q(\sigma) = c_{\sigma}$. The operator ∂_1 maps a real-valued function $f_1 : \{01, 12, 02\} \rightarrow \mathbb{R}$ to a function $f_0 : \{0, 1, 2\} \rightarrow \mathbb{R}$ where $f_0(0) = -f_1(01) - f_1(02)$, $f_0(1) = f_1(01) - f_1(12)$, $f_0(2) = f_1(12) + f_1(02)$. Since $C_2 = 0$, the homology group H_1 is $\ker \partial_1$ and we can verify that $f_1 \in H_1(X)$ implies $f_1(01) = -f_1(02) = f_1(12)$. In other words, the 1-th Betti number is 1.

For the simplicial complex $Y = \{0, 1, 2, 01, 12\}$ (Figure 2.2c), the matrix representation of ∂_1 is

$$\begin{array}{c} 01 \quad 12 \\ 0 \\ 1 \\ 2 \end{array} \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}$$

and we can verify that the only $f_1 \in C_1(Y)$ that satisfies $\partial_1 f_1 = 0$ is the zero function. The intuition behind the difference of $H_1(X)$ and $H_1(Y)$ is that, in X the edges $\{01, 12, 02\}$ constitute a close path, while in Y there are no close paths.

Generally, a sequence of abelian groups and group morphisms

$$\dots \xrightarrow{\partial_{q+2}} A_{q+1} \xrightarrow{\partial_{q+1}} A_q \xrightarrow{\partial_q} A_{q-1} \xrightarrow{\partial_{q-1}} \dots$$

where $\partial_q \partial_{q+1} = 0$ holds, is referred to as a *chain complex*. The q -th homology group $H_q = \ker \partial_q / \text{im } \partial_{q+1}$ is readily defined. If each chain group A_q is an inner product space, the q -th

combinatorial Laplacian or Hodge Laplacian $\Delta_q : A_q \rightarrow A_q$ [43] is defined by

$$\partial_{q+1}\partial_{q+1}^\dagger + \partial_q^\dagger\partial_q,$$

where ∂_q^\dagger denotes the adjoint of ∂_q . The q -th combinatorial Laplacian Δ_q is a positive semi-definite symmetric operator and therefore only has non-negative eigenvalues. We often call $\partial_{q+1}\partial_{q+1}^\dagger$ the q -th up Laplacian and denote it by $\Delta_{q,+}$, and $\partial_q^\dagger\partial_q$ the q -th down Laplacian and denote it by $\Delta_{q,-}$. A sequence of abelian groups and group morphisms

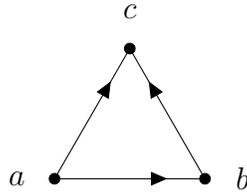
$$\dots \xrightarrow{\partial^{q-2}} A^{q-1} \xrightarrow{\partial^{q-1}} A^q \xrightarrow{\partial^q} A^{q+1} \xrightarrow{\partial^{q+1}} \dots$$

where $\partial^q\partial^{q-1} = 0$, is called a *cochain complex*. The q -th cohomology group is $\ker \partial^q / \text{im } \partial^{q-1}$, and the q -th combinatorial Laplacian for a cochain complex is defined analogously when each cochain group is an inner product space.

Example 2.0.5. The graph Laplacian \mathcal{L} of a graph $G = (V, E)$ is usually defined element-wise by

$$\mathcal{L}_{ij} = \begin{cases} \deg v_i, & \text{if } i = j \\ -1, & \text{if } i \neq j \text{ and } v_i \text{ is adjacent to } v_j \\ 0, & \text{otherwise} \end{cases}$$

The graph Laplacian \mathcal{L} actually coincides with the matrix representation of 0-th combinatorial Laplacian Δ_0 . Take the graph $\{[a, b], [b, c], [c, a]\}$ as an example.



Its graph Laplacian is

$$\begin{array}{c} a \quad b \quad c \\ a \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \\ b \\ c \end{array}.$$

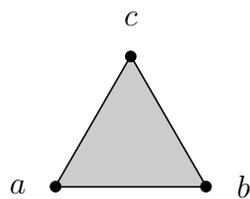
The matrix representation of ∂_1 is

$$\begin{array}{c} ab \quad bc \quad ac \\ a \begin{pmatrix} -1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \\ b \\ c \end{array},$$

and

$$\begin{pmatrix} -1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

Example 2.0.6. Consider the 2-simplex shown below. We compute its Δ_1 .



The matrix representation of ∂_1 is

$$\begin{array}{c} ab \quad bc \quad ac \\ a \begin{pmatrix} -1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \\ b \\ c \end{array},$$

and the matrix representation of ∂_2 is

$$\begin{matrix} & abc \\ ab & \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \\ bc & \\ ac & \end{matrix}.$$

The matrix representation of $\Delta_1 = \partial_2 \partial_2^\dagger + \partial_1^\dagger \partial_1$ is

$$\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \end{pmatrix} + \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Definition 2.0.5. Two q -simplices σ_i and σ_j are said to be lower adjacent, denoted by $\sigma_i \stackrel{L}{\sim} \sigma_j$, if they share a common $(q-1)$ -face. They are said to be upper adjacent, denoted by $\sigma_i \stackrel{U}{\sim} \sigma_j$, if they both are faces of a $(q+1)$ -simplex. The lower degree $\deg_L(\sigma)$ of a q -simplex σ is $q+1$, the number of its $(q-1)$ -faces. The upper degree $\deg_U(\sigma)$ of a q -simplex σ is defined as the number of $(q+1)$ -simplices in K of which σ is a face. The degree of q -simplex σ is defined by

$$\deg_U(\sigma) + \deg_L(\sigma) = \deg_U(\sigma) + q + 1.$$

Now suppose $\sigma_i \stackrel{U}{\sim} \sigma_j$ with a common upper $(q+1)$ -simplex τ . Let's examine the signs of the coefficients of σ_i and σ_j in the boundary $\partial\tau$ of τ . We say that σ_i and σ_j are similarly oriented if the signs of the coefficients of σ_i and σ_j in $\partial\tau$ are the same; They are dissimilarly oriented if the signs are different.

Next suppose $\sigma_i \stackrel{L}{\sim} \sigma_j$ with common lower $(q-1)$ -simplex η . Let's examine the signs of the coefficients of η in $\partial\sigma_i$ and $\partial\sigma_j$. We say that η is a similar common lower simplex of σ_i and σ_j if the signs are the same; η is a dissimilar common lower simplex if the signs are different.

The following proposition gives an explicit formula for \mathcal{L}_q , the matrix representation of Δ_q .

Proposition 2.0.1. [52, Theorem. 3.3.4] Suppose we have a finite simplicial complex and its set of q -simplices is $\{\sigma_1, \dots, \sigma_n\}$.

(1) When $q = 0$,

$$\mathcal{L}_{ij} = \begin{cases} \deg_U \sigma_i, & \text{if } i = j \\ -1, & \text{if } i \neq j \text{ and } \sigma_i \overset{U}{\sim} \sigma_j \\ 0, & \text{if } i \neq j \text{ and } \sigma_i \overset{U}{\not\sim} \sigma_j \end{cases}$$

(2) If $q > 0$, then

$$\mathcal{L}_{ij} = \begin{cases} \deg_U \sigma_i + q + 1, & \text{if } i = j \\ 1, & \text{if } i \neq j, \sigma_i \overset{U}{\sim} \sigma_j \text{ and have a similar} \\ & \text{common lower simplex} \\ -1, & \text{if } i \neq j, \sigma_i \overset{U}{\not\sim} \sigma_j \text{ and have a dissimilar} \\ & \text{common lower simplex} \\ 0, & \text{if } i \neq j \text{ and either } \sigma_i \text{ and } \sigma_j \text{ are upper} \\ & \text{adjacent or are not lower adjacent} \end{cases}$$

Next we are going to show that the kernel of Δ_q is isomorphic to the q -th homology group. We first need the following lemma.

Proposition 2.0.2. [85] If U, V, W are finite-dimensional inner product spaces and $f : U \rightarrow V$, $g : V \rightarrow W$ are two linear morphisms such that $gf = 0$,

$$U \xrightarrow{f} V \xrightarrow{g} W$$

then $\ker(g^\dagger g + f f^\dagger) \cong \ker g / \text{im } f$, and $V = \text{im } g^\dagger \oplus \ker(g^\dagger g + f f^\dagger) \oplus \text{im } f$.

Proof. We give the reader an outline of the proof. Since $\langle g^\dagger g + f f^\dagger v, v \rangle = \langle gv, gv \rangle + \langle f^\dagger v, f^\dagger v \rangle$, $v \in \ker(g^\dagger g + f f^\dagger)$ is equivalent to $v \in \ker g \cap \ker f^\dagger$. Since $\ker f^\dagger$ is orthogonal to $\text{im } f$, and $\text{im } f \subset \ker g$, we know that $\ker(g^\dagger g + f f^\dagger)$ is isomorphic to the orthogonal complement of $\text{im } f$ in $\ker g$. □

The condition that $v \perp \text{im } f$ implies that $\langle v, v \rangle \leq \langle v + fu, v + fu \rangle$ for any $u \in U$. Conversely, if $\langle v + fu, v + fu \rangle - \langle v, v \rangle = 2\langle v, fu \rangle + \langle fu, fu \rangle \geq 0$ for any $u \in U$, $\langle v, fu \rangle$ must be 0, otherwise we can always multiply a coefficient k to u to make the sum $2k\langle v, fu \rangle + k^2\langle fu, fu \rangle < 0$. Therefore, for a equivalent class $v' + \text{im } f$ in $\ker g / \text{im } f$, its corresponding element v in $\ker(g^\dagger g + ff^\dagger)$ is the element that minimizes its ‘size’ $\langle v, v \rangle$.

Each chain group of a simplicial chain complex has a canonical inner product structure. We can just let $\langle \sigma, \tau \rangle = 0$ if $\sigma \neq \tau$, and $\langle \sigma, \tau \rangle = 1$ if $\sigma = \tau$. Therefore the q -th combinatorial Laplacian is readily defined for a simplicial complex. Apply Proposition 2.0.2 to a simplicial chain complex, we know that the kernel of the q -th combinatorial Laplacian Δ_q is isomorphic to the q -th homology group H_q [43] (sometimes called the Hodge theorem), and A_q admits a *Hodge decomposition*

$$C_q(X) = \text{im } \partial_q^\dagger \oplus \ker \Delta_q \oplus \text{im } \partial_{q+1}.$$

Proposition 2.0.3. *The nonzero spectrum of Δ_q is the union of the nonzero spectra of $\Delta_{q,+}$ and $\Delta_{q,-}$.*

Proof. This is derived from the Hodge decomposition. Since $\Delta_q|_{\text{im } \partial_q^\dagger} = \Delta_{q,+}|_{\text{im } \partial_q^\dagger} \subset \text{im } \partial_q^\dagger$ and $\Delta_q|_{\text{im } \partial_{q+1}} = \Delta_{q,-}|_{\text{im } \partial_{q+1}} \subset \text{im } \partial_{q+1}$, Δ_q is indeed the orthogonal direct sum

$$0|_{\ker \Delta_q} \oplus \Delta_{q,+}|_{\text{im } \partial_q^\dagger} \oplus \Delta_{q,-}|_{\text{im } \partial_{q+1}}.$$

□

The operator $\Delta_0 = \partial_1 \partial_1^\dagger$ is more commonly known as the *graph Laplacian*, and there is a vast amount of work regarding the relation between the spectrum of graph Laplacian and the shape of a graph [37]. For a connected graph, it is well known that the minimal nonzero eigenvalue of the graph Laplacian reflects the graph connectivity [47]. Note that graphs that have the same homology groups may have different graph Laplacians (Figure 2.3).

The intuition behind the Hodge theorem when $q = 0$ is as follows. For a simple graph (V, E) , let f_0 be a function that sends every vertex to a real number. If we view the simple

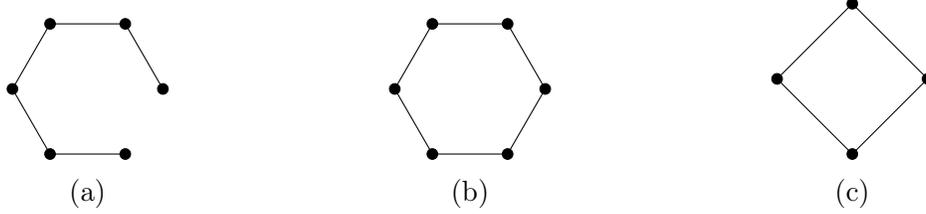


Figure 2.3 Simplicial homology can distinguish (a) from (b) and (c), but cannot distinguish between (b) and (c). The graph Laplacian can distinguish among all of them. Indeed for a cycle graph with n vertices, the spectrum of the graph Laplacian is $\{2 - 2 \cos(2k\pi/n) \mid k = 1, \dots, n\}$.

graph as a simplicial complex, ∂_1^\dagger maps f_0 to a real valued function whose domain is E . The sum

$$\sum_{ij \in E} |f_0(i) - f_0(j)|^2 = \langle \partial_1^\dagger f_0, \partial_1^\dagger f_0 \rangle = \langle f_0, \partial_1 \partial_1^\dagger f_0 \rangle$$

is called the *Dirichlet energy* of f_0 , and it measures how f_0 varies over V . Any $f_0 \in \ker \Delta_0 = \ker \partial_1 \partial_1^\dagger$ is a function with zero Dirichlet energy. For a connected graph, if f_0 has zero Dirichlet energy then $f_0(a) = f_0(b)$ for any two vertices a and b , because there is always a path that starts from a and ends at b . In other words, f_0 is a constant function. If a graph has more than one connected components, f_0 only needs to be locally constant on any connected components. In other words, the dimension of $\ker \Delta_0$ is equal to the number of connected components, which is also the dimension of H_0 .

Persistent homology

Now suppose X is a subcomplex of Y . Then the q -th chain group of X is a subspace of the q -th chain group of Y , as shown in the following diagram

$$\begin{array}{ccccccc} \dots & \xrightarrow{\partial_{q+2}} & C_{q+1}(X) & \xrightarrow{\partial_{q+1}} & C_q(X) & \xrightarrow{\partial_q} & C_{q-1}(X) & \xrightarrow{\partial_{q-1}} & \dots \\ & & \downarrow \text{hooked dashed} & & \downarrow \text{hooked dashed} & & \downarrow \text{hooked dashed} & & \\ \dots & \xrightarrow{\partial_{q+2}} & C_{q+1}(Y) & \xrightarrow{\partial_{q+1}} & C_q(Y) & \xrightarrow{\partial_q} & C_{q-1}(Y) & \xrightarrow{\partial_{q-1}} & \dots \end{array}$$

where hooked dashed arrows represent inclusion maps $\iota : C_q(X) \hookrightarrow C_q(Y)$. The inclusion ι is a chain map and induces a map $\iota^\bullet : H_q(X) \rightarrow H_q(Y)$. Intuitively, ι^\bullet sends a cycle, i.e., an element of $\ker \partial_q(X)$, to itself in $\ker \partial_q(Y)$. The q -th persistent homology for the pair (X, Y)

is the image

$$\iota^\bullet(H_q(X)).$$

Proposition 2.0.4.

$$\iota^\bullet(H_q(X)) \cong \frac{\ker \partial_q(X)}{\ker \partial_q(X) \cap \text{im } \partial_{q+1}(Y)}.$$

Proof. For any $\alpha \in \ker \partial_q(X)$, $\alpha + \text{im } \partial_{q+1}(X) \in \ker \iota^\bullet$ is equivalent to $\alpha \in \text{im } \partial_{q+1}(Y)$. $\alpha \in \ker \partial_q(X) \cap \text{im } \partial_{q+1}(Y)$, is equivalent to

$$\alpha + \text{im } \partial_{q+1}(X) \in \frac{\ker \partial_q(X) \cap \text{im } \partial_{q+1}(Y)}{\text{im } \partial_{q+1}(X)}.$$

Therefore,

$$\ker \iota^\bullet = \frac{\ker \partial_q(X) \cap \text{im } \partial_{q+1}(Y)}{\text{im } \partial_{q+1}(X)},$$

and

$$\iota^\bullet(H_q(X)) \cong \frac{\ker \partial_q(X) / \text{im } \partial_{q+1}(X)}{\ker \partial_q(X) \cap \text{im } \partial_{q+1}(Y) / \text{im } \partial_{q+1}(X)} \cong \frac{\ker \partial_q(X)}{\ker \partial_q(X) \cap \text{im } \partial_{q+1}(Y)}.$$

□

The quotient space $\ker \partial_q(X) / (\ker \partial_q(X) \cap \text{im } \partial_{q+1}(Y))$ has an intuitive interpretation. The space $\ker \partial_q(X) \cap \text{im } \partial_{q+1}(Y)$ actually corresponds to all the $(q + 1)$ -simplices in Y whose boundaries are in X . When X evolves into Y (more simplices are added), some topological features of X will be lost (a cycle maybe filled). We can say that the quotient space $\ker \partial_q(X) / (\ker \partial_q(X) \cap \text{im } \partial_{q+1}(Y))$ captures the persistent topological features of X .

Persistence modules

Given a point cloud, a *filtration* of simplicial complexes can be constructed in various ways. Recall that a filtration is a sequence of simplicial complexes $\{X_t\}$, where $X_s \subset X_t$ if $s \leq t$.

Example 2.0.7. One popular construction is called the Rips filtration, where X_t is the Rips complex with t as the threshold: a simplex is in X_t if and only if the distance between any pair of its vertices is at most t . Because there are only finitely many possible pairwise distances, X_t will change only for a finite number of times. Consider the point cloud $\{x = (1, 0), y = (0, 1), z = (-1, 0), w = (0, -1)\} \subset \mathbb{R}^2$ shown in 2.4a. When $t = 0$, there are no edges in X_t . When $t = \sqrt{2}$, X_t changes for the first time and becomes $\{x, y, z, w, xy, yz, zw, xw\}$. If t goes from $\sqrt{2}$ to 2, $X_2 = X_{\sqrt{2}} \cup \{xz, yw, yzw, xzw, xyw, xyz, xyzw\}$. As t becomes bigger, X_t contains more and more simplices.

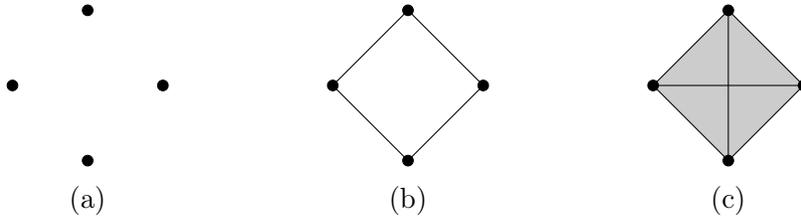


Figure 2.4 (a) $X_0 = \{x, y, z, w\}$. (b) $X_{\sqrt{2}} = \{x, y, z, w, xy, yz, zw, xw\}$. (c) $X_2 = X_{\sqrt{2}} \cup \{xz, yw, yzw, xzw, xyw, xyz, xyzw\}$.

After a filtration is constructed, each inclusion map $X_s \subset X_t$ induces a map $\iota_{s,t}^\bullet : H_q(X_s) \rightarrow H_q(X_t)$ for each $q \leq 0$. These homology groups and maps form a sequence

$$\dots \longrightarrow H_q(X_{t_i}) \xrightarrow{\iota_{t_i, t_{i+1}}^\bullet} H_q(X_{t_{i+1}}) \xrightarrow{\iota_{t_{i+1}, t_{i+2}}^\bullet} H_q(X_{t_{i+2}}) \longrightarrow \dots$$

Note that $\iota_{t_{i+1}, t_{i+2}}^\bullet \iota_{t_i, t_{i+1}}^\bullet = \iota_{t_i, t_{i+2}}^\bullet$. Using the language of category theory, we can say that such a sequence is a functor from a totally order set $\{t_i\}$ ($t_i \leq t_{i+1}$ is thought of as a morphism $t_i \rightarrow t_{i+1}$) to the category of vector spaces. Such a functor is called a persistence module. This viewpoint will pave the way for further generalizations.

It is known that a persistence module has a unique decomposition into fundamental building blocks [151]. We usually view the filtration $\{X_t\}$ as a temporal evolution, so homology class are ‘born’ and ‘killed’ at certain timestamps. Each fundamental building block in the decomposition of a persistence module will be interpreted as the life of a homology class. This unique decomposition is often represented as a persistence diagram or barcodes. The

number of homology classes that are born and killed at certain timestamps can actually be calculated from persistent Betti numbers [44].

Persistent Laplacians

It is possible to construct a symmetric semi-definite operator whose kernel is isomorphic to a given persistent homology group. In this section we slightly generalize the notion of persistent homology to the setting of *differential graded inner product spaces*², and give the definition of a *persistent Laplacian*.

Definition 2.0.6. A *differential graded inner product space* (V, d^V) is just a chain complex

$$\cdots \xrightarrow{d_{q+2}^V} V_{q+1} \xrightarrow{d_{q+1}^V} V_q \xrightarrow{d_q^V} V_{q-1} \xrightarrow{d_{q-1}^V} \cdots$$

whose chain groups are inner product spaces. One can think of V as the direct sum of all V_q . When we say (V, d^V) is a subspace of (W, d^W) , we mean that V_q is a subspace of W_q for any q , and the inner space structure of V_q and boundary operator d of (V, d^V) are inherited from (W, d^W) .

For a pair of differential graded inner product spaces $(V, d^V) \subset (W, d^W)$, the q -th persistent homology group is defined analogously by

$$\iota^\bullet(H_q(V)) \cong \frac{\ker d_q^V}{\ker d_q^V \cap \operatorname{im} d_{q+1}^W}.$$

Observe that $\ker d_q^V \cap \operatorname{im} d_{q+1}^W = V_q \cap \operatorname{im} d_{q+1}^W$. The preimage of $V_q \cap \operatorname{im} d_{q+1}^W$ under d_{q+1}^W is just $(d_{q+1}^W)^{-1}(V_q) = \{w \in W_{q+1} \mid d_{q+1}^W w \in V_q\}$. Hence, $\ker d_q^V \cap \operatorname{im} d_{q+1}^W$ is the image of $\pi d_{q+1}^W|_{(d_{q+1}^W)^{-1}(V_q)} : (d_{q+1}^W)^{-1}(V_q) \rightarrow V_q$, where $\pi = \iota^\dagger$, the projection map from W to V . We denote $\pi d_{q+1}^W|_{(d_{q+1}^W)^{-1}(V_q)}$ by $d_{q+1}^{V,W}$, and $(d_{q+1}^W)^{-1}(V_q)$ by $\Theta_{q+1}^{V,W}$. These maps are shown in the

²The extension of persistent Laplacians to the setting of differential graded inner product spaces first appeared in [88].

following diagram

$$\begin{array}{ccccc}
V_{q+1} & \xrightarrow{d_{q+1}^V} & V_q & \xrightleftharpoons[(d_q^V)^\dagger]{d_q^V} & V_{q-1} \\
\downarrow \text{hooked} & & \nearrow d_{q+1}^{V,W} & & \downarrow \text{hooked} \\
& & \Theta_{q+1}^{V,W} & & \\
& & \nwarrow (d_{q+1}^{V,W})^\dagger & & \\
W_{q+1} & \xrightarrow{d_{q+1}^W} & W_q & & \\
\downarrow \text{hooked} & & \downarrow \text{hooked} & & \\
& & & &
\end{array}$$

where hooked dashed arrows represent inclusion maps. We define the q -th persistent Laplacian $\Delta_q^{V,W} : V_q \rightarrow V_q$ by

$$(d_q^V)^\dagger d_q^V + d_{q+1}^{V,W} (d_{q+1}^{V,W})^\dagger.$$

Because $d_q^V d_{q+1}^{V,W} = 0$, by Proposition 2.0.2 we see that

$$\ker \Delta_q^{V,W} \cong \frac{\ker d_q^V}{\ker d_q^V \cap \text{im } d_{q+1}^W}.$$

This relation is sometimes referred to as the persistent Hodge theorem. The operators $d_{q+1}^{V,W} (d_{q+1}^{V,W})^\dagger$ and $(d_q^V)^\dagger d_q^V$ are sometimes referred to as the up persistent Laplacian and the down persistent Laplacian, respectively.

If there is an inner product preserving chain map $f : (V, d^V) \rightarrow (W, d^W)$, the q -th persistent Laplacian can also be defined [88]. The concept of a persistent Laplacian was first introduced by André Lieutier in 2014 [84]. Later in 2020, it was rediscovered independently by Wang et al [133]. Persistent Laplacians are originally defined for simplicial chain complexes $(C_\bullet(X), \partial)$ and $(C_\bullet(Y), \partial)$ where X is a subcomplex of Y .

The matrix representation of a persistent Laplacian

In this section we will explain the calculation of the matrix representation of a persistent Laplacian with simple examples.

Example 2.0.8. We compute the persistent Laplacian $\Delta_1^{X,Y}$ for X and Y in Figure 2.5. Since each chain group has a canonical orthonormal basis, the matrix representation of ∂^\dagger is

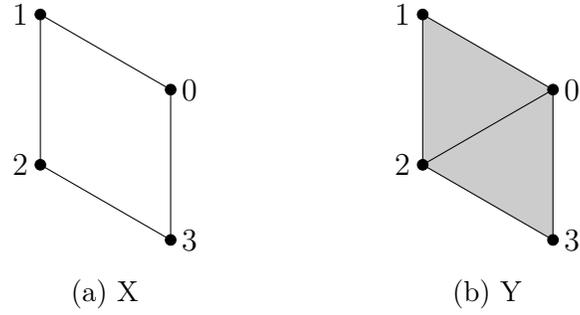


Figure 2.5 $X = \{0, 1, 2, 3, 01, 12, 23, 03\}$ and $Y = \{0, 1, 2, 3, 01, 12, 23, 03, 02, 012, 023\}$.

the transpose of the matrix representation of ∂ . The matrix representation of ∂_1^X is

$$\begin{array}{c}
 \begin{array}{cccc}
 & 01 & 12 & 23 & 03 \\
 0 & \left(\begin{array}{cccc}
 -1 & 0 & 0 & -1 \\
 1 & 1 & -1 & 0 & 0 \\
 2 & 0 & 1 & -1 & 0 \\
 3 & 0 & 0 & 1 & 1
 \end{array} \right) \\
 \end{array}
 \end{array}
 ,$$

and we easily get the matrix representation of $(\partial_1^X)^\dagger$

$$\begin{array}{c}
 \begin{array}{cccc}
 & 0 & 1 & 2 & 3 \\
 01 & \left(\begin{array}{cccc}
 -1 & 1 & 0 & 0 \\
 12 & 0 & -1 & 1 & 0 \\
 23 & 0 & 0 & -1 & 1 \\
 03 & -1 & 0 & 0 & 1
 \end{array} \right) \\
 \end{array}
 \end{array}
 .$$

The matrix representation of ∂_2^Y is

$$\begin{array}{cc} & \begin{array}{cc} 012 & 023 \end{array} \\ \begin{array}{c} 01 \\ 12 \\ 23 \\ 03 \\ 02 \end{array} & \left(\begin{array}{cc} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ -1 & 1 \end{array} \right) \end{array} \cdot$$

Our goal is to find a basis of the subspace $\Theta_2^{X,Y}$. We first try to make the submatrix

$$\begin{array}{cc} & \begin{array}{cc} 012 & 023 \end{array} \\ 02 & \left(\begin{array}{cc} -1 & 1 \end{array} \right) \end{array}$$

in column echelon form. We apply one column reduction and get

$$\begin{array}{cc} & \begin{array}{cc} 012 & 023 + 012 \end{array} \\ \begin{array}{c} 01 \\ 12 \\ 23 \\ 03 \\ 02 \end{array} & \left(\begin{array}{cc} 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 0 & -1 \\ -1 & 0 \end{array} \right) \end{array} \cdot$$

Therefore, $\Theta_2^{X,Y} = \text{span}(023 + 012)$ and one matrix representation of $\partial_2^{X,Y}$ is

$$\begin{array}{cc} & \begin{array}{cc} 023 + 012 \end{array} \\ \begin{array}{c} 01 \\ 12 \\ 23 \\ 03 \end{array} & \left(\begin{array}{c} 1 \\ 1 \\ 1 \\ -1 \end{array} \right) \end{array} \cdot$$

Generally, for $f : V \rightarrow W$, if we choose arbitrary bases of V and W and take a matrix representation $[f]$ of f , then the matrix representation $[f^\dagger]$ of f^\dagger is $P^{-1}[f]^T Q$, where P and Q are inner product matrices of V and W respectively. If we use $\{023 + 012\}$ as the basis of $\Theta_2^{X,Y}$, then the inner product matrix of $\Theta_2^{X,Y}$ is 2 (the square of the norm of $023 + 012$). The corresponding matrix representation of $(\partial_2^{X,Y})^\dagger$ is

$$\frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & -1 \end{pmatrix}$$

and the matrix representation of the up Laplacian is

$$\begin{array}{cccc} & 01 & 12 & 23 & 03 \\ \begin{array}{l} 01 \\ 12 \\ 23 \\ 03 \end{array} & \begin{pmatrix} 1/2 & 1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 & -1/2 \\ -1/2 & -1/2 & -1/2 & 1/2 \end{pmatrix} \end{array}.$$

We can also find an orthonormal basis for $\Theta_2^{X,Y}$ at first and just take the transpose. After the calculation of the up persistent Laplacian and the down persistent Laplacian, we only need to add them to get the persistent Laplacian.

One of the main contributions of [63, 96] is that the up persistent Laplacian can be calculated via the Schur complement. Let's first recall the definition.

Definition 2.0.7. For a square matrix $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, the Schur complement [23] of D in M , denoted by M/D , is given by $A - BD^{-1}C$ where D^{-1} is the Moore-Penrose generalized inverse. We can view M as an operator on \mathbb{R}^n , so M/D is a way of 'restricting' M to the subspace \mathbb{R}^m corresponding to A . Now suppose $L : V \rightarrow V$ is a linear operator on a finite-dimensional real inner product space V , and W is a subspace, we can define the *Schur restriction* of L onto W as follows. We first choose bases for W and W^\perp . With respect

to the chosen bases, the matrix representation of L is a block matrix
$$\begin{matrix} & W & W^\perp \\ \begin{matrix} W \\ W^\perp \end{matrix} & \begin{pmatrix} A & B \\ C & D \end{pmatrix} \end{matrix},$$
 and the Schur restriction $Sch(L, W) : W \rightarrow W$ is naturally defined by the linear operator represented by $A - BD^{-1}C$.

Gulen et al. [63] showed that if L is self-adjoint positive semi-definite, the Schur restriction $Sch(L, W)$ is well defined, i.e., independent of the choice of bases of W and W^\perp . Moreover, they proved the following proposition.

Proposition 2.0.5. *Let $f : \hat{V} \rightarrow V$ be a linear morphism between two finite-dimensional real inner product space, and $L = ff^\dagger : V \rightarrow V$. For any subspace $W \subset V$, let $f_W : f^{-1}W \rightarrow W$ be the restriction of f on $f^{-1}W$ (its codomain is also restricted to W). Then the Schur restriction of L onto W is $f_W(f_W)^\dagger$.*

This proposition ensures that we can first compute $\partial_{q+1}^Y(\partial_{q+1}^Y)^\dagger$ and then compute the Schur complement. In the above example, the matrix representation of $\partial_{q+1}^Y(\partial_{q+1}^Y)^\dagger$ is

$$\begin{matrix} & 01 & 12 & 23 & 03 & 02 \\ \begin{matrix} 01 \\ 12 \\ 23 \\ 03 \\ 02 \end{matrix} & \begin{pmatrix} 1 & 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & -1 & 1 & -1 \\ -1 & -1 & 1 & -1 & 2 \end{pmatrix} \end{matrix}.$$

Note that $C_q(X) = \text{span}\{01, 12, 23, 03\}$. We treat this matrix as a block matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where A is

$$\begin{array}{c} \\ 01 \\ 12 \\ 23 \\ 03 \end{array} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

The Schur complement $A - BD^{-1}C$ is

$$\begin{array}{c} \\ 01 \\ 12 \\ 23 \\ 03 \end{array} \begin{pmatrix} 1/2 & 1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 & -1/2 \\ -1/2 & -1/2 & -1/2 & 1/2 \end{pmatrix}.$$

Example 2.0.9. Consider two complexes $X = \{0, 1, 2, 3, 4, 02, 04, 14, 23, 34\}$ and $Y = X \cup \{024, 234\}$. Let's first calculate the matrix representation of $\Delta_1^{X,Y}$. The matrix representation of ∂_1^X is

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} 0 & -1 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

The matrix representation of ∂_2^Y is

$$\begin{array}{c} 024 \quad 234 \\ 14 \\ 02 \\ 23 \\ 04 \\ 34 \\ 24 \end{array} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix}.$$

Note that $24 \notin C_1(X)$. It is easy to see that $\text{span}\{024 + 234\} = \Theta_2^{X,Y}$. So $\partial_2^{X,Y}$ is equal to

$$\begin{array}{c} 024 + 234 \\ 14 \\ 02 \\ 23 \\ 04 \\ 34 \end{array} \begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}$$

and P is equal to $\langle 024 + 234, 024 + 234 \rangle = 2$. So $\Delta_1^{X,Y}$ is equal to

$$\begin{pmatrix} 2 & 0 & 0 & 1 & 1 \\ 0 & 2.5 & -0.5 & 0.5 & 0.5 \\ 0 & -0.5 & 2.5 & -0.5 & -0.5 \\ 1 & 0.5 & -0.5 & 2.5 & 0.5 \\ 1 & 0.5 & -0.5 & 0.5 & 2.5 \end{pmatrix}.$$

Example 2.0.10. Consider the complex Y



and let $X = \{1, 2\}$. We compute $\Delta_0^{X,Y}$. The matrix representation of ∂_1^X is

$$\begin{array}{c} 13 \quad 34 \quad 24 \\ 1 \left(\begin{array}{ccc} -1 & 0 & 0 \\ 2 \left(\begin{array}{ccc} 0 & 0 & -1 \\ 3 \left(\begin{array}{ccc} 1 & -1 & 0 \\ 4 \left(\begin{array}{ccc} 0 & 1 & 1 \end{array} \right) \end{array} \right) \end{array} \right) \end{array} \right) \end{array}.$$

After a few steps of Gauss elimination we get

$$\begin{array}{c} 13 \quad 13 + 34 \quad -13 - 34 + 24 \\ 1 \left(\begin{array}{ccc} -1 & -1 & 1 \\ 2 \left(\begin{array}{ccc} 0 & 0 & -1 \\ 3 \left(\begin{array}{ccc} 1 & 0 & 0 \\ 4 \left(\begin{array}{ccc} 0 & 1 & 0 \end{array} \right) \end{array} \right) \end{array} \right) \end{array} \right) \end{array}.$$

It is clear that $\Theta_1^{X,Y} = \text{span}\{-13 - 34 + 24\}$, $P = 3$ and the matrix representation of $\partial_1^{X,Y}$ is

$$\begin{array}{c} -13 - 34 + 24 \\ 1 \left(\begin{array}{c} 1 \\ 2 \left(\begin{array}{c} -1 \end{array} \right) \end{array} \right) \end{array}.$$

Then the matrix representation of $\Delta_0^{X,Y}$ is

$$\begin{pmatrix} 1/3 & -1/3 \\ -1/3 & 1/3 \end{pmatrix}.$$

Its spectrum is $\{0, 2/3\}$.

Eigenvalues and eigenvectors of a Laplacian

There are already some results concerning the relation between spectra of Laplacians and the shape of a simplicial complex [52, 72]. How do we interpret eigenvectors of a Laplacian?

For an eigenvector of a q -th combinatorial Laplacian, we can look at the shape of q -simplices where the eigenvector has support (signs are arbitrary because they are affected by the fixed ordering of vertices). Empirical observations [79, 98, 138] suggest that:

(a) harmonic eigenvectors (eigenvectors of zero eigenvalues) have support near q -dimensional “holes” (or vertices in a connected component when $q = 0$).

(b) non-harmonic eigenvectors (eigenvectors of nonzero eigenvalues) have support near “clusters” of q -simplices. As to persistent Laplacians, very little is known about the geometrical/topological interpretation of eigenvalues and eigenvectors.

The workflow of the (persistent) Laplacian approach

Given a point cloud, the usual workflow of topological data analysis is to first generate a filtration $\{X_t\}$ (t is often associated with distance) and then compute persistent homology (often in the form of barcodes or a persistence diagram). If we want to employ Laplacians, we need to select some pairs of X_t and X_s in the filtration and compute persistent Laplacians. After we compute some Laplacians, we have to featurize Laplacians. Here featurization of Laplacians is the process of transforming a set of Laplacians to a vector of a fixed size. As the spectrum of a Laplacian is not affected by the global ordering of points in the point cloud, most featurization methods focus on the spectrum. Some featurization methods are summarized in [97]. So far the choices of pairs (X_t, X_s) and featurization methods require some domain-knowledge and experience about the specific problem, and we wonder if any data-driven (or self-learning) approach is possible.

Homotopy continuation and persistent Laplacians

Seeking new ways to calculate the spectrum of an operator is an active research topic [3]. In addition to the traditional methods of numerical linear algebra, one may alternatively calculate the spectrum by finding the roots of the characteristic polynomial associated with the operator.

Homotopy continuation is a method for solving a single polynomial or systems of polynomial equations. The essential idea is to build a homotopy between the system to be solved

(called the target system) and an easier system with known roots (called the start system) and track down the known roots of the start system to the roots of the target system. As systems of polynomial equations arise in mathematics, science, and engineering, homotopy continuation methods have found applications in various areas, such as algebraic geometry [69, 81], robot kinematics [130], optimal control [6], differential equations [1, 67], and biology [62, 66, 113]. Several software packages implement homotopy continuation methods, such as Bertini [7], HomotopyContinuation.jl [12], Hom4PS-3 [33], and PHCpack [128]. [140] verified that at least for some simple polytopes and small molecules in the three-dimensional space, the minimal nonzero eigenvalues of persistent Laplacians calculated by homotopy continuation are very close to the result from HERMES.

CHAPTER 3

(CO)SHEAVES AND PERSISTENT SHEAF LAPLACIANS

The goal of this chapter is to introduce the theory of cellular (co)sheaves and the extension of persistent Laplacians to the setting of cellular sheaves. We will first introduce weighted simplicial complexes, which can be viewed as cellular cosheaves.

Weighted simplicial complexes

Generally speaking, any simplicial complex whose simplices have weights can be called a *weighted simplicial complex*. The weights can be geometrical, such as angles between simplices, volumes of simplices, or non-geometrical such as numbers of scientific papers coauthored by groups of people. Many theories and models involving weighted simplicial complexes exist (e.g., [5, 9, 38, 103, 119]). Here we focus on the theory of weighted simplicial complexes proposed by Robert J. MacG. Dawson [41], and developed in [14, 15, 83, 109, 111, 142, 143].

Definition 3.0.1. A weighted simplicial complex is a simplicial complex where each simplex σ has a weight $w(\sigma)$ valued in an integral domain R , such that if σ is a face of τ , then $w(\tau)$ is divisible by $w(\sigma)$.

The *weighted chain complex* of a weighted simplicial complex X is defined as follows. Let the q -th chain group $C_q(X, w)$ be the set of formal sums of nonzero weighted q -simplices with coefficients in R . For $\sigma = [v_{a_0}, \dots, v_{a_q}]$, we denote the face $[v_{a_0}, \dots, \hat{v}_{a_i}, \dots, v_{a_q}]$ by $d_i\sigma$. The *weighted boundary operator* ∂ is defined by

$$\partial\sigma = \sum_{i=0}^q \frac{w(\sigma)}{w(d_i\sigma)} (-1)^i d_i\sigma.$$

We still have $\partial^2 = 0$, because for $0 \leq i < j \leq q$,

$$\frac{w(\sigma)}{w(d_i\sigma)} \frac{w(d_i\sigma)}{w(d_{j-1}d_i\sigma)} = \frac{w(\sigma)}{w(d_j\sigma)} \frac{w(d_j\sigma)}{w(d_id_j\sigma)} = \frac{w(\sigma)}{w(d_id_j\sigma)};$$

hence, weighted homology groups can be defined analogously. Wu et al. [142] pointed out that in the proof of $\partial^2 = 0$, what really matters is the quotient of weights. If we write

$w(\tau)/w(\sigma)$ as $\phi(\sigma, \tau)$, then the equality

$$\frac{w(\sigma)}{w(d_i\sigma)} \frac{w(d_i\sigma)}{w(d_{j-1}d_i\sigma)} = \frac{w(\sigma)}{w(d_j\sigma)} \frac{w(d_j\sigma)}{w(d_id_j\sigma)}$$

becomes

$$\phi(d_i\sigma, d_{j-1}d_i\sigma)\phi(\sigma, d_i\sigma) = \phi(d_j\sigma, d_id_j\sigma)\phi(\sigma, d_j\sigma),$$

which means that any $\phi : X \times X \rightarrow R$ satisfying this equality induces a (ϕ -weighted) boundary operator

$$\partial_q\sigma = \sum_{i=0}^q (-1)^i \phi(\sigma, d_i\sigma) d_i\sigma.$$

such that $\partial^2 = 0$. A simplicial complex paired with such a generalized weight function ϕ is called a ϕ -weighted simplicial complex.

Example 3.0.1. [142] A weighted polygon is a polygon with $\phi(\{v_i, v_j\}, v_i) = \alpha_i \in \mathbb{Z}$ (Figure 3.1). The matrix representation of ∂_1 is

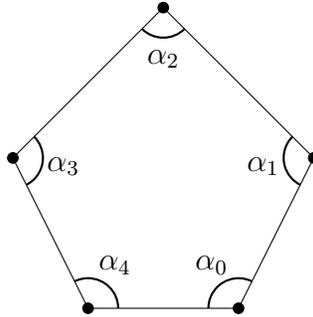


Figure 3.1 A weighted polygon.

$$\begin{array}{c}
 v_0v_1 \quad v_0v_4 \quad v_1v_2 \quad v_2v_3 \quad v_3v_4 \\
 v_0 \begin{pmatrix} -\alpha_0 & -\alpha_0 & 0 & 0 & 0 \\ \alpha_1 & 0 & -\alpha_1 & 0 & 0 \\ 0 & 0 & \alpha_2 & -\alpha_2 & 0 \\ 0 & 0 & 0 & \alpha_3 & -\alpha_3 \\ 0 & \alpha_4 & 0 & 0 & \alpha_4 \end{pmatrix} \\
 v_1 \\
 v_2 \\
 v_3 \\
 v_4
 \end{array}$$

and the weighted H_0 is dependent on α_i . The weighted homology of weighted polygons might be useful for studying ring structures in biomolecules.

We have emphasized that a point cloud can be studied by building a filtration of simplicial complexes. If we want to distinguish some points from other points, we can assign weights and building a filtration of weighted simplicial complexes [111]. We may also consider weighted versions of combinatorial and persistent Laplacians [142].

Example 3.0.2. Suppose each point v in a point cloud has weight $w(v)$. We can associate any simplex $\{v_{a_0}, \dots, v_{a_q}\}$ the product weight [111]

$$\prod_{i=0}^q w(v_{a_i}).$$

Since the weighted boundary map can be formally given by

$$\partial(\sigma) = \sum_{i=0}^q w(v_{a_i})(-1)^i d_i \sigma,$$

we can just define the q -th chain groups as the space generated by all q -simplices without worrying about zero weights.

Example 3.0.3. Suppose a point cloud contains two types of points $\{A, B\}$. We can assign weights $\{0, 1\}$ to $\{A, B\}$, and compute weighted homology and Laplacians using product weighting. At least when a point cloud is simple, weighted combinatorial Laplacian can be used to differentiate among different patterns of distribution of A and B . For a point cloud of four points $\{(0, 0), (1, 0), (1, 1), (0, 1)\}$ there are five configurations (shown in Figure 3.2) that include at least one point whose weight is 1. The weighted Laplacian results are shown in Figure 3.3.

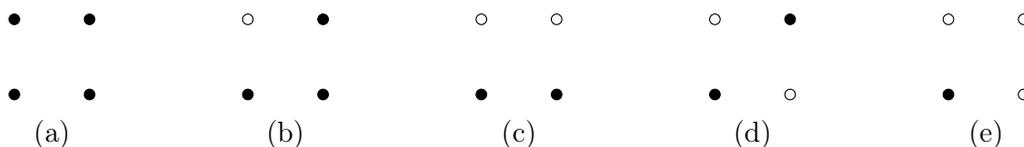


Figure 3.2 Different patterns of A and B .

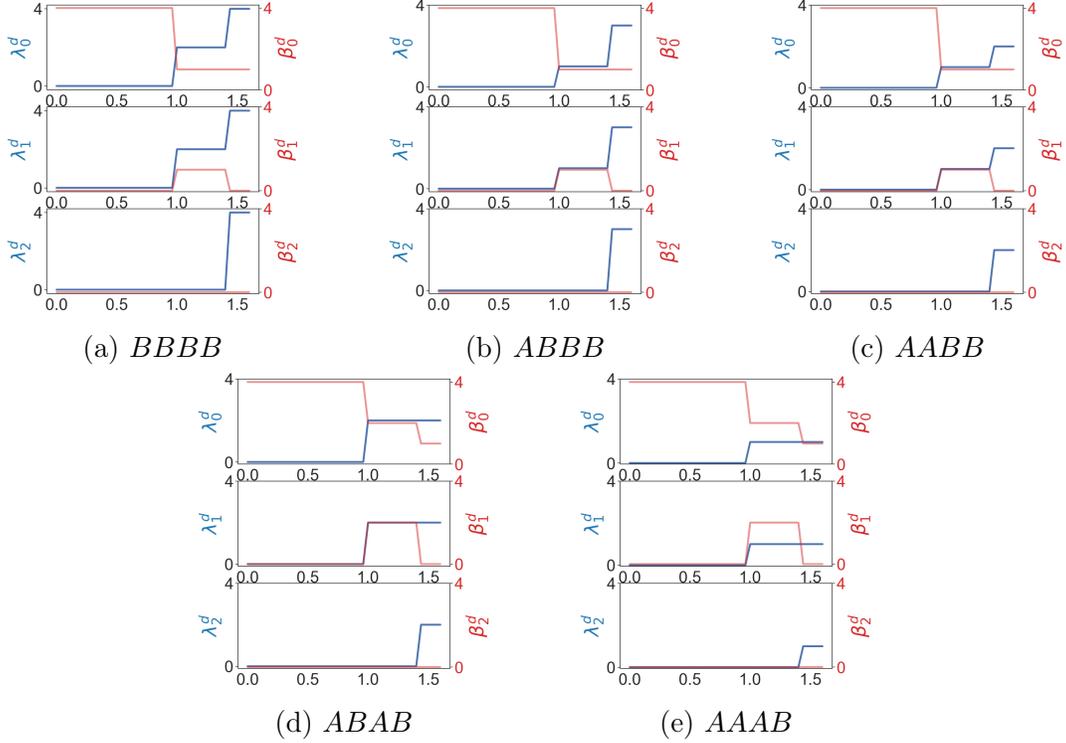


Figure 3.3 Results of weighted homology and Laplacians.

Cellular (co)sheaves

In a ϕ -weighted simplicial complex, we can imagine that a copy of R resides each simplex and $\phi(\tau, \sigma)$ represents a scalar multiplication from the copy on τ to the copy on σ [65]. If we associate each simplex with a vector space and designate a linear morphism for every face relation, we will get a *cellular (co)sheaf*. The theory of cellular (co)sheaves was first introduced in [122] and later gained attention for its application potential [39, 64, 150, 115, 116].

Definition 3.0.2. ¹A *cellular sheaf* \mathcal{S} on a simplicial complex X consists of the following data: an assignment to each simplex σ of X a (finite-dimensional) vector space $\mathcal{S}(\sigma)$ and to each face relation $\sigma \leq \tau$ a linear morphism of vector spaces denoted by $\mathcal{S}_{\sigma \leq \tau}$ or $\mathcal{S}(\sigma \leq \tau) : \mathcal{S}(\sigma) \rightarrow \mathcal{S}(\tau)$, satisfying the rule

$$\rho \leq \sigma \leq \tau \Rightarrow \mathcal{S}_{\rho \leq \tau} = \mathcal{S}_{\sigma \leq \tau} \circ \mathcal{S}_{\rho \leq \sigma}$$

¹For ease of exposition we simplify the definition of a cellular (co)sheaf.

and $\mathcal{S}_{\sigma \leq \sigma}$ is the identity map.

The vector space $\mathcal{S}(\sigma)$ is referred to as the *stalk* of \mathcal{S} over σ , and the linear morphism $\mathcal{S}_{\sigma \leq \tau}$ is referred to as the *restriction map* of the face relation $\sigma \leq \tau$. A *cellular cosheaf* is very similar to a cellular sheaf, and the only difference is that in a cosheaf $\mathcal{S}_{\sigma \leq \tau}$ is a morphism from $\mathcal{S}(\tau)$ to $\mathcal{S}(\sigma)$, referred to as the *extension map* of the face relation $\sigma \leq \tau$.

Example 3.0.4. Let X be a finite simplicial complex. We attach to every simplex of X a fixed vector space V and let every restriction map be the identity map. This sheaf is referred to as the *constant sheaf* \underline{V} on X .

Example 3.0.5. Let X be a finite simplicial complex. We attach to every simplex of X the zero space, except a simplex σ , the stalk of which is a vector space V . All restriction maps have to be zero except the identity map required by the definition. This sheaf is called a *skyscraper sheaf*.

Definition 3.0.3. Suppose that $f : X \rightarrow Y$ is a simplicial map [99] and that \mathcal{S} is a cellular sheaf on Y . The *pullback sheaf* $f^*\mathcal{S}$ on X is given by

$$(f^*\mathcal{S})(\sigma) = \mathcal{S}(f(\sigma)),$$

and for the face relation $\sigma \leq \tau$ of X ,

$$(f^*\mathcal{S})_{\sigma \leq \tau} = \mathcal{S}_{f(\sigma) \leq f(\tau)}.$$

The pullback of a cosheaf is defined analogously.

Example 3.0.6. Suppose that X is a subcomplex of Y , and \mathcal{S} is a sheaf on Y . We can define a sheaf \mathcal{T} on X using the data of Y . For $\sigma \in X$, let $\mathcal{T}(\sigma) = \mathcal{S}(\sigma)$. For the face relation $\sigma \leq \tau$ in X , let $\mathcal{T}(\sigma \leq \tau) = \mathcal{S}(\sigma \leq \tau)$. The sheaf \mathcal{T} is a pullback of \mathcal{S} .

Definition 3.0.4. A *global section* s of a sheaf \mathcal{S} is an assignment to each simplex σ an element $s_\sigma \in \mathcal{S}(\sigma)$ such that $\mathcal{S}_{\sigma \leq \tau}(s_\sigma) = s_\tau$ for any face relation $\sigma \leq \tau$. The set of global sections is denoted by $\Gamma(X; \mathcal{S})$.

It is possible to construct the sheaf cochain complex of a sheaf or cosheaf chain complex of a cosheaf. For a sheaf \mathcal{S} on a finite simplicial complex X , let the q -th cochain group $C^q(X; \mathcal{S})$ be the direct sum of $\mathcal{S}(\sigma)$ over all q -simplices σ . To define the coboundary map d , we need a *signed incidence relation* [39].

Definition 3.0.5. A *signed incidence relation* is an assignment to every face relation $\sigma \leq \tau$ an integer $[\sigma : \tau]$ satisfying the following conditions: (1) if $\dim \tau - \dim \sigma > 1$, then $[\sigma : \tau] = 0$; and (2) if $\gamma \leq \tau$ and $\dim \tau - \dim \gamma = 2$, the sum $\sum_{\sigma} [\gamma : \sigma][\sigma : \tau] = 0$.

Once a signed incidence relation is given, the coboundary map $d^q : C^q(X; \mathcal{S}) \rightarrow C^{q+1}(X; \mathcal{S})$ is given by

$$d^q|_{\mathcal{S}(\sigma)} = \sum_{\sigma \leq \tau} [\sigma : \tau] \mathcal{S}_{\sigma \leq \tau}.$$

Since d^q is a linear morphism, its action on each stalk $\mathcal{S}(\sigma)$ determines itself. We can verify that $d^q d^{q-1} = 0$ [39, Lemma 6.2.2], so there is the sheaf cochain complex

$$0 \longrightarrow C^0(X; \mathcal{S}) \xrightarrow{d} C^1(X; \mathcal{S}) \xrightarrow{d} C^2(X; \mathcal{S}) \xrightarrow{d} \dots$$

The q -th sheaf cohomology group $H^q(X; \mathcal{S})$ is defined as $\ker d^q / \operatorname{im} d^{q-1}$.

A natural signed incidence relation exists for every oriented simplicial complex. Recall that the orientation of a simplex is determined by the ordering of its vertices. For an oriented simplex $\tau = [v_0, v_1, \dots, v_n]$ and its oriented face $\sigma = [v_0, \dots, \hat{v}_i, \dots, v_n]$, we let $[\sigma : \tau] = (-1)^i$. If σ or τ is oriented alternatively, we let $[\sigma : \tau] = (-1)^{i+1}$. This signed incidence relation is used throughout this paper. We remind the reader that we do not need orientation information to define a sheaf.

Dually, given a cellular cosheaf \mathcal{S} on a simplicial complex X , the q -th cosheaf chain group is the direct sum of all stalks over q -simplices, and the cosheaf boundary map is given by

$$d_q|_{\mathcal{S}(\tau)} = \sum_{\sigma \leq \tau} [\sigma : \tau] \mathcal{S}_{\sigma \leq \tau}.$$

(Co)sheaf Laplacians

Recall that if cochain groups of a cochain complex

$$\dots \xrightarrow{d^{q-2}} A^{q-1} \xrightarrow{d^{q-1}} A^q \xrightarrow{d^q} A^{q+1} \xrightarrow{d^{q+1}} \dots$$

are all finite-dimensional inner product spaces, the q -th combinatorial Laplacian $\Delta_q : A^q \rightarrow A^q$ is defined by

$$\Delta_q = (d^q)^\dagger d^q + d^{q-1}(d^{q-1})^\dagger,$$

where $(d^q)^\dagger$ is the adjoint of d^q , and it is well-known that the kernel of Δ_q is isomorphic to the q -th cohomology group H^q . Hansen and Ghrist [65] applied this construction to sheaf cochain complexes and the resulting new combinatorial Laplacian is referred to as the sheaf Laplacian. If every stalk of a sheaf \mathcal{S} is a finite-dimensional inner product space, we can equip an inner product structure on every $C^q(X; \mathcal{S})$ such that $\mathcal{S}(\sigma)$ and $\mathcal{S}(\sigma')$ are orthogonal if $\sigma \neq \sigma'$.

Example 3.0.7. Suppose there is a sheaf \mathcal{F} over the simplicial complex $\{0, 1, 2, 01, 02, 12\}$, then the sheaf coboundary map d^0 is represented by the block matrix

$$\begin{array}{c} \mathcal{F}_0 \quad \mathcal{F}_1 \quad \mathcal{F}_2 \\ \mathcal{F}_{01} \begin{pmatrix} -\mathcal{F}_{0 \leq 01} & \mathcal{F}_{1 \leq 01} & 0 \\ \mathcal{F}_{02} \begin{pmatrix} -\mathcal{F}_{0 \leq 02} & 0 & \mathcal{F}_{2 \leq 02} \\ \mathcal{F}_{12} \begin{pmatrix} 0 & -\mathcal{F}_{1 \leq 12} & \mathcal{F}_{2 \leq 12} \end{pmatrix} \end{pmatrix} \end{pmatrix} \end{array}.$$

The 0-th sheaf Laplacian is $(d^0)^\dagger d^0$, represented by the block matrix

$$\begin{array}{c} \mathcal{F}_0 \quad \mathcal{F}_1 \quad \mathcal{F}_2 \\ \mathcal{F}_0 \begin{pmatrix} \mathcal{F}_{0 \leq 01}^* \mathcal{F}_{0 \leq 01} + \mathcal{F}_{0 \leq 02}^* \mathcal{F}_{0 \leq 02} & -\mathcal{F}_{0 \leq 01}^* \mathcal{F}_{1 \leq 01} & -\mathcal{F}_{0 \leq 02}^* \mathcal{F}_{2 \leq 02} \\ \mathcal{F}_1 \begin{pmatrix} -\mathcal{F}_{1 \leq 01}^* \mathcal{F}_{0 \leq 01} & \mathcal{F}_{1 \leq 01}^* \mathcal{F}_{1 \leq 01} + \mathcal{F}_{1 \leq 12}^* \mathcal{F}_{1 \leq 12} & -\mathcal{F}_{1 \leq 12}^* \mathcal{F}_{2 \leq 12} \\ \mathcal{F}_2 \begin{pmatrix} -\mathcal{F}_{2 \leq 02}^* \mathcal{F}_{0 \leq 02} & -\mathcal{F}_{2 \leq 12}^* \mathcal{F}_{1 \leq 12} & \mathcal{F}_{2 \leq 02}^* \mathcal{F}_{2 \leq 02} + \mathcal{F}_{2 \leq 12}^* \mathcal{F}_{2 \leq 12} \end{pmatrix} \end{pmatrix} \end{pmatrix} \end{array}.$$

Persistent (co)sheaf Laplacians

Persistent (co)sheaf (co)homology is known to experts [114, 149] and a systematical treatment can be found in [117]. We first show how to construct persistent cosheaf Laplacians. Suppose Y is an oriented simplicial complex and X is a subcomplex of Y whose orientation is identical to Y . If \mathcal{G} is a cosheaf on Y , let \mathcal{F} be the pullback cosheaf on X . We have the following commutative diagram

$$\begin{array}{ccccccc} \cdots & \xrightarrow{d} & C_{q+1}(X; \mathcal{F}) & \xrightarrow{d} & C_q(X; \mathcal{F}) & \xrightarrow{d} & C_{q-1}(X; \mathcal{F}) & \xrightarrow{d} & \cdots \\ & & \downarrow \text{hooked dashed} & & \downarrow \text{hooked dashed} & & \downarrow \text{hooked dashed} & & \\ \cdots & \xrightarrow{d} & C_{q+1}(Y; \mathcal{G}) & \xrightarrow{d} & C_q(Y; \mathcal{G}) & \xrightarrow{d} & C_{q-1}(Y; \mathcal{G}) & \xrightarrow{d} & \cdots \end{array}$$

where hooked dashed arrows represent inclusion maps $\iota : C_q(X; \mathcal{F}) \hookrightarrow C_q(Y; \mathcal{G})$. The q -th persistent cosheaf homology group is defined by

$$\iota^* H_q(X; \mathcal{F}).$$

Consider the following diagram

$$\begin{array}{ccccc} C_{q+1}(X; \mathcal{F}) & \xrightarrow{d_{q+1}^{\mathcal{F}}} & C_q(X; \mathcal{F}) & \xrightleftharpoons[(d_q^{\mathcal{F}})^\dagger]{d_q^{\mathcal{F}}} & C_{q-1}(X; \mathcal{F}) \\ \downarrow \text{hooked dashed} & & \swarrow d_{q+1}^{\mathcal{F}, \mathcal{G}} & \nearrow (d_{q+1}^{\mathcal{F}, \mathcal{G}})^\dagger & \downarrow \text{hooked dashed} \\ & & \Theta_{q+1}^{\mathcal{F}, \mathcal{G}} & & \\ & & \swarrow \text{hooked dashed} & & \\ C_{q+1}(Y; \mathcal{G}) & \xrightarrow{d_{q+1}^{\mathcal{G}}} & C_q(Y; \mathcal{G}) & & \end{array}$$

where $\Theta_{q+1}^{\mathcal{F}, \mathcal{G}} = \{x \in C_{q+1}(Y; \mathcal{G}) \mid d_{q+1}^{\mathcal{G}} x \in C_q(X; \mathcal{F})\}$. Let $d_{q+1}^{\mathcal{F}, \mathcal{G}} : \Theta_{q+1}^{\mathcal{F}, \mathcal{G}} \rightarrow C_q(X; \mathcal{F})$ be $\pi_q d_{q+1}^Y|_{\Theta_{q+1}^{\mathcal{F}, \mathcal{G}}}$ (π is the adjoint of ι), the q -th persistent cosheaf Laplacian $\Delta_q^{\mathcal{F}, \mathcal{G}} : C_q(X; \mathcal{F}) \rightarrow C_q(X; \mathcal{F})$ is

$$(d_q^{\mathcal{F}})^\dagger d_q^{\mathcal{F}} + d_{q+1}^{\mathcal{F}, \mathcal{G}} (d_{q+1}^{\mathcal{F}, \mathcal{G}})^\dagger.$$

As shown in section 2, the persistent Hodge theorem is still true.

We show next that the spectrum of a persistent cosheaf Laplacian doesn't rely on the choice of the orientation of the complex Y . This is important, since a point cloud may not have a canonical ordering.

Proposition 3.0.1. *The spectrum of the q -th persistent cosheaf Laplacian is independent of the orientation of Y .*

Proof. Fixing a choice of orientation for each simplex of Y , it suffices to show that the spectrum of $\Delta_q^{\mathcal{F}, \mathcal{G}}$ is unchanged if the orientation of one simplex of Y is alternated. We first fix some notations. Suppose we change the orientation of a simplex σ , then every morphism defined with respect to this new orientation will have a bar. We also sometimes denote $\Theta_{q+1}^{\mathcal{F}, \mathcal{G}}$ by $\bar{\Theta}$. We define a linear map $I_{\sigma, -}$ such that $I_{\sigma, -}|_{\mathcal{G}(\sigma)} = -I$ and $I_{\sigma, -}|_{\mathcal{G}(\sigma')} = I$ if $\sigma' \neq \sigma$. The adjoint of $I_{\sigma, -}$ is itself. Depending on context, the domain of $I_{\sigma, -}$ will be understood as $C_{q-1}(X; \mathcal{F}), C_q(Y; \mathcal{G})$ or $C_{q+1}(Y; \mathcal{G})$. The proof is divided into cases.

Case I. If $\sigma \in C_{q-1}(X; \mathcal{F})$, then $\overline{d_{q+1}^{\mathcal{G}}} = d_{q+1}^{\mathcal{G}}$ and $\overline{d_q^{\mathcal{F}}} = I_{\sigma, -}d_q^{\mathcal{F}}$. So $(\overline{d_q^{\mathcal{F}}})^\dagger \overline{d_q^{\mathcal{F}}} = (d_q^{\mathcal{F}})^\dagger I_{\sigma, -}I_{\sigma, -}d_q^{\mathcal{F}} = (d_q^{\mathcal{F}})^\dagger d_q^{\mathcal{F}}$. In other words, $\bar{\Delta} = \Delta$.

Case II. If $\sigma \in C^q(Y; \mathcal{G})$, then $\overline{d_q^{\mathcal{F}}} = d_q^{\mathcal{F}}\pi I_{\sigma, -}|_{C_q(X; \mathcal{F})}$ and $\overline{d_{q+1}^{\mathcal{G}}} = I_{\sigma, -}d_{q+1}^{\mathcal{G}}$. As

$$(\overline{d_{q+1}^{\mathcal{G}}})\bar{\Theta} = I_{\sigma, -}(d_{q+1}^{\mathcal{G}})\bar{\Theta} \subset I_{\sigma, -}C_q(X; \mathcal{F}) = C_q(X; \mathcal{F}),$$

we see that $\bar{\Theta} \subset \bar{\Theta}$. Similarly

$$(d_{q+1}^{\mathcal{G}})\bar{\Theta} = I_{\sigma, -}I_{\sigma, -}(d_{q+1}^{\mathcal{G}})\bar{\Theta} = I_{\sigma, -}(\overline{d_{q+1}^{\mathcal{G}}})\bar{\Theta} \subset I_{\sigma, -}C_q(X; \mathcal{F}) = C_q(X; \mathcal{F}),$$

we see that $\bar{\Theta} \supset \bar{\Theta}$, so $\bar{\Theta} = \bar{\Theta}$.

Then

$$\overline{\pi d_{q+1}^{\mathcal{G}}}|_{\bar{\Theta}} = \pi I_{\sigma, -}|_{C_q(X; \mathcal{F})}\pi d_{q+1}^{\mathcal{G}}|_{\bar{\Theta}}.$$

So

$$\begin{aligned} \bar{\Delta} &= \overline{d_q^{\mathcal{F}}{}^\dagger}(\overline{d_q^{\mathcal{F}}}) + \overline{\pi d_{q+1}^{\mathcal{G}}}|_{\bar{\Theta}}(\overline{\pi d_{q+1}^{\mathcal{G}}}|_{\bar{\Theta}})^\dagger \\ &= \pi I_{\sigma, -}|_{C_q(X; \mathcal{F})}\Delta(\pi I_{\sigma, -}|_{C_q(X; \mathcal{F})})^\dagger. \end{aligned}$$

Case III. If $\sigma \in C_{q+1}(Y; \mathcal{G})$, then $\overline{d_q^{\mathcal{F}}} = d_q^{\mathcal{F}}$ and $\overline{d_{q+1}^{\mathcal{G}}} = d_{q+1}^{\mathcal{G}}I_{\sigma, -}$. As $C_q(X; \mathcal{F}) \supset d_{q+1}^{\mathcal{G}}\bar{\Theta} = d_{q+1}^{\mathcal{G}}I_{\sigma, -}\bar{\Theta}$, we see that $I_{\sigma, -}\bar{\Theta} \subset \bar{\Theta}$. Similarly $I_{\sigma, -}\bar{\Theta} \subset \bar{\Theta}$, so $I_{\sigma, -}\bar{\Theta} = \bar{\Theta}$. Denote by $I_{\sigma, -}^\bar{\Theta} :$

$\overline{\Theta} \rightarrow \Theta$ the restriction of $I_{\sigma,-}$ on $\overline{\Theta}$. We have $\overline{\pi d_{q+1}^{\mathcal{G}}}|_{\overline{\Theta}} = \pi d_{q+1}^{\mathcal{G}}|_{\overline{\Theta}} = \pi d_{q+1}^{\mathcal{G}}|_{\Theta} I_{\sigma,-}^{\overline{\Theta}}$. Then

$$\begin{aligned} \overline{\Delta} &= \overline{d_q^{\mathcal{F}} \dagger d_q^{\mathcal{F}}} + \overline{\pi d_{q+1}^{\mathcal{G}}}|_{\overline{\Theta}} (\overline{\pi d_{q+1}^{\mathcal{G}}}|_{\overline{\Theta}}) \dagger \\ &= (d_q^{\mathcal{F}} \dagger d_q^{\mathcal{F}} + \pi d_{q+1}^{\mathcal{G}}|_{\Theta} I_{\sigma,-}^{\overline{\Theta}} (I_{\sigma,-}^{\overline{\Theta}}) \dagger) (\pi d_{q+1}^{\mathcal{G}}|_{\Theta}) \dagger \\ &= (d_q^{\mathcal{F}} \dagger d_q^{\mathcal{F}} + \pi d_{q+1}^{\mathcal{G}}|_{\Theta} (\pi d_{q+1}^{\mathcal{G}}|_{\Theta}) \dagger) \\ &= \Delta. \end{aligned}$$

□

The next corollary is the consequence of the persistent Hodge theorem.

Corollary 1. Persistent cosheaf Betti numbers are independent of the orientation of Y

Now let's discuss sheaves. Given two oriented simplicial complexes X, Y , if $X \subset Y$ and the orientation of X is identical to Y , let sheaf \mathcal{F} on X be the pullback of the sheaf \mathcal{G} on Y , then we have the following commutative diagram

$$\begin{array}{ccccccc} \dots & \xrightarrow{d} & C^{q-1}(X; \mathcal{F}) & \xrightarrow{d} & C^q(X; \mathcal{F}) & \xrightarrow{d} & C^{q+1}(X; \mathcal{F}) & \xrightarrow{d} & \dots \\ & & \pi \uparrow & & \pi \uparrow & & \pi \uparrow & & \\ \dots & \xrightarrow{d} & C^{q-1}(Y; \mathcal{G}) & \xrightarrow{d} & C^q(Y; \mathcal{G}) & \xrightarrow{d} & C^{q+1}(Y; \mathcal{G}) & \xrightarrow{d} & \dots \end{array}$$

where $\pi : C^q(Y; \mathcal{G}) \rightarrow C^q(X; \mathcal{F})$ is a projection map such that $\pi|_{\mathcal{G}(\sigma)}$ is the identity map if $\sigma \in X$, and $\pi|_{\mathcal{G}(\sigma)} = 0$ otherwise. Since π is a cochain map, it induces a map π^\bullet between sheaf cohomology groups of \mathcal{F} and \mathcal{G} , and the q -th persistent sheaf group is defined by

$$\pi^\bullet(H^q(Y; \mathcal{G}))$$

whose dimension is the q -th persistent sheaf Betti number. We can dualize the above diagram (i.e., reverse all arrows) and define the persistent sheaf Laplacian by the persistent cosheaf

Laplacian of the dualized diagram. More specifically, we have the following diagram

$$\begin{array}{ccccc}
C^{q-1}(X; \mathcal{F}) & \xrightleftharpoons[(d_{\mathcal{F}}^{q-1})^\dagger]{d_{\mathcal{F}}^{q-1}} & C^q(X; \mathcal{F}) & & \\
& & \downarrow (d_{\mathcal{F}, \mathcal{G}}^q)^\dagger & \searrow d_{\mathcal{F}, \mathcal{G}}^q & \\
& & C^q(Y; \mathcal{G}) & & \Theta_{\mathcal{F}, \mathcal{G}}^{q+1} \\
& & & & \swarrow \text{---} \\
& & & & C^{q+1}(Y; \mathcal{G}) \\
& & & \xrightleftharpoons[(d_{\mathcal{G}}^q)^\dagger]{d_{\mathcal{G}}^q} &
\end{array}$$

(note that an inner product space is self-dual) where $\Theta_{\mathcal{F}, \mathcal{G}}^{q+1} = \{x \in C^{q+1}(Y; \mathcal{G}) \mid (d_{\mathcal{G}}^q)^\dagger(x) \in C^q(X; \mathcal{F})\}$ and $d_{\mathcal{F}, \mathcal{G}}^q$ is the adjoint of $\pi(d_{\mathcal{G}}^q)^\dagger|_{\Theta_{\mathcal{F}, \mathcal{G}}^{q+1}} : \Theta_{\mathcal{F}, \mathcal{G}}^{q+1} \rightarrow C^q(X; \mathcal{F})$. We define the q -th persistent sheaf Laplacian $\Delta_q^{\mathcal{F}, \mathcal{G}}$ ² by

$$\Delta_q^{\mathcal{F}, \mathcal{G}} = (d_{\mathcal{F}, \mathcal{G}}^q)^\dagger d_{\mathcal{F}, \mathcal{G}}^q + d_{\mathcal{F}}^{q-1} (d_{\mathcal{F}}^{q-1})^\dagger.$$

The nullity of $\Delta_q^{\mathcal{F}, \mathcal{G}}$ is equal to the q -th persistent Betti number of the dualized diagram (since we dualize everything, the two cochain complexes become chain complexes). By the universal coefficient theorem for cohomology, π^\bullet and ι^\bullet have the same rank (here \bullet means the induced map between cohomology or homology groups). So the q -th persistent Betti number of the dualized diagram is equal to the q -th persistent sheaf Betti number. In other words, we have

$$\ker \Delta_q^{\mathcal{F}, \mathcal{G}} \cong \pi^\bullet(H^q(Y; \mathcal{G})).$$

More concretely, given a sheaf cochain complex

$$\dots \longrightarrow V^{q-1} \xrightarrow{d} V^q \xrightarrow{d} V^{q+1} \longrightarrow \dots,$$

we are actually working with the complex

$$\dots \longleftarrow V^{q-1} \xleftarrow{d^\dagger} V^q \xleftarrow{d^\dagger} V^{q+1} \longleftarrow \dots.$$

The universal coefficient theorem relates the above chain complex to the following cochain complex

$$\dots \longrightarrow (V^{q-1})^* \xrightarrow{(d^\dagger)^*} (V^q)^* \xrightarrow{(d^\dagger)^*} (V^{q+1})^* \longrightarrow \dots,$$

²Sometimes we may use the notation $\Delta_q^{X, Y}$.

which can be identified as the original sheaf cochain complex

$$\dots \longrightarrow V^{q-1} \xrightarrow{d} V^q \xrightarrow{d} V^{q+1} \longrightarrow \dots,$$

since there is a natural isomorphism between the dual functor and the adjoint functor.

Example 3.0.8. Consider the 1-dimensional simplicial complex Y



and the constant sheaf \mathbb{R} over Y . We compute $\Delta_0^{X,Y}$ when $X = \{1, 2\}$. The matrix representation of $(d_Y^0)^\dagger$ is

$$\begin{matrix} & 13 & 34 & 24 \\ 1 & \left(\begin{array}{ccc} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{array} \right) \\ 2 & & & \\ 3 & & & \\ 4 & & & \end{matrix}.$$

According to the definition of $\Theta_{X,Y}^0$, we want to find all elements of $C^1(Y; \mathbb{R})$ that are sent to $C^0(X; \mathbb{R}) = \text{span}\{1, 2\}$ by $(d_Y^0)^\dagger$. After a few steps of column Gauss elimination we get a new matrix representation of $(d_Y^0)^\dagger$

$$\begin{matrix} & 13 & 13 + 34 & -13 - 34 + 24 \\ 1 & \left(\begin{array}{ccc} -1 & -1 & 1 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{array} \right) \\ 2 & & & \\ 3 & & & \\ 4 & & & \end{matrix}.$$

From this representation, we see that for any vector $v = a13 + b(13 + 34) + c(-13 - 34 + 24)$ (a, b , and c are coefficients), $(d_Y^0)^\dagger v \in \text{span}\{1, 2\}$ if and only if a and b are both zero. In

other words, $\Theta_{X,Y}^0 = \text{span}\{-13 - 34 + 24\}$, $P = 3$, and the matrix representation of $(d_{X,Y}^0)^\dagger$ is

$$\frac{-13 - 34 + 24}{2} \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}.$$

Then, the matrix representation of $\Delta_0^{X,Y}$ is

$$\begin{pmatrix} 1/3 & -1/3 \\ -1/3 & 1/3 \end{pmatrix}$$

and its spectrum is $\{0, 2/3\}$.

Cellular sheaves on a labeled simplicial complex

In this section we construct a class of sheaves on a so called labeled simplicial complex. A labeled simplicial complex is a simplicial complex where each vertex is associated with a real number. We pay attention to label simplicial complexes because in application we often have a point cloud where each point is associated with some kind of quantity. For example, the atoms of a molecule can be seen as a point cloud, and each atom has its partial charge. If we build a Rips filtration from a labeled point cloud, then each complex in the filtration will be a label simplicial complex.

We first give a simple example. Suppose that there is a 1-dimensional labeled simplicial complex X where each vertex v_i is associated with a quantity $q_i \in \mathbb{R}$. Denote the edge connecting v_i and v_j by e_{ij} . We can define a sheaf \mathcal{S} on X where each stalk is \mathbb{R} , and for the face relation $v_i \leq e_{ij}$, the morphism $\mathcal{S}_{v_i \leq e_{ij}}$ is the scalar multiplication by q_j/r_{ij} where r_{ij} is the length of e_{ij} . The assignment $q_i \rightarrow v_i$ and $q_i q_j / r_{ij} \rightarrow e_{ij}$ is a global section, since $\mathcal{S}_{v_i \leq e_{ij}}(q_i) = \mathcal{S}_{v_j \leq e_{ij}}(q_j) = q_i q_j / r_{ij}$. If we think of q_i and q_j as partial charges on atoms, then the quantity $q_i q_j / r_{ij}$ is the potential energy.

The above sheaf can be generalized to high-dimensional labeled simplicial complexes. Let $F : X \rightarrow \mathbb{R}$ be a nowhere zero function. We can define a sheaf where each stalk is \mathbb{R} , and

for the face relation $[v_0, \dots, v_n] \leq [v_0, \dots, v_n, v_{n+1}, \dots, v_m]$ (here orientation is not relevant), the linear morphism $\mathcal{S}([v_0, \dots, v_n] \leq [v_0, \dots, v_n, v_{n+1}, \dots, v_m])$ is the scalar multiplication by

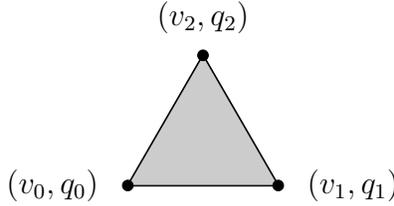
$$\frac{F([v_0, \dots, v_n])q_{n+1} \cdots q_m}{F([v_0, \dots, v_n, v_{n+1}, \dots, v_m])}.$$

This is indeed a sheaf since if we have $[v_0, \dots, v_n] \leq [v_0, \dots, v_m] \leq [v_0, \dots, v_l]$, then

$$\frac{F([v_0, \dots, v_m])q_{m+1} \cdots q_l}{F([v_0, \dots, v_l])} \frac{F([v_0, \dots, v_n])q_{n+1} \cdots q_m}{F([v_0, \dots, v_m])} = \frac{F([v_0, \dots, v_n])q_{n+1} \cdots q_l}{F([v_0, \dots, v_l])}.$$

The assignment $q_{i_0} \cdots q_{i_n} / F([v_{i_0}, \dots, v_{i_n}]) \rightarrow [v_{i_0}, \dots, v_{i_n}]$ is a nontrivial global section.

Example 3.0.9. Consider the oriented simplicial complex $\{v_0, v_1, v_2, v_0v_1, v_1v_2, v_0v_2, v_0v_1v_2\}$



where $q_i \in \mathbb{R}$ is associated to v_i . Let r_{01}, r_{12}, r_{02} be the lengths of e_{01}, e_{12}, e_{02} . We can define the above sheaf on this complex where F maps every vertex to 1, every edge e_{ij} to its length r_{ij} , and the 2-simplex $[v_0, v_1, v_2]$ to $r_{01}r_{12}r_{02}$. The matrix representation of d^0 is

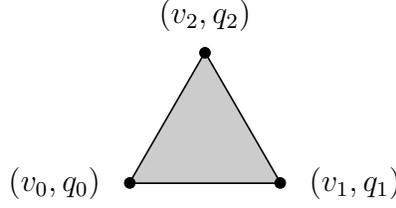
$$\begin{array}{c} v_0 \quad v_1 \quad v_2 \\ v_0v_1 \begin{pmatrix} -q_1/r_{01} & q_0/r_{01} & 0 \\ 0 & -q_2/r_{12} & q_1/r_{12} \\ -q_2/r_{02} & 0 & q_0/r_{02} \end{pmatrix}, \\ v_1v_2 \\ v_0v_2 \end{array}$$

and the matrix representation of d^1 is

$$v_0v_1v_2 \begin{pmatrix} v_0v_1 & v_0v_2 & v_1v_2 \\ \frac{q_2}{r_{02}r_{12}} & \frac{-q_1}{r_{01}r_{12}} & \frac{q_0}{r_{01}r_{02}} \end{pmatrix}.$$

Note that many alternative sheaf constructions are available by appropriate choices of F .

Example 3.0.10. We consider the F defined in the same way as in Example 3.0.9 to evaluate the spectra of sheaf Laplacians. Consider the 2-simplex



whose edges are all of length 1. The Δ_0 is

$$\begin{pmatrix} q_1^2 + q_2^2 & -q_0q_1 & -q_0q_2 \\ -q_0q_1 & q_0^2 + q_2^2 & -q_1q_2 \\ -q_0q_2 & -q_1q_2 & q_0^2 + q_1^2 \end{pmatrix}$$

and its eigenvalues are $\{q_0^2 + q_1^2 + q_2^2, q_0^2 + q_1^2 + q_2^2, 0\}$ and the corresponding eigenvectors are $(-q_1/q_0, 1, 0)^T$, $(-q_2/q_0, 0, 1)^T$, and $(q_0/q_2, q_1/q_2, 1)^T$. Moreover, Δ_1 is

$$\begin{pmatrix} q_0^2 + q_1^2 + q_2^2 & 0 & 0 \\ 0 & q_0^2 + q_1^2 + q_2^2 & 0 \\ 0 & 0 & q_0^2 + q_1^2 + q_2^2 \end{pmatrix}$$

and its only eigenvalue is $q_0^2 + q_1^2 + q_2^2$. This example shows that the eigenvalues of Δ_0 and Δ_1 are dependent on the amplitude of q_i , which allows the embedding of non-geometric information in practical applications. However, they are not sensitive to the sign of q_i . Therefore, a (persistent) sheaf Dirac theory as an extension of recent Dirac formulation or quantum persistent homology [2] may enable us to further eliminate the sign degeneracy.

Experiments of persistent sheaf Laplacians

Given a labeled point cloud P (i.e., a point cloud with a nonzero quantity q_i associated with each point v_i), we can build a Rips or Alpha filtration from it and construct a sheaf \mathcal{S}_t on each X_t consistently as described in section 3 provided a suitable global $F : 2^P \rightarrow \mathbb{R}$ is chosen. If we take F_t to be the restriction of $F : 2^P \rightarrow \mathbb{R}$ on X_t and construct \mathcal{S}_t , then we ensure that the stalks of σ and τ and the restriction map between σ and τ remain the

same for any \mathcal{S}_t containing $\sigma \leq \tau$. Since \mathcal{S}_t is the pullback of \mathcal{S}_{t+p} for any t and p , we get a persistent module of sheaf cochain complexes and can compute the spectra of persistent sheaf Laplacians. In this section, we calculate the spectra of persistent sheaf Laplacians for a few examples of point clouds in this way. Some examples are the vertices of simple geometrical shapes and some are the coordinates of atoms of molecules. We assign the quantities q_i to simple geometrical shapes, and take the partial charges as q_i for molecules. An Alpha filtration $\{X_r\}$ is built for each labeled point cloud, parametrized by radius r . We choose F such that F maps every vertex v_i to 1, every edge $v_i v_j$ to the length of itself r_{ij} , and every 2-cell $v_i v_j v_k$ to the product of lengths of its edges $r_{ij} r_{ik} r_{jk}$. The spectrum of $\Delta_d^{X_r, X_{r+p}}$ for $d = 0, 1$ and selected r, p will be calculated. The radius r will be a multiple of 0.01 or 0.01Å. Many information can be extracted from a spectrum, but for simplicity here we only plot the minimal nonzero eigenvalue and the nullity against the radius r . The minimal nonzero eigenvalue of the persistent sheaf Laplacian $\Delta_d^{X_r, X_{r+p}}$ is denoted by $\lambda_d^{r,p}$ and the d -th persistent sheaf Betti number of the pair (X_r, X_{r+p}) (i.e., the nullity of $\Delta_d^{X_r, X_{r+p}}$) is denoted by $\beta_d^{r,p}$.

The examples of simple shapes are the vertex sets of a 2-dimensional square or a 2-dimensional trapezoid as shown in Figure C.1 with different choices of local property q_i . More specifically, we consider four labeled point clouds and two point clouds in \mathbb{R}^2 . The two point clouds are $\{(0, 0), (1, 0), (1, 1), (0, 1)\}$ and $\{(0, 0), (1, 0), (3/4, \sqrt{15}/4), (1/4, \sqrt{15}/4)\}$. For $\{(0, 0), (1, 0), (1, 1), (0, 1)\}$ we assign $q = \pm 1$ to $(0, 1)$ and $q = 1$ to the rest. For $\{(0, 0), (1, 0), (3/4, \sqrt{15}/4), (1/4, \sqrt{15}/4)\}$ we assign $q = \pm 1$ to $(1/4, \sqrt{15}/4)$ and $q = 1$ to the rest. For the two point clouds we construct the constant sheaf and compute persistent sheaf Laplacians, whose spectra coincide with persistent Laplacians. For the four labeled point clouds we construct filtrations of sheaves in the way described earlier in this section. The results are shown in Figures C.2 and C.3. The first thing we can infer is that, the minimal nonzero eigenvalue and the nullity usually change significantly when the topology of X_r changes (when p is nonzero, both of them change at $r - p$ rather than r). When we consider

labeled point clouds instead of point clouds, sometimes the minimal nonzero eigenvalue or the nullity show less change (compare (a)(c) and (b)(d) in Figure C.2 and Figure C.3). If we compare the results of the square and the trapezoid, we see more changes in the results of the trapezoid. This is because the filtration constructed from the vertex set of trapezoid contains more different complexes. We also observe that the change of signs of q_i does not affect the minimal nonzero eigenvalue and the nullity, though the eigenspaces of Laplacians are different.

Next we study the molecule CB8 [118] shown in Figure C.4. We associate each atom with the corresponding partial charge (obtained using [107]). The results for CB8 are shown in Figures C.5 and C.6. Due to complexity of the molecule, it is very difficult to explain the spectral details of the system. However, this information can be very useful for machine learning analysis.

Finally, to demonstrate our method for practical problems, we study a small protein called bacteriocin AS-48 (PDB ID: 1E68) [54]. We select the model 1 of AS-48 and compute the pqr file by PDB2PQR with the Amber force field [77]. For the sake of faster computation, we only use the coordinates of carbon atoms as the point cloud. Results are shown in Figures C.8 and C.9.

CHAPTER 4

OTHER PERSISTENT TOPOLOGICAL LAPLACIANS

Persistent Laplacians for simplicial maps

The classical filtration of simplicial complexes only represents one type of shape evolution. We also need tools to study more general shape evolution, such as sparsification of a simplicial complex. This requires us to consider general simplicial maps rather than inclusion maps. Gülen et al. [63] developed a theory of persistent Laplacians for a simplicial map $f : X \rightarrow Y$ (the chain map induced by a simplicial map is not necessarily injective). Suppose $f : X \rightarrow Y$ is a simplicial map,

$$\begin{array}{ccccccc}
 \cdots & \longrightarrow & C_{q+1}(X) & \xrightarrow{\partial_{q+1}^X} & C_q(X) & \xrightarrow{\partial_q^X} & C_{q-1}(X) & \longrightarrow & \cdots \\
 & & \downarrow f_{q+1} & & \downarrow f_q & & \downarrow f_{q-1} & & \\
 \cdots & \longrightarrow & C_{q+1}(Y) & \xrightarrow{\partial_{q+1}^Y} & C_q(Y) & \xrightarrow{\partial_q^Y} & C_{q-1}(Y) & \longrightarrow & \cdots
 \end{array}$$

where $f_q : C_q(X) \rightarrow C_q(Y)$ is induced by f . Different from the original q -th persistent Laplacian, we need to define two subspaces

$$C_{q+1}(Y) \supset \Theta_{q+1}^{Y \leftarrow X} = \{c \in C_{q+1}(Y) \mid \partial_{q+1}^Y(c) \in f_q(\ker \partial_q^X)\}$$

and

$$C_{q-1}(X) \supset \Theta_{q-1}^{X \rightarrow Y} = \{c \in C_{q-1}(X) \mid (\partial_q^X)^*(c) \in (\ker f_q)^\perp\},$$

and then use the restrictions of ∂_{q+1}^Y and $(\partial_q^X)^*$ to them to construct the q -th persistent Laplacian for f . The q -th persistent Laplacian for a simplicial map has a more symmetric expression, and the proof of the persistent Hodge theorem is more complicated.

Digraphs and path homology

The motivation behind path homology is to construct a homology theory of digraphs such that directional information of edges is encoded and higher dimensional homology groups are less likely to be non-trivial. Path homology was proposed by Grigor'yan et al. [57] and developed in [58, 59, 60, 61, 86]. A summary of recent advances in path homology

of digraphs can be found in [55]. There are also other (co)homology theories of digraphs [22, 92, 95, 108]. Recall that a digraph is a pair $G = (V, E)$ where E is a set of ordered pairs of vertices. We only consider digraphs without self-loops. In path homology, the focus is on the paths in a digraph. An *allowed q -path* is an ordered finite sequence of vertices $\{x_0, \dots, x_q\}$ such that $(x_i, x_{i+1}) \in E$ for all $i = 0, \dots, q - 1$. The length of an allowed path is its natural dimension, so we can try to define the q -th chain group as the vector space consisting of formal linear combinations of allowed q -paths with coefficients in \mathbb{R} , denoted by \mathcal{A}_q , and define the boundary map ∂_q by

$$\partial_q\{x_0, \dots, x_q\} = \sum_{i=0}^q (-1)^i \{x_0, \dots, \hat{x}_i, \dots, x_q\}$$

then formally we can show that $\partial^2 = 0$. However, $\partial_q\{x_0, \dots, x_q\}$ may include paths that are not allowed paths. To solve this problem, we need to introduce some general concepts first.

Definition 4.0.1. Suppose X is a finite set. An elementary p -path is a sequence $[x_0, \dots, x_p]$ of $p + 1$ elements of X . The space generated by all elementary p -paths with coefficient in \mathbb{R} is denoted by $\Lambda_p(X)$. The q -th non-regular boundary map is given by

$$\partial_q^{\text{nr}}[x_0, \dots, x_q] = \sum_{i=0}^p [x_0, \dots, \hat{x}_i, \dots, x_q].$$

One can prove this is a chain complex. Among all the paths, a path that lingers at a vertex (for some i , $x_i = x_{i+1}$) is considered a degenerate path since we are not interested in self-loops.

Definition 4.0.2. A path $[x_0, \dots, x_q]$ over X where $x_i \neq x_{i+1}$ for each i is called regular. The space generated by all regular q -paths is denoted by \mathcal{R}_q .

We define a new boundary operator ∂_q between regular paths. When computing $\partial_q([x_0, \dots, x_q])$, we first compute $\partial_q^{\text{nr}}([x_0, \dots, x_q])$ and treat all irregular paths arising from it as zeros. One can still verify that $\partial^2 = 0$ [56].

Now given a digraph $G = (V, E)$, every \mathcal{A}_q is a subspace of \mathcal{R}_q .

$$\begin{array}{ccccccc} \dots & \xrightarrow{\partial_{q+2}} & \mathcal{R}_{q+1} & \xrightarrow{\partial_{q+1}} & \mathcal{R}_q & \xrightarrow{\partial_q} & \mathcal{R}_{q-1} & \xrightarrow{\partial_{q-1}} & \dots \\ & & \uparrow \text{---} & & \uparrow \text{---} & & \uparrow \text{---} & & \\ & & \mathcal{A}_{q+1} & & \mathcal{A}_q & & \mathcal{A}_{q-1} & & \end{array}$$

One way to make $\partial_q : \mathcal{A}_q \rightarrow \mathcal{A}_{q-1}$ well defined is to restrict ∂_q to the subspace $\mathcal{A}_q \cap \partial_q^{-1} \mathcal{A}_{q-1}$. We have to verify that $\partial_q(\mathcal{A}_q \cap \partial_q^{-1} \mathcal{A}_{q-1}) \subset \mathcal{A}_{q-1} \cap \partial_{q-1}^{-1} \mathcal{A}_{q-2}$. $\partial_q(\mathcal{A}_q \cap \partial_q^{-1} \mathcal{A}_{q-1}) \subset \mathcal{A}_{q-1}$ is true by definition, and $\partial_q(\mathcal{A}_q \cap \partial_q^{-1} \mathcal{A}_{q-1}) \subset \partial_{q-1}^{-1} \mathcal{A}_{q-2}$ is true because of $\partial^2 = 0$. Therefore, we have the chain complex

$$\dots \longrightarrow \mathcal{A}_{q+1} \cap \partial_{q+1}^{-1} \mathcal{A}_q \xrightarrow{\partial_{q+1}} \mathcal{A}_q \cap \partial_q^{-1} \mathcal{A}_{q-1} \xrightarrow{\partial_q} \mathcal{A}_q \cap \partial_q^{-1} \mathcal{A}_{q-1} \longrightarrow \dots$$

and the definition of a path homology group is straightforward. The q -th chain group $\mathcal{A}_q \cap \partial_q^{-1} \mathcal{A}_{q-1}$ is called the space of ∂ -invariant q -paths on G , denoted by Ω_q ¹.

As to the geometrical interpretation of path homology, we only know for sure that non-reduced H_0 is the number of connected components of the underlying undirected graph. It is not easy to relate higher dimensional path homology groups to features of the digraph. Some characterizations of path homologies of certain families of small digraphs were obtained by Chowdhury et al [34]. Since directional information of edges is encoded in path homology, path homology can be used to distinguish network motifs [35] and isomers in molecular and materials sciences [25]. We can also quantify the significance of a node in a network by observing the changes of path homology after the deletion of the node [25].

Since Ω_q inherits the inner product structure from \mathcal{A}_q , the so-called path Laplacian² can be defined. We can use path Laplacians [53, 55, 134] to distinguish among digraphs that path homology cannot. For example, according to [57, Theorem 5.4], the following two digraphs G_L and G_R (see Figure 4.1) have the same path homology. But the spectrum of the 0-th path Laplacian of G_L is $\{0, 3, 3\}$ and that of G_R is $\{0, 2, 4, 4\}$.

¹If a digraph is not simple, there will be two choices of ∂_q [57] that might be suitable for different problems [74].

²Another type of path Laplacians was proposed by Estrada [46], and was generalized and applied in molecular biology in [89].

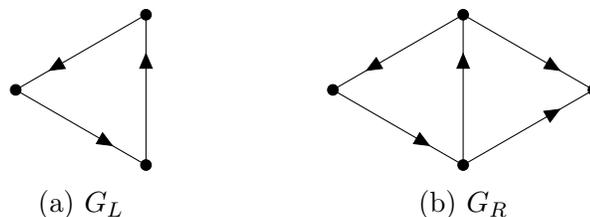


Figure 4.1 Two graphs that have the same path homology.

Persistent path homology was proposed by Chowdhury and Mémoli [35] to study a digraph where each edge e has a weight $w(e)$. A filtration of digraphs $\{G_d\}$ is constructed such that $e \in G_d$ if and only if $w(e) \leq d$. Wang and Wei [134] introduced persistent path Laplacians based on Chowdhury and Mémoli’s work. They suggested that persistent path Laplacians can be applied to study molecules, where much information can be encoded in a digraph.

Example 4.0.1. Given a weighted digraph, we can build a filtration $\{G_d\}$ such that $e \in G_d$ iff $w(e) \leq d$. Two weighted graphs whose path Betti numbers are the same for every G_d may have different path Laplacians. This is shown in Figure 4.2.

Besides path homology, there is another homology theory of digraphs based on clique complexes. A q -clique of a digraph $G = (V, E)$ is an ordered subset of vertices $\sigma = (x_0, x_1, \dots, x_{q-1})$ such that for $i < j$ we have $(x_i, x_j) \in E$. The directed flag complex $\text{dFl}(G)$ is the simplicial complex on V with q -simplices the $(q+1)$ -cliques of G . A persistent Laplacian theory of directed flag complexes was proposed by Jones and Wei [75].

Hypergraphs and hyperdigraphs

Hypergraphs can be thought of as a generalization of simplicial complexes. A hypergraph is a pair (V, E) where E is a subset of the power set of V . A element of E is called a hyperedge, and a hyperedge consisting of $q+1$ elements is called a q -hyperedge. To define a chain complex for hypergraphs, the problem here is identical to what we encounter in the study of digraphs. If we define the q -th chain group to be the vector space generated by q -hyperedges, the boundary map is not well-defined. One solution is to consider the *associated simplicial complex (simplicial closure)* of a hypergraph [102], i.e., the minimal

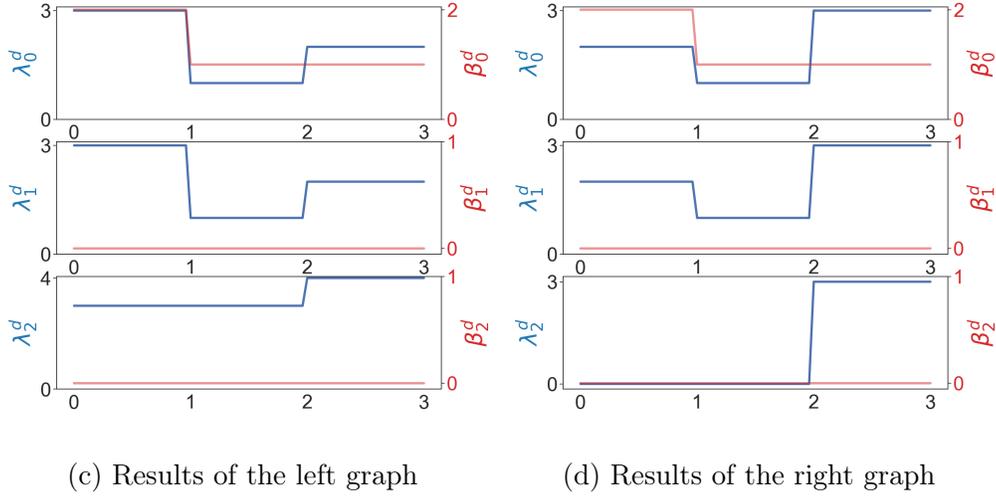


Figure 4.2 The x axis represents the weight. As usual, λ and β represent the minimal eigenvalues and Betti numbers.

simplicial complex that contains the given hypergraph. Another solution inspired by the path homology is the *embedded homology* [13]. If we look at the chain complex of the associated simplicial complex, each simplicial chain group C_q contains D_q , the vector space generated by hyperedges. We only need to restrict the domain of the simplicial boundary operator to

$$\text{Inf}_q = D_q \cap \partial_q^{-1}(D_q),$$

and then the boundary operator is well-defined.

A hyperdigraph is a hypergraph where each hyperedge is ordered³, the embedded homology of which can be defined similarly [24]. Persistent homology and persistent Laplacians of hypergraphs and hyperdigraphs are studied in [13, 24, 110, 112]. Other approaches regarding the homology and Laplacian of hypergraph includes [36, 45, 73, 76, 100, 101].

³There are other definitions of a hyperdigraph [4, 127].

$$\begin{array}{ccccccc}
\cdots & \xrightarrow{\partial_{q+2}} & C_{q+1} & \xrightarrow{\partial_{q+1}} & C_q & \xrightarrow{\partial_q} & C_{q-1} & \xrightarrow{\partial_{q-1}} & \cdots \\
& & \uparrow \text{---} \downarrow & & \uparrow \text{---} \downarrow & & \uparrow \text{---} \downarrow & & \\
& & D_{q+1} & & D_q & & D_{q-1} & & \\
& & \uparrow \text{---} \downarrow & & \uparrow \text{---} \downarrow & & \uparrow \text{---} \downarrow & & \\
& & D_{q+1} \cap \partial_{q+1}^{-1}(D_q) & & D_q \cap \partial_q^{-1}(D_{q-1}) & & D_{q-1} \cap \partial_{q-1}^{-1}(D_{q-2}) & &
\end{array}$$

Figure 4.3 C_q is the q -th chain group of the associated simplicial complex of H , and D_q is the vector space generated by q -hyperedges.

Mayer homology

We can also define persistent homology and persistent Laplacians for the so-called N -chain complexes [121]. An N -chain complex is a sequence of abelian groups and group morphisms (V, d) where $d^N = 0$. A simplicial complex actually gives rise to a N -chain complex. Recall that in a simplicial chain complex the boundary operator is given by

$$\partial[v_{a_0}, \dots, v_{a_q}] = \sum_i (-1)^i [v_{a_0}, \dots, \hat{v}_{a_i}, \dots, v_{a_q}].$$

For a prime number N , let $\xi = e^{2\pi i/N}$, we can define another boundary operator d by

$$d[v_{a_0}, \dots, v_{a_q}] = \sum_i \xi^i [v_{a_0}, \dots, \hat{v}_{a_i}, \dots, v_{a_q}]$$

and prove that $d^N = 0$. Even though N -chain complex is not a chain complex, observe that

$$C_{N+q}(X; \mathbb{C}) \xrightarrow{d^n} C_{N+q-n}(X; \mathbb{C}) \xrightarrow{d^{N-n}} C_N(X; \mathbb{C})$$

resembles a part of chain complex, we can define the Mayer homology group $H_{N+q-n, N-n}(X)$ by

$$H_{N+q-n, N-n}(X) = \ker d^{N-n} / \text{im } d^n.$$

Therefore, it is not surprising that the theory of persistent homology and persistent Laplacians can be extended to the setting of N -chain complexes. One advantage of using N -chain complexes is that the number of Betti numbers and Laplacians is much larger than traditional persistent homology and persistent Laplacians, and we can fine tune N for a specific problem.

Dirac operators are closely related to combinatorial Laplacians. If we think of all combinatorial Laplacians as a single operator $dd^* + d^*d = (d + d^*)^2$ on V , then the q -th Dirac operator is the restriction of the square root $d + d^*$ on $V_0 \oplus \cdots \oplus V_{q+1}$. We can also see this by direct computation. The square of D_q is

$$\begin{array}{c} V_0 \quad V_1 \quad V_2 \quad \cdots \quad V_q \quad V_{q+1} \\ \begin{array}{c} V_0 \\ V_1 \\ V_2 \\ \vdots \\ V_q \\ V_{q+1} \end{array} \left(\begin{array}{cccccc} [\Delta_0] & 0 & 0 & \cdots & 0 & 0 \\ 0 & [\Delta_1] & 0 & \cdots & 0 & 0 \\ 0 & 0 & [\Delta_2] & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & [\Delta_q] & 0 \\ 0 & 0 & 0 & \cdots & 0 & [\Delta_{q+1,-}] \end{array} \right) \end{array}$$

where Δ_q is the q -th combinatorial Laplacian. Therefore, the square of any eigenvalue λ of a Dirac operator must be an eigenvalue of a combinatorial Laplacian.

Recall that to define persistent Laplacians, we construct an auxiliary subspace $\Theta_{q+1}^{X,Y}$ of $C_{q+1}(Y)$ and a map $\partial_{q+1}^{X,Y} : \Theta_{q+1}^{X,Y} \rightarrow C_q(X)$. Since $C_q(X)$ is actually a subspace of $\Theta_q^{X,Y}$, we have a boundary map $\iota \circ \partial_q^{X,Y} : \Theta_{q+1}^{X,Y} \rightarrow \Theta_q^{X,Y}$. All $\Theta_q^{X,Y}$ and $\iota \partial_q^{X,Y}$ constitute an auxiliary chain complex

$$\begin{array}{ccccccc} & & C_{q+1}^X & & C_q^X & & C_{q-1}^X & & \cdots \\ & \nearrow & \downarrow \iota & \nearrow \partial_{q+1}^{X,Y} & \downarrow \iota & \nearrow \partial_q^{X,Y} & \downarrow \iota & \nearrow \partial_{q-1}^{X,Y} & \\ \cdots & & \Theta_{q+1}^{X,Y} & & \Theta_q^{X,Y} & & \Theta_{q-1}^{X,Y} & & \end{array}$$

The q -th persistent Dirac operator of simplicial complexes $X \subset Y$ is defined by the q -th Dirac operator on this auxiliary complex, and it is easy to generalize persistent Dirac operators to other settings such as digraphs and hyperdigraphs [126]. The square of a persistent Dirac operator is not necessarily a block matrix consisting of persistent Laplacians.

CHAPTER 5

UNDERSTANDING DOMINANT VARIANTS OF SARS-COV-2 USING LAPLACIANS

The aim of this chapter is to demonstrate the applications of Laplacians in biological sciences through the study on dominant variants of SARS-CoV-2. SARS-CoV-2 enters the host cell via either endosomes or plasma membrane fusion. In both ways, the S protein of SARS-CoV-2 first attaches to the host cell-surface protein, angiotensin converting enzyme 2 (ACE2). The receptor binding domain (RBD) of the S protein is essential for the entry. This is why the binding free energy (BFE) of the RBD-ACE2 complex is a measure of viral infectivity. A mutation of viral RBD induces BFE changes. A positive (resp. negative) BFE change indicates the strengthening (resp. weakening) of the protein-protein binding. If we know how to predict BFE changes, we can predict infectivity of new variants. TDA methods such as persistent homology and persistent Laplacian are useful because they are able to encode viral RBD structural information.

Biological background

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of the ongoing global coronavirus disease 2019 (COVID-19) pandemic. Its evolution and future direction are of major concern. It was well established that the emergence of SARS-CoV-2 new variants is dictated by mutation-induced infectivity strengthening [30] and antibody resistance (or vaccine breakthrough) [132], two molecular mechanisms that determined the natural selection at the population scale. More specifically, the binding of the viral spike protein, particularly the receptor-binding domain (RBD), to the human receptor angiotensin-converting enzyme 2 (ACE2) facilitates the entry of the virus into host cells [71, 129]. In early 2020, it was hypothesized that natural selection favors those SARS-CoV-2 RBD mutations that strengthen the RBD-ACE2 binding, which leads to higher viral infectivity [30]. The hypothesis was initially supported by the frequency analysis of 89 single RBD mutations found from the genotyping of 15,140 complete SARS-CoV-2 genome samples [30] and later confirmed

beyond doubt by the evolution pattern of 651 RBD mutations found from the genotyping of 506,768 SARS-CoV-2 genomes extracted from COVID-19 patients up to early 2021 [131].

The vaccine breakthrough mechanism was not discovered until vaccines became widely available in industrialized countries in the summer of 2021. It was found that an RBD mutation that weakens the viral infectivity had an unusually high observed frequency in 2,298,349 complete SARS-CoV-2 genomes isolated from patients. This abnormal statistics was found to strongly correlate with the vaccination rates in a few industrialized countries, including Denmark, the United Kingdom, France, Bulgaria, the United States, etc [132]. To understand this correlation, the mutational impact of a set of 130 antibodies extracted from Covid patients that targets the RBD was studied. It was found that the abnormal mutation on the RBD has a very strong ability to disrupt the binding of most antibody-RBD complexes, which gives rise to antibody resistance (or vaccine breakthrough) at the population scale [132].

As discussed above, the reveal of the natural selection mechanisms of SARS-CoV-2 evolution is a typical example of a data-driven discovery that cannot be achieved by individual experimental laboratories. In fact, the discovery utilized results from tens of thousands of experimental laboratories around the world [30, 132]. Machine learning, including deep learning and data-driven approach, played an essential role in the discovery. Deep learning methods can offer some of the most accurate predictions of biomolecular properties, including the binding affinity of protein-protein interactions (PPIs). This approach becomes particularly advantageous and outperforms other methods when good-quality experimental data are available. However, structure-based machine learning, including deep learning methods encounter difficulties in PPI predictions due to their intricate structural complexity and high dimensionality.

Although sequence-based approaches offer good predictions of mutational impacts on proteins, structure-based methods outperform other approaches [106]. In machine-learning-assisted directed evolution and protein engineering and machine-learning-based PPI and

protein folding stability predictions, mutant structures are typically not available and are conventionally created by computational means for the machine learning predictions [19, 16, 26, 27, 91], which is a source of errors. It is interesting and important to quantify such errors. Fortunately, since SARS-COV-2 variants are some of the most studied subjects, some of their three-dimensional (3D) structures are available in the literature, which offers an opportunity for in-depth analysis and comparison.

PTL analysis of structural changes

We are interested in both the structural changes of the wild type RBD induced by mutations and the structural changes of the wild type RBD or mutant RBDs induced by their binding to ACE2. To quantify structural changes we first perform alignment of structures and calculate the distances between corresponding atoms (e.g., C_α). Then, we compute PTLs of different structures to further characterize their structural changes.

5.0.1 PTL analysis of RBD structural changes induced by mutations

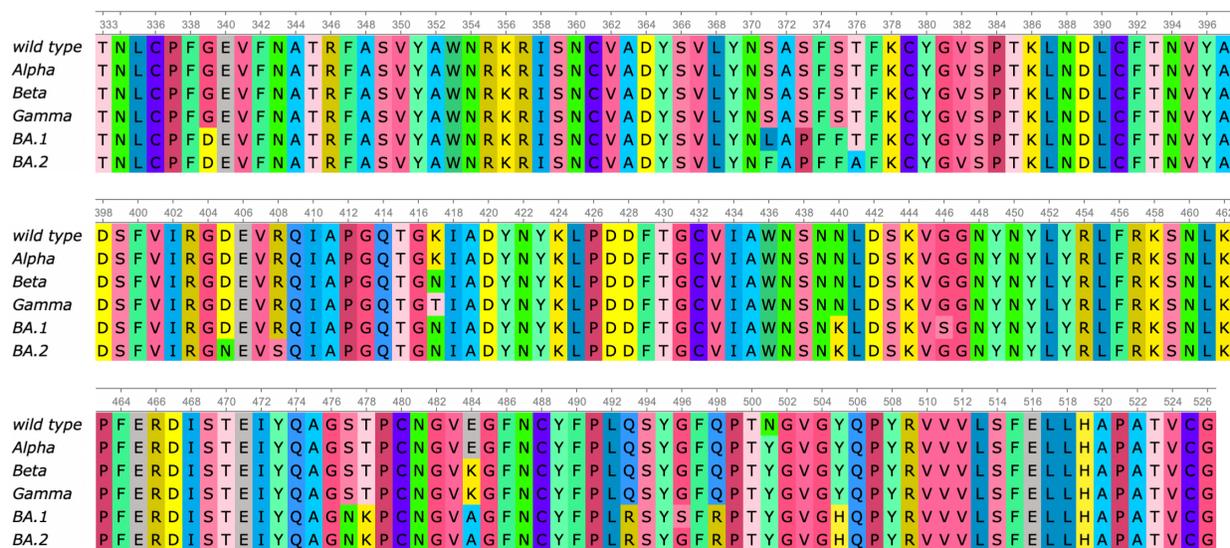


Figure 5.1 Sequence alignment of RBDs of the wild type, Alpha, Beta, Gamma, BA.1, and BA.2. Alpha has one RBD mutation N501Y. Beta has three RBD mutations K417N, E484K, and N501Y. Gamma has three RBD mutations K417T, E484K, and N501Y. BA.1 has 15 RBD mutations G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, and Y505H. BA.2 has 16 RBD mutations G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, and Y505H.

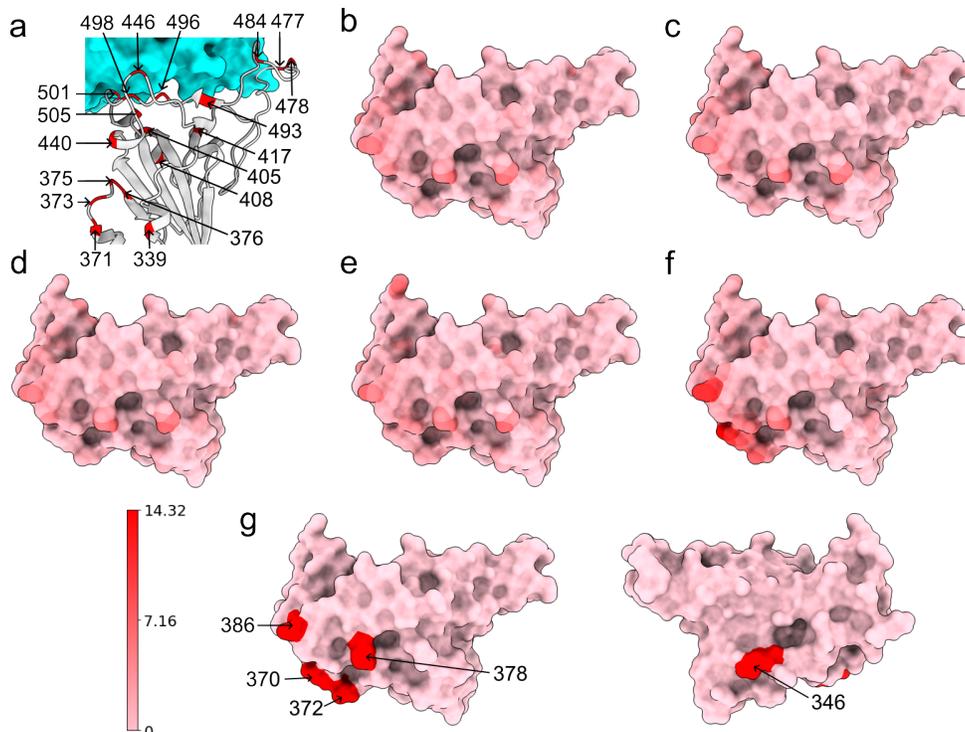


Figure 5.2 (a) Wild type RBD-ACE2 complex. The RBD is colored by light grey and mutated residues in Alpha, Beta, Gamma, BA.1 and BA.2 are marked. (b, c, d, e, f) Atoms of the wild type RBD are colored by their distances to corresponding atoms in a mutant RBD. Subfigures (a), (b), (c), (b), and (f) corresponds to the Alpha, Beta, Gamma, BA.1, and BA.2 variants, respectively. Pink and red corresponds to 0\AA and 14.32\AA respectively. For each mutant we record the residues that have at least one atom whose distance to the corresponding atom in the wild type RBD is larger than 7.16\AA . In Alpha, Beta, and Gamma, such residue is R346. In BA.1, such residue is K386. In BA.2, such residues are N370, A372, K378, and K386. These residues are marked in (g). (Plots generated by ChimeraX [104].)

To understand the structural differences of RBD between the wild type and mutants in RBD-ACE2 complex, we align the RBDs of SARS-CoV-2 variants Alpha (PDB ID: 8DLK[93]), Beta (PDB ID: 8DLN[93]), Gamma (PDB ID: 8DLQ[93]), BA.1 (PDB ID: 7T9L[94]), and BA.2 (PDB ID: 7XB0[82]) along with the wild type RBD (PDB ID: 6M0J[80]) in Figures 5.2 and 5.3. For Alpha, Beta, Gamma, BA.1, and BA.2, the maximal distances between corresponding atoms of mutant RBDs and the wild-type RBD are 9.14\AA , 9.33\AA , 9.87\AA , 7.44\AA , and 14.32\AA respectively. For each mutant, the residues are recorded if they have at least one atom whose distance to the corresponding atom in wild-type RBD is more than 7.16\AA , which is half of the maximal distance, 14.32\AA . For variants Alpha, Beta, and

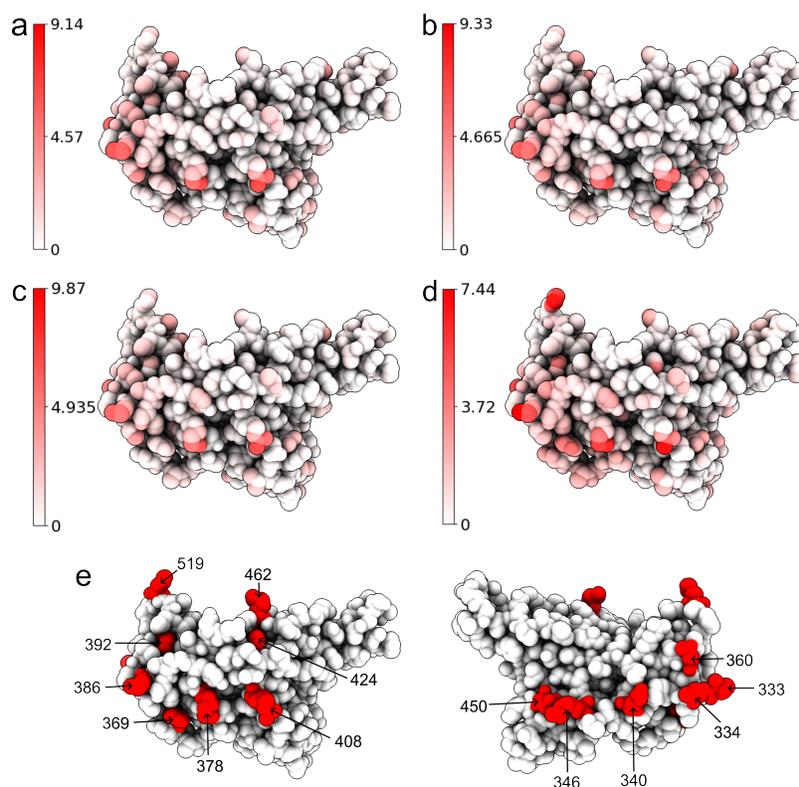


Figure 5.3 Atoms of the wild type RBD are colored by their distances to corresponding atoms in a mutant RBD. Subfigures (a), (b), (c), and (b) corresponds to the Alpha, Beta, Gamma, and BA.1, variants, respectively. Each alignment has its own color range. For each mutant, we record the residues that have at least one atom whose distance to the corresponding atom in the wild type RBD is more than half of the maximal distance (4.57\AA , 4.67\AA , 4.94\AA , and 3.72\AA) between corresponding atoms. In Alpha, such residues are T333, R346, K378, K386, R408, and N450. In Beta and Gamma, such residues are T333, R346, K378, K386, and R408. In BA.1, such residues are T333, N334, E340, R346, N360, D364, Y369, K378, K386, F392, R408, K424, N450, K462, and H519. These residues are marked in (e). (Plots generated by ChimeraX [104].)

Gamma, such a residue is R346, while in BA.1 such residue is K386. BA.2 has most such residues, which are N370, A372, K378, and K386, containing atoms deviating from the wild type. However, these residues are not in the receptor-binding motif (RBM, residues 438-506) that interacts directly with ACE2.

Alternatively, for Alpha, Beta, Gamma, and BA.1 variants, we can change the threshold from 7.16\AA to the half of maximal distance (4.57\AA , 4.67\AA , 4.94\AA , and 3.72\AA , respectively). Then in the Alpha variant, such residues are T333, R346, K378, K386, R408, and N450. In Beta and Gamma variants, such residues are T333, R346, K378, K386, and R408. In BA.1 such residues are T333, N334, E340, R346, N360, D364, Y369, K378, K386, F392, R408, K424, N450, K462, and H519. Also most large C_α structural changes occur at the coil regions of the RBD. For the BA.2 variant, the half of maximal distance is 7.16\AA and we have recorded such residues that have at least one atom whose distance to the corresponding atom in the wild-type RBD is more than 7.16\AA .

To quantify the total structural differences between the wild type and mutants, we calculate the sum of squares of distances between corresponding C_α atoms. The results of Alpha, Beta, Gamma, BA.1, and BA.2 are 69\AA^2 , 70\AA^2 , 67\AA^2 , 93\AA^2 , and 255\AA^2 , respectively as shown in Figure 5.4. The large values for BA.1 and BA.2 are consistent with fact that BA.1 and BA.2 are strongly antibody disruptive [29, 31]. The large structural changes induced by BA.2 mutations create significant mismatch between antibodies and antigens, making BA.2 one of the most antibody resistant variants [31]. Arguably, the amount of mutation-induced structural changes in RBD-ACE2 complexes also strongly correlates with viral infectivity changes.

Now we turn to the topological characterization of the mutation-induced conformational changes. To this end, we employ persistent Laplacians (PL) and persistent sheaf Laplacians (PSL) to examine the local RBD structural changes induced by the mutation N501Y (a common mutation that exists in Alpha, Beta, Gamma, BA.1, and BA.2). For the wild type and mutants, the residue 501 mutation site is defined as the set of neighborhood heavy

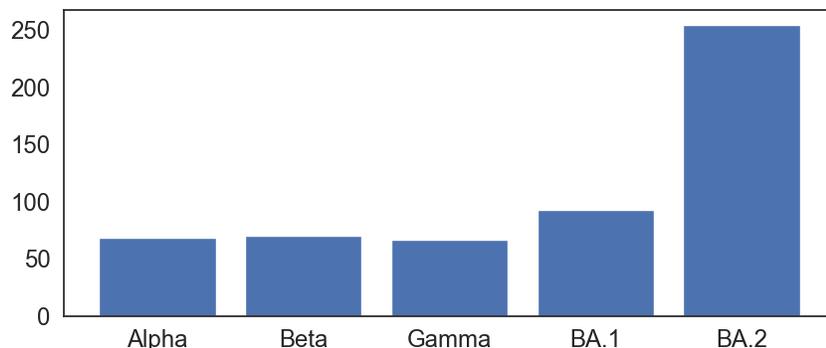


Figure 5.4 The total structural changes of RBD between the wild type and mutants in RBD-ACE2 complex. Given an alignment of a mutant RBD to the wild type RBD, the total structural changes is defined to be the sum of squares of distances between corresponding C_α atoms in RBD.

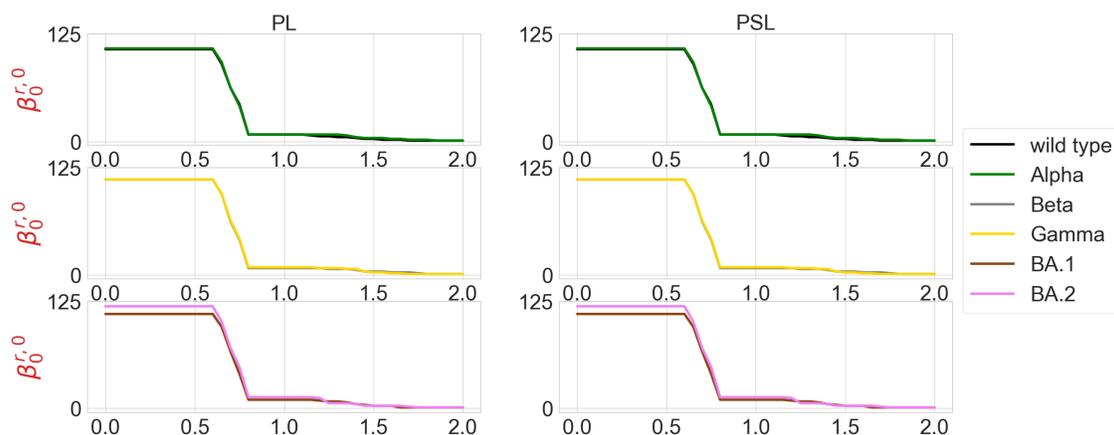


Figure 5.5 Illustration of persistent (sheaf) Betti numbers of element nonspecific persistent Laplacian (PL) and persistent sheaf Laplacian (PSL) of the residue 501 mutation site at different filtration values, i.e., radii (unit: Å). The wild type (PDB ID: 6M0J) and Alpha (PDB ID: 8DLK) are given in the first row. The Beta (PDB ID: 8DLN) and Gamma (PDB ID: 8DLQ) are given in the second row. BA.1 (PDB ID: 7T9L) and BA.2 (PDB ID: 7XB0) are given in the third row.

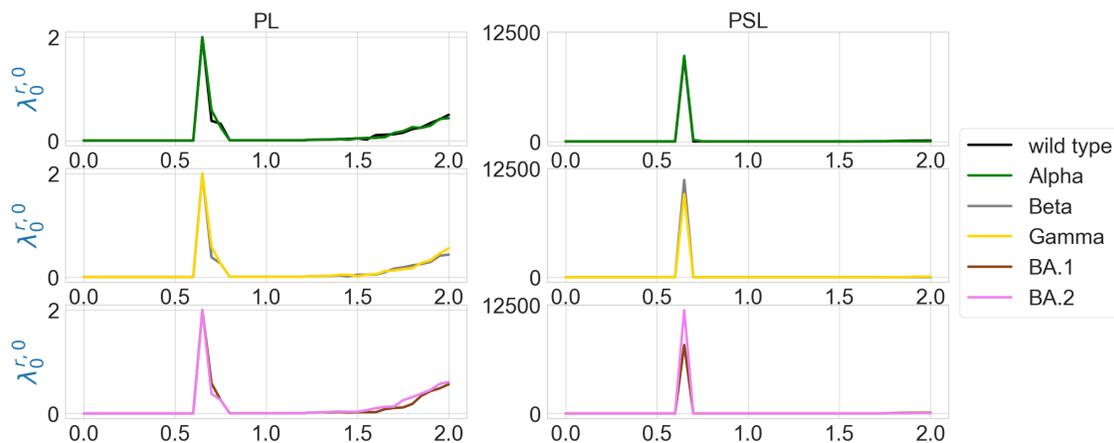


Figure 5.6 Illustration of the first nonzero eigenvalues of element nonspecific persistent Laplacian (PL) and persistent sheaf Laplacian (PSL) of the residue 501 mutation site at different filtration values, i.e., radii (unit: Å). The wild type (PDB ID: 6M0J) and Alpha (PDB ID: 8DLK) are given in the first row. The Beta (PDB ID: 8DLN) and Gamma (PDB ID: 8DLQ) are given in the second row. BA.1 (PDB ID: 7T9L) and BA.2 (PDB ID: 7XB0) are given in the third row.

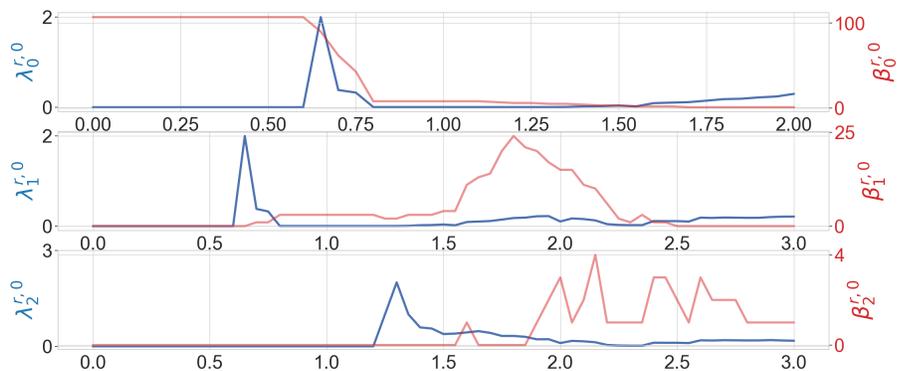


Figure 5.7 Illustration of persistent Betti numbers (red line) and the first nonzero eigenvalues (blue line) of element nonspecific persistent Laplacians of the wild type N501 mutation site at different filtration values, i.e., radii (unit: Å). Alpha filtration is used. The graphs from top to bottom represent the results of dimension-0, dimension-1, and dimension-2 Laplacians.

atoms (C, N, and O) in RBD such that the distance of any atom in the set to the residue 501 C $_{\alpha}$ is smaller than 10Å. We calculate persistent Laplacians and persistent sheaf Laplacians for mutation sites of the wild type and variants and compare the persistent (sheaf) Betti numbers and the smallest nonzero eigenvalues of spectra at different filtration values. Persistent Laplacians and persistent sheaf Laplacians can be calculated as either element non-specifically or element specifically (i.e., considering carbon, nitrogen, and oxygen atoms separately). We first employ the element non-specific approach and compare the results of the wild type and variants. The results of persistent Laplacians and persistent sheaf Laplacians are shown in Figures 5.5 and 5.6. The x axis represents the filtration values of Rips filtration, such that at a filtration value r the Rips complex is constructed by considering balls of radius r . The sudden changes of persistent (sheaf) Betti numbers and the first nonzero eigenvalues near $r = 0.65\text{\AA}$ reflect the fact that most neighboring atoms are about 1.3\AA away from each other. In Figure 5.5, The number of atoms is reflected in the initial 0-th Betti numbers. The 0-th Betti number dramatically decreases around 0.65\AA because covalent bond distances are about 1.5\AA . The 0-th Betti number decreases further from 1.2\AA to 1.7\AA due to other many non-covalent bonds.

In Figure 5.6, the results of the wild type and mutants almost coincide, except that the first nonzero eigenvalues of persistent sheaf Laplacians of BA.1 and BA.2 near $r = 0.65\text{\AA}$ have very different values. The results of persistent Laplacians are quite different from those of persistent sheaf Laplacians at large filtration values. The significant changes around $r = 0.65\text{\AA}$ are due to the topological changes.

We are also interested in understanding whether higher dimensional persistent Laplacians can offer an additional characterization of biomolecules. Figure 5.7 presents the higher dimensional persistent Laplacian analysis of the wide type RBD near the N501 residue. Obviously, higher dimensional persistent Laplacian offers significant structural information about the distributions of circles and cavities of the macromolecule. Most dimension-1 circles occur in the range of $1.5\text{-}2.4\text{\AA}$, whereas most 2-dimensional cavities locate around $1.8\text{-}2.8\text{\AA}$.

2-dimensional cavities are short-lived in the filtration, indicating the lack of multiple large cavities in the structure (at most one large cavity in the structure). This distribution can be used to understand interaction forces. For example, the length of hydrogen bonds ranges from 2-3.6Å (corresponding to 1-1.8 Å in the filtration radii). This information is valuable for the design of machine learning representations, including the selection of the set of filtration intervals. We also note that the peak of $\lambda_2^{r,0}$ is at the left of $\beta_2^{r,0}$. It's possible that when r is in the range of 1.2Å-1.5Å, many 2-simplices are born but no 2-cycles are formed yet.

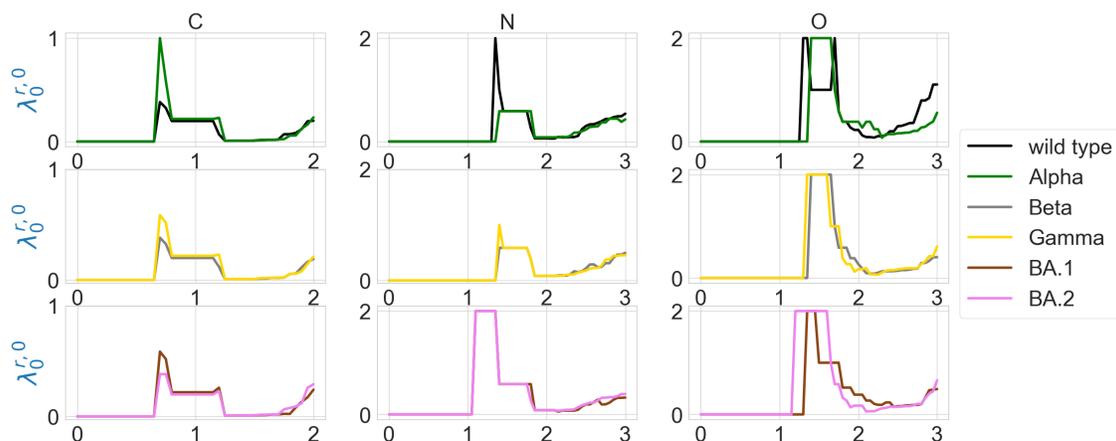


Figure 5.8 Illustration of the first nonzero eigenvalues of element-specific persistent Laplacian of the residue 501 mutation site at different filtration values, i.e., radii (unit: Å). The wild type (PDB ID: 6M0J) and Alpha (PDB ID: 8DLK) are given in the first row. The Beta (PDB ID: 8DLN) and Gamma (PDB ID: 8DLQ) are given in the second row. BA.1 (PDB ID: 7T9L) and BA.2 (PDB ID: 7XB0) are given the third row.

The element-specific results of the residue 501 mutation site of the wild type, and variants Alpha, Beta, Gamma, BA.1, and BA.2 are shown in Figures 5.8 and 5.9, as well as in Figures D.1 and D.2 in the Appendix. We observe that the difference between the first nonzero eigenvalues is much more obvious. For instance, in Figure 5.8 there is a higher spike near 0.7Å in the graph of Alpha carbon atoms, and two spikes near 1.3Å and 1.7Å disappear in the graph of the Alpha variant's oxygen atoms. In Figure 5.8, all results of carbon atoms have similar shapes, implying a relatively stable RBD carbon atom structure. In the results of nitrogen atoms, we notice that the results of Alpha, Beta, and Gamma variants resemble each other, and the same can be said of the results of BA.1 and BA.2

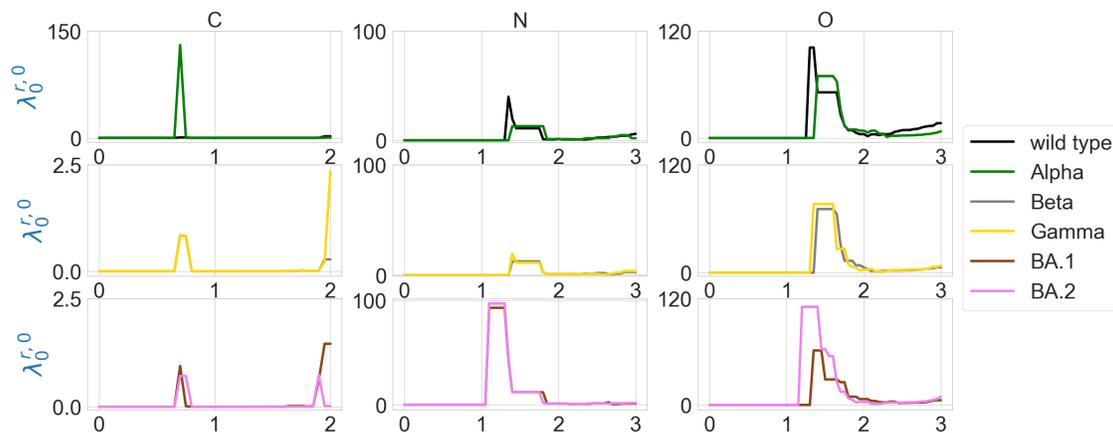


Figure 5.9 Illustration of the first nonzero eigenvalues of element-specific persistent sheaf Laplacian of the residue 501 mutation site at different filtration values, i.e., radii (unit: Å). The wild type (PDB ID: 6M0J) and Alpha (PDB ID: 8DLK) are given in the first row. The Beta (PDB ID: 8DLN) and Gamma (PDB ID: 8DLQ) are given in the second row. BA.1 (PDB ID: 7T9L) and BA.2 (PDB ID: 7XB0) are given the third row.

variants. In the results of oxygen atoms, the results of Alpha, Beta, and Gamma still resemble each other, but the results of BA.1 and BA.2 are quite different. The results of the wild type are unique in the sense that it has one or two spikes near 1.3Å or 1.7Å . These results indicate that element-specific persistent Laplacians and element-specific persistent sheaf Laplacians are better approaches in characterizing SARS-CoV-2 variants than element-nonspecific approaches. We know that nitrogen and oxygen atoms are sparser in a protein, so if we use element nonspecific approach, nitrogen atoms and oxygen atoms will first form edges with neighboring carbon atoms, and we are not able to infer distances between nitrogen atoms or oxygen atoms. This explains why element specific approach outperforms element nonspecific approach.

5.0.2 PTL analysis of RBD structural changes induced by its binding to ACE2

We investigate how binding to ACE2 changes the spike protein RBD structure from the closed state to the open state for the wild type, Alpha, Beta, BA.1, and BA.2 variants. The PDB IDs of the spike protein of wild type, Alpha, Beta, BA.1 and BA.2 used in this section are 7DF3 [148], 7LWS [51], 7LYM [51], 7TF8 [50] and 7XIX [21]. The analysis of the Gamma variant is eliminated due to the lack of experimental structure. We first align each

of the three RBDs in the closed-state spike protein to the RBD in the RBD-ACE2 complex. The maximal distances between corresponding atoms in the RBM of the three alignments of BA.1 are 8.76Å, 13.49Å, and 9.44Å, which are larger than those of alignments of the wild type and other mutants. For each alignment, we record the RBM residues that have at least one atom whose distance to the corresponding atom is larger than 5.28Å, i.e., half of the mean maximal distances between corresponding atoms in RBM of the three alignments of BA.1. In wild-type RBD, such residues are K444 and K458. In Alpha there are no such residues; In Beta, chains A and B have K458; chain C has T478 and P479. In BA.1, each chain has different such residues: chain A has K440, Y453, K458, K478, and F486; chain B has K440, Y453, R457, K458, R466, Y473, Q474, K478, F486, F490, R493; and chain C has K440, Y453, Y473, K478, F486. In BA.2 such residues are E465, K478, and G482.

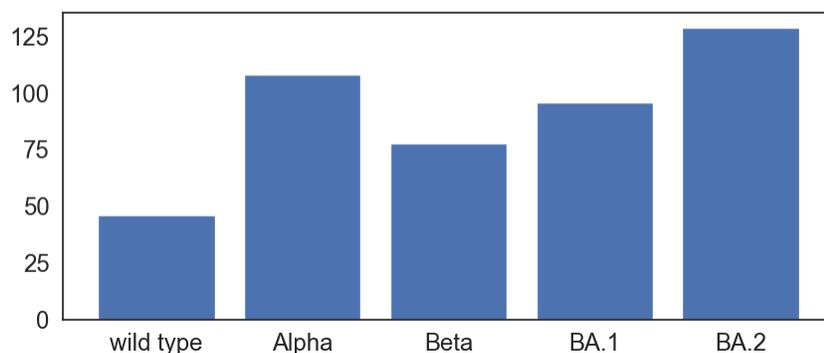


Figure 5.10 The total structural changes of the RBM between the closed state RBD and the open state RBD induced by ACE2 binding. Here the total structural changes are defined to be the sum of squares of distances between C_{α} atoms in the RBM.

We also calculate the total structural changes of the RBM between the closed state RBD and the open state RBD induced by its binding to the human ACE2. Here, the total structural changes are defined to be the sum of squares of distances between C_{α} atoms in the RBM. Since spike protein is a trimer, we calculate the total structural changes for each chain and report the average (see Figure 5.10). It turns out that the average total structural changes induced by binding to ACE2 do not increase too much with respect to the number of RBD mutations.

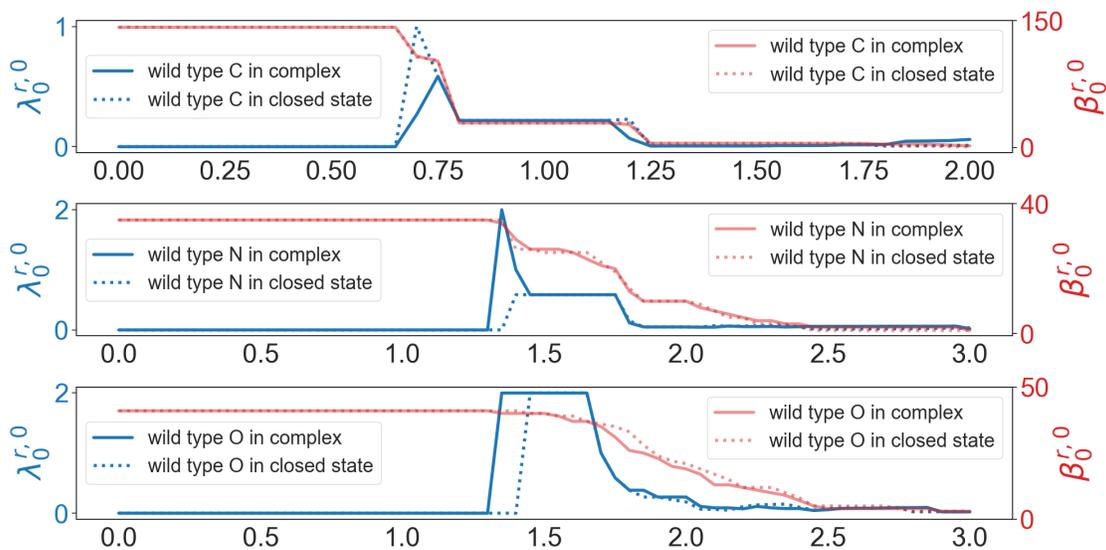


Figure 5.11 Illustration of persistent Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent Laplacian of the RBD binding site of the wild type RBD-ACE2 complex (PDB ID: 6M0J) and closed state spike protein (PDB ID: 7DF3, Chain ID: A) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of carbon atoms, nitrogen atoms, and oxygen atoms, respectively.

Now, we calculate persistent Laplacians and persistent sheaf Laplacians for the RBD binding site in the closed state spike protein and the RBD-ACE2 complex. For the wild type and mutants, we define the RBD binding site as the set of RBD residues whose C_{α} s are within 10\AA from the C_{α} s of ACE2 residues. We choose 10\AA as the cutoff distance, because if we used 11\AA then the RBD binding site would include non-RBM residues. Spike protein as a trimer has three chains. In the results of alignments, the recorded residues of the wild type, Alpha, and BA.2 are the same for the three chains. Therefore, for the wild type, Alpha and BA.2 we only use chain A, and for Beta and BA.1, we use all three chains. The study was carried out in an element-specific manner for carbon atoms, nitrogen atoms, and oxygen atoms. The results of the wild type are shown in Figure 5.11. We noted that persistent Betti numbers cannot distinguish two structures. However, the first nonzero eigenvalues of the persistent Laplacians capture the difference, demonstrating the advantage of persistent Laplacians over persistent homology in protein structure analysis.

Additional analysis is presented in Figures D.3, D.4, D.5, D.6, D.7, D.8, D.9, D.10, D.11, and D.12 in the Appendix. In Figure D.3, the results of the wild type, Alpha, Beta, BA.1,

and BA.2 RBD binding sites are quite similar except that the wild type RBD binding site has relatively lower first nonzero eigenvalues near $r = 0.7\text{\AA}$. A peak appears or disappears in the graph of the nitrogen atoms, whereas for BA.1 and BA.2, the results of the nitrogen atoms resemble each other, sometimes even coincide.

The results of persistent Laplacians and persistent sheaf Laplacians are similar in this work. However, this similarity is due to the specific implementation of persistent sheaf Laplacians. In general, persistent sheaf Laplacians enable the embedding of non-geometric chemical and physical information of biomolecules in topological and spectral representations.

TDA assisted supervised learning

The workflow of TDA-assisted supervised learning is shown as below. Suppose our dataset includes molecules (e.g., PDB files) and corresponding labels (a numerical number representing a biochemical property), one can use TDA methods to obtain a representation (a feature vector) of each molecule and input all representation and labels to train a machine learning model. Persistent homology was first employed in TDA-assisted supervised learning. For example, [20] used persistent homology to predict protein-ligand binding affinity and state-of-the-art results were achieved.

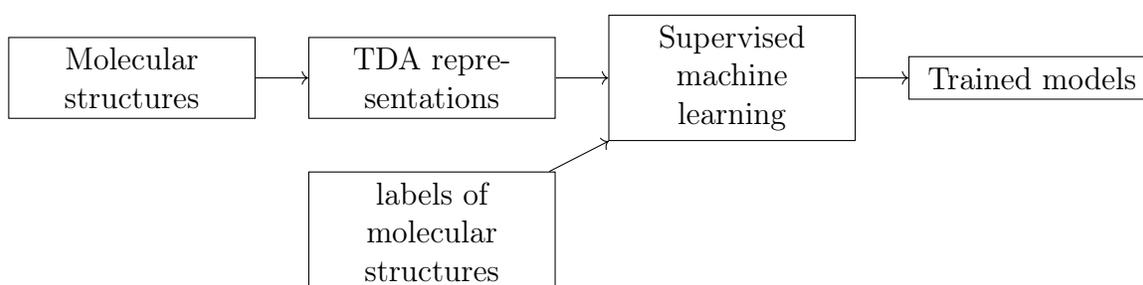


Figure 5.12 The general workflow of TDA-assisted supervised learning.

TDA-assisted supervised learning has also been used to predict the impact of mutation on protein-protein interaction [18, 28]. Suppose a wild type protein-protein complex structure and a dataset including single amino acid mutations (chain ID, residue ID, mutant residue, etc.) and corresponding binding free energy (BFE) changes caused by mutations

are available¹. Mutant structures can be computationally generated based on experimentally determined structures of the wild-type antibody-antigen complexes and mutation information. Since we are interested in the prediction of binding free energy changes caused by mutation, we input both the TDA representations (i.e., feature vectors) of wild type structure and mutant structures to a machine learning model, so that the model will learn the structural change induced by a mutation². Laplacian representations have been employed in TDA-assisted supervised learning [28]. Their performance on benchmark datasets are better than previous models. The only difference is that TopLapGBT uses gradient boosting trees whereas TopLapNet employs neural networks.

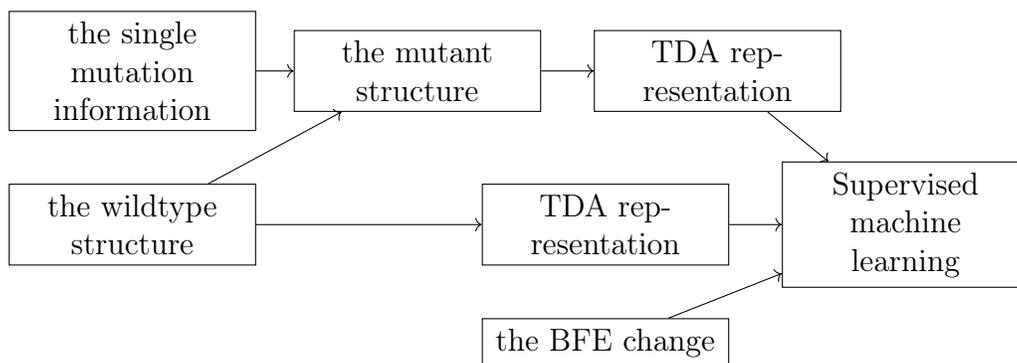


Figure 5.13 The workflow to study the impact of mutation on a wildtype structure.

To do prediction with a trained TopLapGBT or TopLapNet, one has to transform the wildtype structure and the mutant structure to TDA representations and input them to the trained model.

When we apply persistent homology and persistent Laplacians to the study of protein-protein interactions, we always extract the atoms within a certain cutoff distance r of the binding site³ and construct a distance matrix such that if two atoms are in the same protein then the distance between them is an extremely large constant number (to ensure that

¹For instance, the AB-Bind S645 dataset [123] includes 645 mutants with experimentally determined BFE changes across 29 antibody-antigen complexes.

²We also input the difference of the feature vectors of the wild type and the mutant. This trick enhances the performance.

³Suppose an atom a is in protein A . If the distance of atom a to the other protein in the complex is not larger than r , then we say atom a is within distance r of the binding site.

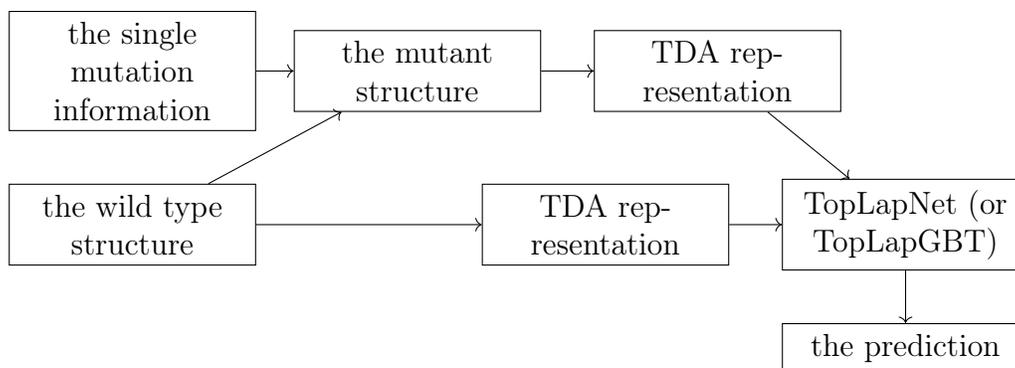


Figure 5.14 How to predict using TopLapGBT or TopLapNet.

atomic interaction within a single protein is ignored). To further characterize the interaction between atoms of certain elements E_1 and E_2 , we can consider the point cloud formed by the atoms of an element E_1 of protein A within r of the binding site, and the atoms of element E_2 of protein B within r of the binding site. After the calculation of persistent homology and persistent Laplacians, the next step is to transform the barcodes of persistent homology or spectra of persistent Laplacians into vector representations of fixed lengths. For barcodes, there are at least two ways: either we divide the interval $[0, r]$ into bins of even length and count the occurrence of bars, birth values, and death values in each bin, or we simply compute statistics such as sum, maximum, minimum, mean, and standard deviation for bar lengths, birth values, and death values. The former method is often applied to 0-dimensional barcodes and the latter to 1-dimensional and 2-dimensional barcodes. For the spectrum of a persistent Laplacian, we separate zero eigenvalues (harmonic spectra) and nonzero eigenvalues (non-harmonic spectra). We use the number of zero eigenvalues, the sum, the minimum, the maximum, the mean, the standard deviation, the variance, and the sum of squares of nonzero eigenvalues.

Impacts of computationally generated structures on PTL-assisted machine learning predictions

We mentioned earlier that a wild type structure is needed in TopLapGBT and TopLapNet. It is natural to ask if structural perturbation to the wild type structure will impact the two models. To explore this, we use a SARS-CoV-2 BA.2 RBD deep mutational scanning

dataset which involves the systematical mutations of each residue on the BA.2 RBD to 19 other residues and records corresponding binding affinity changes [125]. The deep mutational scanning covers the RBD residues from 333 to 527. In order to apply machine learning models, such as TopLapGBT and TopLapNet [28], to this dataset, BA.2 RBD mutants need to be computationally generated based on a BA.2 RBD structure and the choice of the BA.2 RBD structure can affect the performance of machine learning models. We can choose either an experimentally determined BA.2 RBD-ACE2 complex structure or a BA.2 RBD-ACE2 complex structure computationally generated based on an experimentally determined BA.1 RBD-ACE2 complex structure. When the given BA.2 RBD structure is experimentally determined (PDB ID: 7XB0), the resulting models are referred to as ExpTopLapGBT (experimental TopLapGBT) and ExpTopLapNet (experimental TopLapNet). When the BA.2 RBD structure is computationally generated from BA.1 RBD (PDB ID: 7T9L) by Jackal [147], the resulting model is referred to as ComTopLapGBT (computational TopLapGBT) or ComTopLapNet (computational TopLapGBT). The distances of corresponding atoms between the experimentally determined RBD (PDB ID: 7XB0) and the RBD generated computationally from BA.1 RBD (PDB ID: 7T9L) is shown in Figure 5.15.

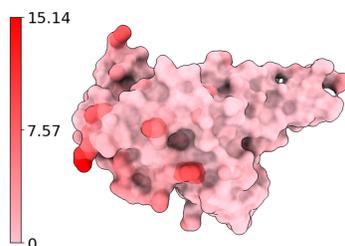


Figure 5.15 Atoms of BA.2 RBD (PDB ID: 7XB0) are colored by their distances to corresponding atoms in the computationally generated structure. We record the residues that have at least one atom whose distance to the corresponding atom in wild type RBD is more than 7.57\AA . Such residues are 370, 375, 378, 386, 387, and 519.

We compare the results of ExpTopLapGBT and ComTopLapGBT, on the predictions of the RBD deep mutational scanning dataset. We split the dataset into 10 folds, and for each fold, we use the other 9 folds as the training set to build a machine learning model, which is used to predict ACE2-binding affinity changes for the fold. Therefore, for a given

10-fold splitting we get the ExpTopLapGBT and ComTopLapGBT predictions of RBD-ACE2 binding affinity changes for the deep mutational scanning dataset. We denote by $R_p(Exp, True)$ the Pearson correlation coefficient between ExpTopLapGBT predicted binding affinity changes and experimental binding affinity changes. Similarly, $R_p(Com, True)$ (or $R_p(Exp, Com)$) is the Pearson correlation coefficient between ComTopLapGBT predicted binding affinity changes and experimental binding affinity changes (or ExpTopLapGBT predicted binding affinity changes). We also do the same analysis for TopLapNet

Method	$R_p(Exp, True)$	$R_p(Com, True)$	$R_p(Exp, Com)$
TopLapGBT	0.901	0.898	0.990
TopLapNet	0.879	0.849	0.925

Table 5.1 $R_p(Exp, True)$ is the correlation coefficient between predictions of ExpTopLapGBT (or ExpTopLapNet) and true affinity changes. Here, $R_p(Com, True)$ is the correlation coefficient between predictions of ComTopLapGBT (or ComTopLapNet) and true affinity changes. $R_p(Exp, Com)$ is Pearson the correlation coefficient between the predictions of ExpTopLapGBT and ComTopLapGBT (or between ExpTopLapNet and ComTopLapNet). A random state affects the 10-fold splitting and the training of GBT and neural networks.

The results of TopLapGBT and TopLapNet are shown in Table 5.1. Generally, the performance of models using experimentally determined structures is better than that of models using the computationally generated structure. This is not surprising since the computationally generated structure is an approximation of the experimental structure. The performance of ExpTopLapGBT and ComTopLapGBT are extremely close, whereas the performance of ComTopLapNet differs very much from that of ExpTopLapNet. We also see that ExpTopLapGBT outperforms ExpTopLapNet.

In this study, we use scikit-learn to build a gradient boosting tree whose parameters are `n_estimators=20000`, `learning_rate = 0.005`, `max_features = 'sqrt'`, `max_depth = 9`,

`min_samples_split = 3`, `subsample = 0.4`, and `n_iter_no_change=500`. Additionally, we use PyTorch to build a neural network with 7 hidden layers where each layer has 8000 neurons.

CHAPTER 6

THESIS CONTRIBUTION AND FUTURE WORK

The main contributions of this dissertation are listed as follows:

- In section 2, we propose to calculate of spectra of persistent Laplacians using homotopy continuation methods.
- In chapter 3, we extend persistent Laplacians to the setting of cellular sheaves and discuss how persistent sheaf Laplacians can be applied to analyze biomolecules.
- In chapter 4, we review the recent generalizations of persistent Laplacians.
- In chapter 5, we perform analysis of RBD structural changes induced by mutations and stability of persistent Laplacian assisted machine learning models.

The contents of this dissertation are mostly adopted from the following publications and preprints:

- X. Wei and G.-W. Wei. Homotopy continuation for the spectra of persistent Laplacians. *Foundations of Data Science*, 3(4):677, 2021.
- X. Wei and G.-W. Wei. Persistent sheaf Laplacians. *Foundations of Data Science*, 2024.
- X. Wei and G.-W. Wei. Persistent Topological Laplacians—a Survey. arXiv preprint arXiv:2312.07563, 2023.
- X. Wei, J. Chen, and G.-W. Wei. Persistent topological Laplacian analysis of SARS-CoV-2 variants. *Journal of computational biophysics and chemistry*, 22(5):569, 2023.

Many future directions are available, including:

- It is challenging to understand the relationship between the geometry/topology of the data and PTLs. The understanding of this relationship is crucial for the application of PTLs to real world problems.

- To a certain extent, the success of persistent homology can be attributed to its integration with machine learning, particularly with the introduction of topological deep learning [19]. The featurization of Laplacians typically requires domain knowledge and experience. Since self learning representations of persistent diagrams have been proposed [70], we wonder if self learning representations of (persistent) Laplacians are possible. It is also interesting to featurize the eigenvectors of Laplacians.
- Despite efforts in software development [96, 135], the computation of PTLs remains slow, particularly for problems involving large datasets. Since the primary value of TDA lies in its ability to analyze data, one of the most pressing needs will be the development of efficient and robust PTL software packages. The development of finite field PTLs will be also valuable.
- The invention of cellular sheaves for different scenarios is crucial for successful applications of persistent sheaf Laplacians.
- One can also extend PTLs to settings such as the Hochschild complex [49], quantum homology [11], multiparameter persistent homology [68], and interaction homotopy and interaction homology [87].
- As discussed in [126], persistent Dirac operators can be formulated for flag complexes, digraphs, hyperdigraphs, etc. It is possible that a persistent sheaf Dirac operator can be devised to distinguish certain point clouds.
- It will be interesting to generalize various PTLs on point clouds to the manifold and knot-type data settings.
- Persistent Mayer homology and persistent Mayer Laplacians have been introduced on N -chain complexes [121]. These formulations encompass persistent homology and persistent Laplacians as special cases. The potential for future developments on these subjects is widely open.

BIBLIOGRAPHY

- [1] E. L. Allgower, D. J. Bates, A. J. Sommese, and C. W. Wampler. Solution of polynomial systems derived from differential equations. *Computing*, 76:1–10, 2006.
- [2] B. Ameneiro, V. Maroulas, and G. Siopsis. Quantum persistent homology. *Journal of Applied and Computational Topology*, pages 1–20, 2024.
- [3] D. N. Arnold, G. David, M. Filoche, D. Jerison, and S. Mayboroda. Computing spectra without solving eigenvalue problems. *SIAM Journal on Scientific Computing*, 41(1):B69–B92, 2019.
- [4] G. Ausiello and L. Laura. Directed hypergraphs: Introduction and fundamental algorithms—a survey. *Theoretical Computer Science*, 658:293–306, 2017.
- [5] F. Baccini, F. Geraci, and G. Bianconi. Weighted simplicial complexes and their representation power of higher-order network data and topology. *Physical Review E*, 106(3):034319, 2022.
- [6] D. J. Bates, I. A. Fotiou, and P. Rostalski. A numerical algebraic geometry approach to nonlinear constrained optimal control. In *2007 46th IEEE Conference on Decision and Control*, pages 6256–6261. IEEE, 2007.
- [7] D. J. Bates, J. D. Hauenstein, A. J. Sommese, and C. W. Wampler. Bertini: Software for numerical algebraic geometry. Available at bertini.nd.edu with permanent doi: dx.doi.org/10.7274/R0H41PB5.
- [8] D. J. Bates, A. J. Sommese, J. D. Hauenstein, and C. W. Wampler. *Numerically Solving Polynomial Systems with Bertini*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013.
- [9] C. Battiloro, S. Sardellitti, S. Barbarossa, and P. Di Lorenzo. Topological signal processing over weighted simplicial complexes. *arXiv preprint arXiv:2302.08561*, 2023.
- [10] G. Bianconi. The topological dirac equation of networks and simplicial complexes. *Journal of Physics: Complexity*, 2(3):035022, 2021.
- [11] P. Biran and O. Cornea. A lagrangian quantum homology. *New perspectives and challenges in symplectic field theory*, 49:1–44, 2009.
- [12] P. Breiding and S. Timme. HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia. In *International Congress on Mathematical Software*, pages 458–465. Springer, 2018.
- [13] S. Bressan, J. Li, S. Ren, and J. Wu. The embedded homology of hypergraphs and applications. *Asian Journal of Mathematics*, 23(3):479–500, 2019.
- [14] A. Bura, Q. He, and C. Reidys. Weighted homology of bi-structures over certain discrete valuation rings. *Mathematics*, 9(7):744, 2021.

- [15] A. C. Bura, N. S. Dutta, T. J. Li, and C. M. Reidys. A computational framework for weighted simplicial homology. *arXiv preprint arXiv:2206.04612*, 2022.
- [16] Z. Cang, L. Mu, and G.-W. Wei. Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening. *PLoS computational biology*, 14(1):e1005929, 2018.
- [17] Z. Cang, L. Mu, K. Wu, K. Opron, K. Xia, and G.-W. Wei. A topological approach for protein classification. *Computational and Mathematical Biophysics*, 3(1), 2015.
- [18] Z. Cang and G.-W. Wei. Analysis and prediction of protein folding energy changes upon mutation by element specific persistent homology. *Bioinformatics*, 33(22):3549–3557, 2017.
- [19] Z. Cang and G.-W. Wei. Topologynet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions. *PLoS computational biology*, 13(7):e1005690, 2017.
- [20] Z. Cang and G.-W. Wei. 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):e2914, 2018.
- [21] Y. Cao, A. Yisimayi, F. Jian, W. Song, T. Xiao, L. Wang, S. Du, J. Wang, Q. Li, X. Chen, et al. BA. 2.12. 1, BA. 4 and BA. 5 escape antibodies elicited by Omicron infection. *Nature*, 608(7923):593–602, 2022.
- [22] L. Caputi and H. Riihimäki. Hochschild homology, and a persistent approach via connectivity digraphs. *Journal of Applied and Computational Topology*, pages 1–50, 2023.
- [23] D. Carlson, E. Haynsworth, and T. Markham. A generalization of the Schur complement by means of the Moore–Penrose inverse. *SIAM Journal on Applied Mathematics*, 26(1):169–175, 1974.
- [24] D. Chen, J. Liu, J. Wu, and G.-W. Wei. Persistent hyperdigraph homology and persistent hyperdigraph Laplacians. *Foundations of Data Science*, 5(4):558–588, 2023.
- [25] D. Chen, J. Liu, J. Wu, G.-W. Wei, F. Pan, and S.-T. Yau. Path topology in molecular and materials sciences. *The Journal of Physical Chemistry Letters*, 14(4):954–964, 2023.
- [26] J. Chen, K. Gao, R. Wang, and G.-W. Wei. Prediction and mitigation of mutation threats to COVID-19 vaccines and antibody therapies. *Chemical Science*, 12(20):6929–6948, 2021.
- [27] J. Chen, K. Gao, R. Wang, and G.-W. Wei. Revealing the threat of emerging SARS-CoV-2 mutations to antibody therapies. *Journal of Molecular Biology*, 433(7744), 2021.

- [28] J. Chen, Y. Qiu, R. Wang, and G.-W. Wei. Persistent Laplacian projected Omicron BA. 4 and BA. 5 to become new dominating variants. *Computers in Biology and Medicine*, 151:106262, 2022.
- [29] J. Chen, R. Wang, N. B. Gilby, and G.-W. Wei. Omicron variant (B. 1.1. 529): infectivity, vaccine breakthrough, and antibody resistance. *Journal of chemical information and modeling*, 62(2):412–422, 2022.
- [30] J. Chen, R. Wang, M. Wang, and G.-W. Wei. Mutations strengthened SARS-CoV-2 infectivity. *Journal of molecular biology*, 432(19):5212–5226, 2020.
- [31] J. Chen and G.-W. Wei. 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.
- [32] J. Chen, R. Zhao, Y. Tong, and G.-W. Wei. Evolutionary de rham-Hodge method. *Discrete and continuous dynamical systems. Series B*, 26(7):3785, 2021.
- [33] T. Chen, T.-L. Lee, and T.-Y. Li. Hom4ps-3: A parallel numerical solver for systems of polynomial equations based on polyhedral homotopy continuation methods. In H. Hong and C. Yap, editors, *Mathematical Software – ICMS 2014*, pages 183–190, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.
- [34] S. Chowdhury, S. Huntsman, and M. Yutin. Path homologies of motifs and temporal network representations. *Applied Network Science*, 7(1):4, 2022.
- [35] S. Chowdhury and F. Mémoli. Persistent path homology of directed networks. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1152–1169. SIAM, 2018.
- [36] F. R. Chung. The Laplacian of a hypergraph. In *Expanding graphs*, pages 21–36, 1992.
- [37] F. R. Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.
- [38] O. T. Courtney and G. Bianconi. Weighted growing simplicial complexes. *Physical Review E*, 95(6):062301, 2017.
- [39] J. Curry. *Sheaves, cosheaves and applications*. PhD thesis, University of Pennsylvania, 2014.
- [40] Y. Dabaghian, F. Mémoli, L. Frank, and G. Carlsson. A topological paradigm for hippocampal spatial map formation using persistent homology. 2012.
- [41] R. J. M. Dawson. Homology of weighted simplicial complexes. *Cahiers de Topologie et Géométrie Différentielle Catégoriques*, 31(3):229–243, 1990.
- [42] M. Desbrun, E. Kanso, and Y. Tong. Discrete differential forms for computational modeling. In *ACM SIGGRAPH 2006 Courses*, pages 39–54. 2006.

- [43] B. Eckmann. Harmonische funktionen und randwertaufgaben in einem komplex. *Commentarii Mathematici Helvetici*, 17(1):240–255, 1944.
- [44] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2010.
- [45] E. Emtander. Betti numbers of hypergraphs. *Communications in algebra*, 37(5):1545–1571, 2009.
- [46] E. Estrada. Path laplacian matrices: introduction and application to the analysis of consensus in networks. *Linear algebra and its applications*, 436(9):3373–3391, 2012.
- [47] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, 23(2):298–305, 1973.
- [48] M. Gameiro, Y. Hiraoka, S. Izumi, M. Kramar, K. Mischaikow, and V. Nanda. A topological measurement of protein compressibility. *Japan Journal of Industrial and Applied Mathematics*, 32(1):1–17, 2015.
- [49] M. Gerstenhaber and A. Voronov. Higher operations on the hochschild complex. *Funct. Anal. Appl*, 29(1):3, 1995.
- [50] S. M.-C. Gobeil, R. Henderson, V. Stalls, K. Janowska, X. Huang, A. May, M. Speakman, E. Beaudoin, K. Manne, D. Li, et al. Structural diversity of the SARS-CoV-2 Omicron spike. *Molecular cell*, 2022.
- [51] S. M.-C. Gobeil, K. Janowska, S. McDowell, K. Mansouri, R. Parks, V. Stalls, M. F. Kopp, K. Manne, D. Li, K. Wiehe, et al. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science*, 373(6555):eabi6226, 2021.
- [52] T. E. Goldberg. Combinatorial Laplacians of simplicial complexes. *Senior Thesis, Bard College*, 2002.
- [53] A. Gomes and D. Miranda. Path cohomology of locally finite digraphs, Hodge’s theorem and the p -lazy random walk. *arXiv preprint arXiv:1906.04781*, 2019.
- [54] C. González, G. M. Langdon, M. Bruix, A. Gálvez, E. Valdivia, M. Maqueda, and M. Rico. Bacteriocin as-48, a microbial cyclic polypeptide structurally and functionally related to mammalian nk-lysin. *Proceedings of the National Academy of Sciences*, 97(21):11221–11226, 2000.
- [55] A. Grigor’yan. Advances in path homology theory of digraphs. 2022.
- [56] A. Grigor’yan, R. Jimenez, Y. Muranov, and S.-T. Yau. Homology of path complexes and hypergraphs. *Topology and its Applications*, 267:106877, 2019.
- [57] A. Grigor’yan, Y. Lin, Y. Muranov, and S.-T. Yau. Homologies of path complexes and digraphs. *arXiv preprint arXiv:1207.2834*, 2012.

- [58] A. Grigor'yan, Y. Lin, Y. Muranov, and S.-T. Yau. Homotopy theory for digraphs. *Pure and Applied Mathematics Quarterly*, 10(4):619–674, 2014.
- [59] A. Grigor'yan, Y. Lin, Y. Muranov, and S.-T. Yau. Cohomology of digraphs and (undirected) graphs. *Asian Journal of Mathematics*, 19(5):887–932, 2015.
- [60] A. Grigor'yan, Y. Lin, Y. V. Muranov, and S.-T. Yau. Path complexes and their homologies. *Journal of Mathematical Sciences*, 248:564–599, 2020.
- [61] A. Grigor'yan, Y. Muranov, and S.-T. Yau. Homologies of digraphs and künneth formulas. *Communications in Analysis and Geometry*, 25(5):969–1018, 2017.
- [62] E. Gross, B. Davis, K. L. Ho, D. J. Bates, and H. A. Harrington. Numerical algebraic geometry for model selection and its application to the life sciences. *Journal of The Royal Society Interface*, 13(123):20160256, 2016.
- [63] A. B. Gülen, F. Mémoli, Z. Wan, and Y. Wang. A generalization of the persistent Laplacian to simplicial maps. In E. W. Chambers and J. Gudmundsson, editors, *39th International Symposium on Computational Geometry (SoCG 2023)*, volume 258 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 37:1–37:17, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [64] J. Hansen. *Laplacians of Cellular Sheaves: Theory and Applications*. PhD thesis, University of Pennsylvania, 2020.
- [65] J. Hansen and R. Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3(4):315–358, 2019.
- [66] W. Hao, J. D. Hauenstein, B. Hu, Y. Liu, A. J. Sommese, and Y.-T. Zhang. Multiple stable steady states of a reaction-diffusion model on zebrafish dorsal-ventral patterning. *Discrete & Continuous Dynamical Systems - S*, 4:1413, 2011.
- [67] W. Hao, B. Hu, and A. J. Sommese. Numerical algebraic geometry and differential equations. In *Future Vision and Trends on Shapes, Geometry and Algebra*, pages 39–53. Springer, 2014.
- [68] H. A. Harrington, N. Otter, H. Schenck, and U. Tillmann. Stratifying multiparameter persistent homology. *SIAM Journal on Applied Algebra and Geometry*, 3(3):439–471, 2019.
- [69] J. Hauenstein, J. Rodriguez, and B. Sturmfels. Maximum likelihood for matrices with rank constraints, 2013.
- [70] C. D. Hofer, R. Kwitt, and M. Niethammer. Learning representations of persistence barcodes. *J. Mach. Learn. Res.*, 20(126):1–45, 2019.
- [71] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181(2):271–280, 2020.

- [72] D. Horak and J. Jost. Spectra of combinatorial laplace operators on simplicial complexes. *Advances in Mathematics*, 244:303–336, 2013.
- [73] S. Hu and L. Qi. The Laplacian of a uniform hypergraph. *Journal of Combinatorial Optimization*, 29(2):331–366, 2015.
- [74] S. Huntsman. Path homology as a stronger analogue of cyclomatic complexity. *arXiv preprint arXiv:2003.00944*, 2020.
- [75] B. Jones and G.-W. Wei. Persistent flag Laplacians. *arXiv preprint arXiv:2203.12965*, 2023.
- [76] J. Jost and R. Mulas. Hypergraph Laplace operators for chemical reaction networks. *Advances in mathematics*, 351:870–896, 2019.
- [77] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M. J. Holst, J. A. McCammon, and N. A. Baker. Improvements to the apbs biomolecular solvation software suite. *Protein science : a publication of the Protein Society*, 27(1):112–128, Jan 2018. 28836357[pmid].
- [78] P. M. Kasson, A. Zomorodian, S. Park, N. Singhal, L. J. Guibas, and V. S. Pande. Persistent voids: a new structural metric for membrane fusion. *Bioinformatics*, 23(14):1753–1759, 2007.
- [79] S. Krishnagopal and G. Bianconi. Spectral detection of simplicial communities via Hodge laplacians. *Physical Review E*, 104(6):064303, 2021.
- [80] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807):215–220, 2020.
- [81] A. Leykin and F. Sottile. Galois groups of schubert problems via homotopy computation. *Mathematics of Computation*, 78(267):1749–1765, 2009.
- [82] L. Li, H. Liao, Y. Meng, W. Li, P. Han, K. Liu, Q. Wang, D. Li, Y. Zhang, L. Wang, et al. Structural basis of human ACE2 higher binding affinity to currently circulating Omicron SARS-CoV-2 sub-variants BA. 2 and BA. 1.1. *Cell*, 185(16):2952–2960, 2022.
- [83] T. J. Li and C. M. Reidys. On weighted simplicial homology. *arXiv preprint arXiv:2205.03435*, 2022.
- [84] A. Lieutier. Talk: Persistent harmonic forms, 2014.
- [85] L.-H. Lim. Hodge Laplacians on graphs. *Siam Review*, 62(3):685–715, 2020.
- [86] Y. Lin, S. Ren, C. Wang, and J. Wu. Weighted path homology of weighted digraphs and persistence. *arXiv preprint arXiv:1910.09891*, 2019.

- [87] J. Liu, D. Chen, and G.-W. Wei. Interaction homotopy and interaction homology. *arXiv preprint arXiv:2311.16322*, 2023.
- [88] J. Liu, J. Li, and J. Wu. The algebraic stability for persistent Laplacians. *arXiv preprint arXiv:2302.03902*, 2023.
- [89] R. Liu, X. Liu, and J. Wu. Persistent path-spectral (pps) based machine learning for protein–ligand binding affinity prediction. *Journal of Chemical Information and Modeling*, 2023.
- [90] X. Liu, H. Feng, J. Wu, and K. Xia. Persistent spectral hypergraph based machine learning (psh-ml) for protein-ligand binding affinity prediction. *Briefings in Bioinformatics*, 22(5):bbab127, 2021.
- [91] X. Liu, H. Feng, J. Wu, and K. Xia. 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.
- [92] D. Lütgehetmann, D. Govc, J. P. Smith, and R. Levi. Computing persistent homology of directed flag complexes. *Algorithms*, 13(1):19, 2020.
- [93] D. Mannar, J. W. Saville, Z. Sun, X. Zhu, M. M. Marti, S. S. Srivastava, A. M. Berezuk, S. Zhou, K. S. Tuttle, M. D. Sobolewski, et al. SARS-CoV-2 variants of concern: spike protein mutational analysis and epitope for broad neutralization. *Nature Communications*, 13(1):1–12, 2022.
- [94] D. Mannar, J. W. Saville, X. Zhu, S. S. Srivastava, A. M. Berezuk, K. S. Tuttle, A. C. Marquez, I. Sekirov, and S. Subramaniam. SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein–ACE2 complex. *Science*, 375(6582):760–764, 2022.
- [95] P. Masulli and A. E. Villa. The topology of the directed clique complex as a network invariant. *SpringerPlus*, 5:1–12, 2016.
- [96] F. Mémoli, Z. Wan, and Y. Wang. Persistent Laplacians: Properties, algorithms and implications. *SIAM Journal on Mathematics of Data Science*, 4(2):858–884, 2022.
- [97] Z. Meng and K. Xia. Persistent spectral–based machine learning (perspect ml) for protein-ligand binding affinity prediction. *Science Advances*, 7(19):eabc5329, 2021.
- [98] A. Muhammad and M. Egerstedt. Control using higher order Laplacians in network topologies. In *Proc. of 17th International Symposium on Mathematical Theory of Networks and Systems*, pages 1024–1038. Citeseer, 2006.
- [99] J. R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley Publishing Company, Inc., 1984.
- [100] Y. Muranov, A. Szczepkowska, and V. Vershinin. Path homology of directed hypergraphs. *Homology, Homotopy and Applications*, 24(2):347–363, 2022.

- [101] A. Myers, C. Joslyn, B. Kay, E. Purvine, G. Roek, and M. Shapiro. Topological analysis of temporal hypergraphs. In *Algorithms and Models for the Web Graph: 18th International Workshop, WAW 2023, Toronto, ON, Canada, May 23–26, 2023, Proceedings*, pages 127–146. Springer, 2023.
- [102] A. D. Parks and S. L. Lipscomb. Homology and hypergraph acyclicity: a combinatorial invariant for hypergraphs. Technical report, NAVAL SURFACE WARFARE CENTER DAHLGREN VA, 1991.
- [103] G. Petri, M. Scolamiero, I. Donato, and F. Vaccarino. Topological strata of weighted complex networks. *PloS one*, 8(6):e66506, 2013.
- [104] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, and T. E. Ferrin. Ucsf chimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1):70–82, 2021.
- [105] Y. Qiu and G.-W. Wei. Persistent spectral theory-guided protein engineering. *Nature Computational Science*, 3(2):149–163, 2023.
- [106] Y. Qiu and G.-W. Wei. Persistent spectral theory-guided protein engineering. *Nature Computational Science*, 2023.
- [107] T. Raček, O. Schindler, D. Toušek, V. Horský, K. Berka, J. Koča, and R. Svobodová. Atomic charge calculator II: web-based tool for the calculation of partial atomic charges. *Nucleic acids research*, 48(W1):W591–W596, 2020.
- [108] M. W. Reimann, M. Nolte, M. Scolamiero, K. Turner, R. Perin, G. Chindemi, P. Dłotko, R. Levi, K. Hess, and H. Markram. Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in computational neuroscience*, page 48, 2017.
- [109] S. Ren and C. Wu. Weighted simplicial complexes and weighted analytic torsions. *arXiv preprint arXiv:2103.04252*, 2021.
- [110] S. Ren, C. Wu, and J. Wu. Hodge decompositions for weighted hypergraphs. *arXiv preprint arXiv:1805.11331*, 2018.
- [111] S. Ren, C. Wu, and J. Wu. Weighted persistent homology. *Rocky Mountain Journal of Mathematics*, 48(8):2661 – 2687, 2018.
- [112] S. Ren and J. Wu. Stability of persistent homology for hypergraphs. *arXiv preprint arXiv:2002.02237*, 2020.
- [113] Y. Ren, J. W. Martini, and J. Torres. Decoupled molecules with binding polynomials of bidegree $(n, 2)$. *Journal of mathematical biology*, 78:879–898, 2019.
- [114] M. Robinson. How do we deal with noisy data?
- [115] M. Robinson. *Topological signal processing*, volume 81. Springer.

- [116] M. Robinson. Sheaves are the canonical data structure for sensor integration. *Information Fusion*, 36:208–224, 2017.
- [117] F. Russold. Persistent sheaf cohomology. *arXiv preprint arXiv:2204.13446*, 2022.
- [118] Samplchallenges. `Sampl6/cb8.mol2` at master · samplchallenges/sampl6.
- [119] A. Sharma, T. J. Moore, A. Swami, and J. Srivastava. Weighted simplicial complex: A novel approach for predicting small group evolution. In *Advances in Knowledge Discovery and Data Mining: 21st Pacific-Asia Conference, PAKDD 2017, Jeju, South Korea, May 23-26, 2017, Proceedings, Part I 21*, pages 511–523. Springer, 2017.
- [120] L. Shen, J. Liu, and G.-W. Wei. Evolutionary Khovanov homology. *AIMS Mathematics*, 9(9):26139–26165, 2024.
- [121] L. Shen, J. Liu, and G.-W. Wei. Persistent Mayer homology and persistent Mayer Laplacian. *Foundations of Data Science*, 6(4):584–612, 2024.
- [122] A. D. Shepard. *A cellular description of the derived category of a stratified space*. PhD thesis, Brown University, 1985.
- [123] S. Sirin, J. R. Apgar, E. M. Bennett, and A. E. Keating. AB-Bind: antibody binding mutational database for computational affinity predictions. *Protein Science*, 25(2):393–409, 2016.
- [124] A. J. Sommese and C. W. Wampler. *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*. WORLD SCIENTIFIC, 2005.
- [125] T. N. Starr, A. J. Greaney, C. M. Stewart, A. C. Walls, W. W. Hannon, D. Veessler, and J. D. Bloom. Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA. 1 and BA. 2 receptor-binding domains. *PLoS pathogens*, 18(11):e1010951, 2022.
- [126] F. Suwayyid and G.-W. Wei. Persistent Dirac of paths on digraphs and hypergraphs. *Foundations of Data Science*, 6(2):124–153, 2024.
- [127] M. Thakur and R. Tripathi. Linear connectivity problems in directed hypergraphs. *Theoretical Computer Science*, 410(27-29):2592–2618, 2009.
- [128] J. Verschelde. Algorithm 795: Phcpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Trans. Math. Softw.*, 25(2):251–276, June 1999.
- [129] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veessler. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, 181(2):281–292, 2020.
- [130] C. W. Wampler and A. J. Sommese. Numerical algebraic geometry and algebraic kinematics. *Acta Numerica*, 20:469–567, 2011.

- [131] R. Wang, J. Chen, K. Gao, and G.-W. Wei. Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries. *Genomics*, 113(4):2158–2170, 2021.
- [132] R. Wang, J. Chen, and G.-W. Wei. Mechanisms of SARS-CoV-2 evolution revealing vaccine-resistant mutations in Europe and America. *The Journal of Physical Chemistry Letters*, 12:11850–11857, 2021.
- [133] R. Wang, D. D. Nguyen, and G.-W. Wei. Persistent spectral graph. *International Journal for Numerical Methods in Biomedical Engineering*, 36(9):e3376, 2020.
- [134] R. Wang and G.-W. Wei. Persistent path Laplacian. *Foundations of Data Science*, 5(1):26–55, 2023.
- [135] R. Wang, R. Zhao, E. Ribando-Gros, J. Chen, Y. Tong, and G.-W. Wei. Hermes: Persistent spectral graph software. *Foundations of Data Science*, 3(1):67–97, 2020.
- [136] J. Wee, G. Bianconi, and K. Xia. Persistent Dirac for molecular representation. *Scientific Reports*, 13(1):11183, 2023.
- [137] J. Wee and K. Xia. Ollivier persistent Ricci curvature-based machine learning for the protein–ligand binding affinity prediction. *Journal of Chemical Information and Modeling*, 61(4):1617–1626, 2021.
- [138] R. K. J. Wei, J. Wee, V. E. Laurent, and K. Xia. Hodge theory-based biomolecular data analysis. *Scientific Reports*, 12(1):9699, 2022.
- [139] X. Wei, J. Chen, and G.-W. Wei. Persistent topological Laplacian analysis of SARS-CoV-2 variants. *Journal of computational biophysics and chemistry*, 22(5):569, 2023.
- [140] X. Wei and G.-W. Wei. Homotopy continuation for the spectra of persistent Laplacians. *Foundations of Data Science*, 3(4):677, 2021.
- [141] X. Wei and G.-W. Wei. Persistent sheaf Laplacians. *Foundations of Data Science*, 2024.
- [142] C. Wu, S. Ren, J. Wu, and K. Xia. Weighted (co)homology and weighted Laplacian. *arXiv preprint arXiv:1804.06990*, 2018.
- [143] C. Wu, S. Ren, J. Wu, and K. Xia. Discrete morse theory for weighted simplicial complexes. *Topology and its Applications*, 270:107038, 2020.
- [144] K. Xia, X. Feng, Y. Tong, and G. W. Wei. Persistent homology for the quantitative prediction of fullerene stability. *Journal of computational chemistry*, 36(6):408–422, 2015.
- [145] K. Xia and G.-W. Wei. Persistent homology analysis of protein structure, flexibility, and folding. *International journal for numerical methods in biomedical engineering*, 30(8):814–844, 2014.

- [146] K. Xia and G.-W. Wei. Persistent topology for cryo-EM data analysis. *International Journal for Numerical Methods in Biomedical Engineering*, 31(8), 2015.
- [147] Z. Xiang and B. Honig. Extending the accuracy limits of prediction for side-chain conformations. *Journal of molecular biology*, 311(2):421–430, 2001.
- [148] C. Xu, Y. Wang, C. Liu, C. Zhang, W. Han, X. Hong, Y. Wang, Q. Hong, S. Wang, Q. Zhao, et al. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Science advances*, 7(1):eabe5575, 2021.
- [149] K. Yegnesh. Persistence and sheaves. *arXiv preprint arXiv:1612.03522*, 2016.
- [150] H. R. Yoon. *Cellular sheaves and cosheaves for distributed topological data analysis*. PhD thesis, University of Pennsylvania, 2018.
- [151] A. Zomorodian and G. Carlsson. Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 347–356, 2004.

APPENDIX A

STATISTICS OF NONZERO EIGENVALUES OF PERSISTENT LAPLACIANS

Here we present some statistics of nonzero eigenvalues of persistent Laplacians. An Alpha filtration is constructed from the atoms of molecule CB8 [118]. The spectrum of $\Delta_d^{X_r, X_{r+p}}$ for $d = 0, 1$ and selected r, p will be calculated. The radius r will be a multiple of 0.01\AA .

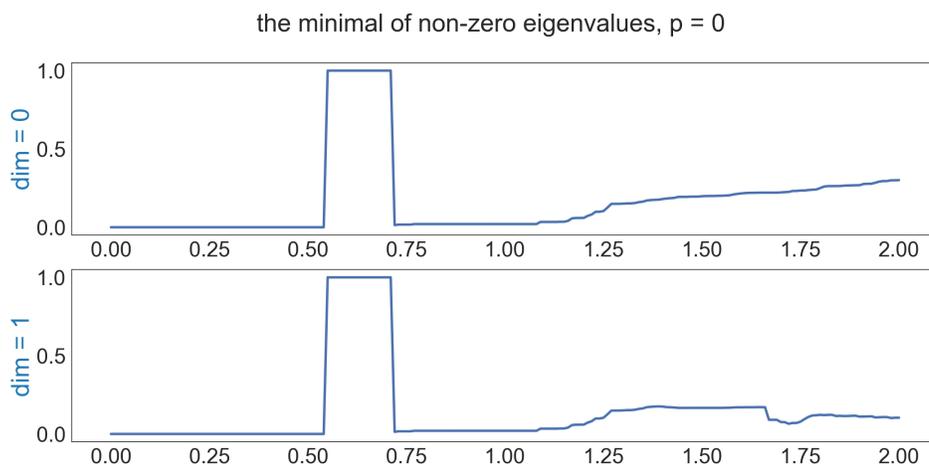


Figure A.1 The minimal of non-zero eigenvalues when $p = 0$.

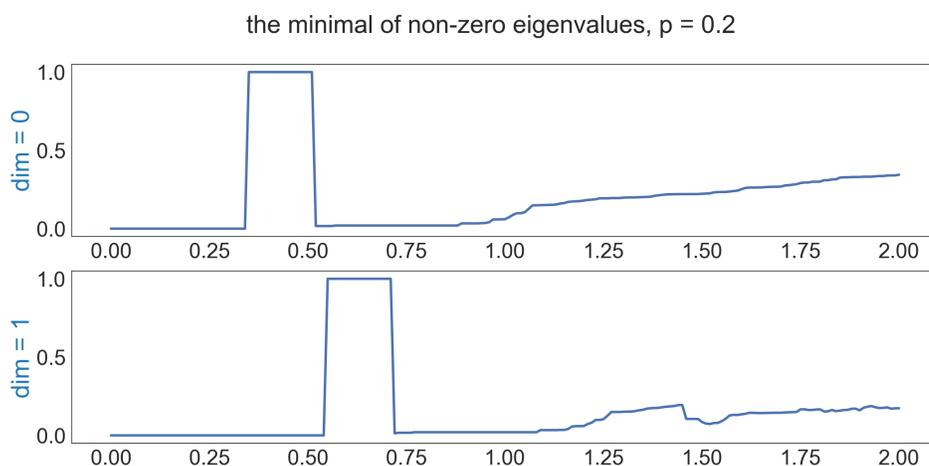


Figure A.2 The minimal of non-zero eigenvalues when $p = 0.2$.

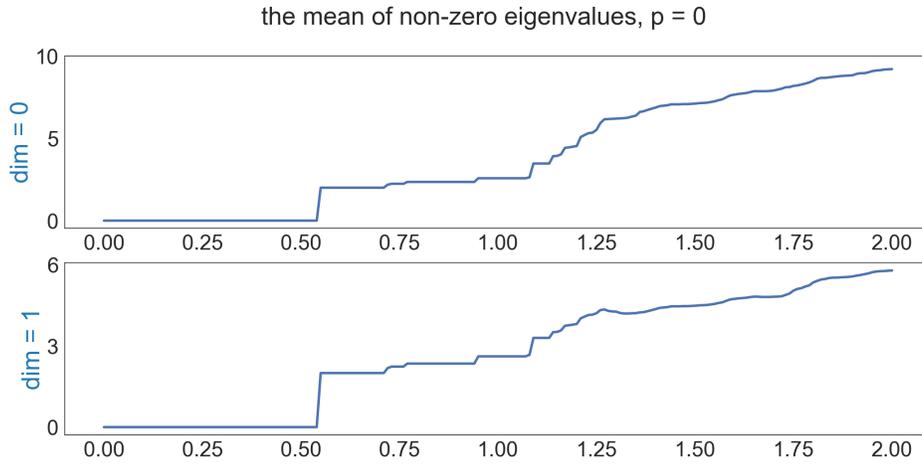


Figure A.3 The mean of non-zero eigenvalues when $p = 0$.

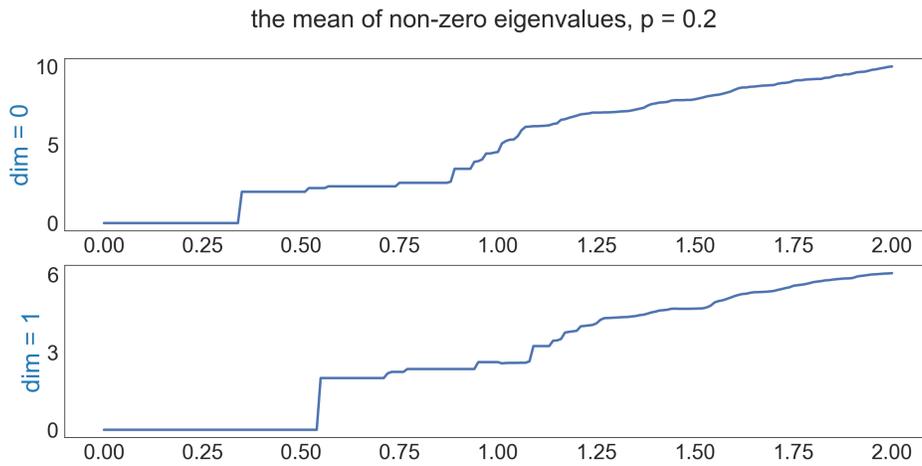


Figure A.4 The mean of non-zero eigenvalues when $p = 0.2$.

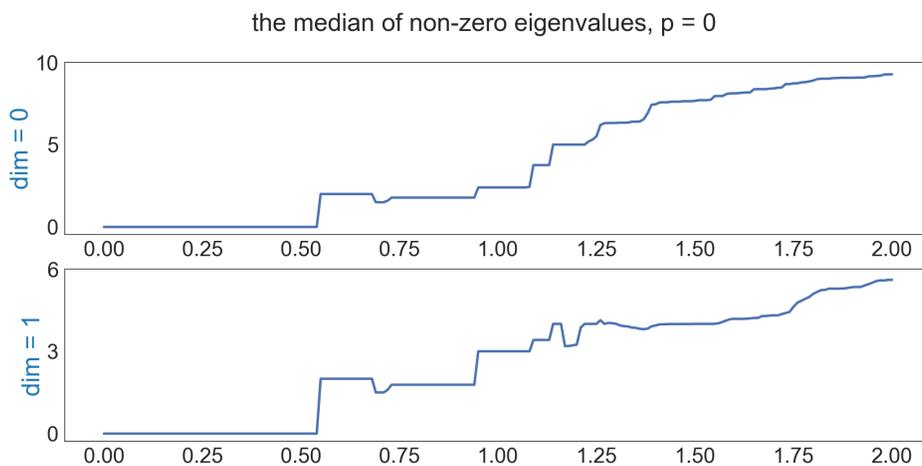


Figure A.5 The median of non-zero eigenvalues when $p = 0$.

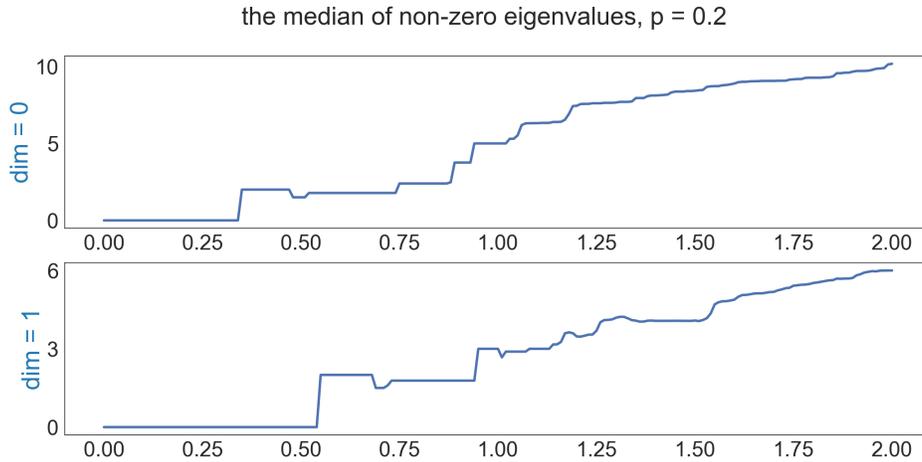


Figure A.6 The median of non-zero eigenvalues when $p = 0.2$.

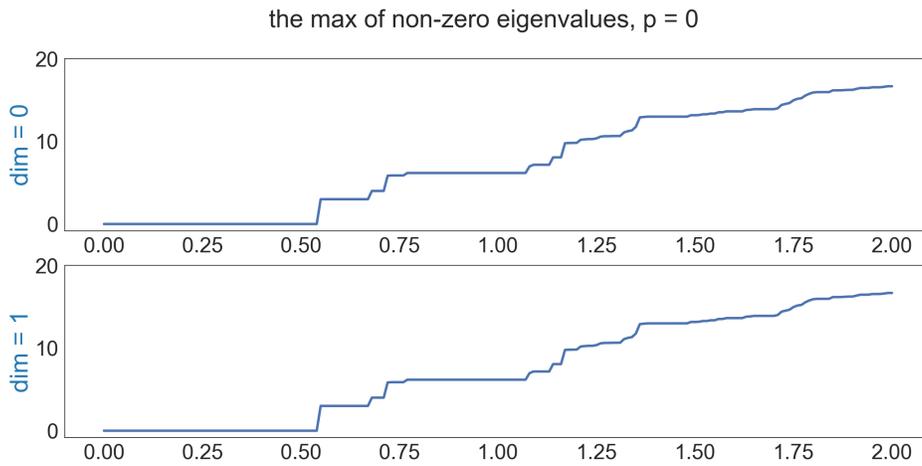


Figure A.7 The maximal of non-zero eigenvalues when $p = 0$.

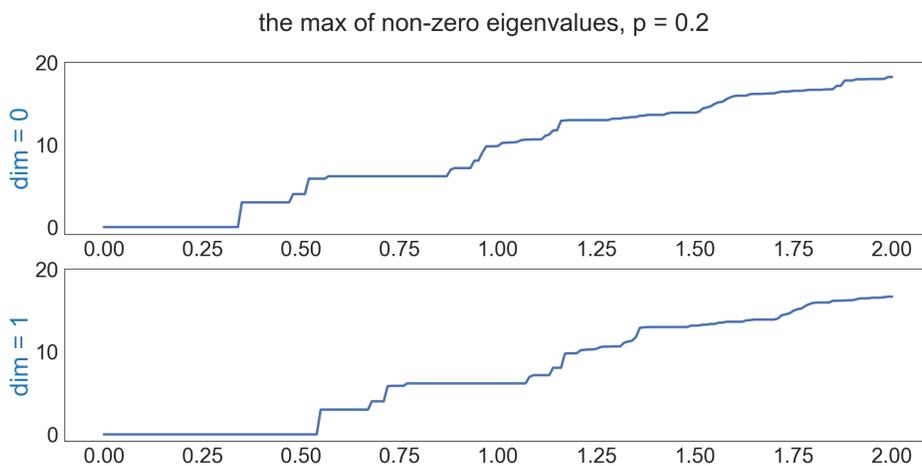


Figure A.8 The maximal of non-zero eigenvalues when $p = 0.2$.

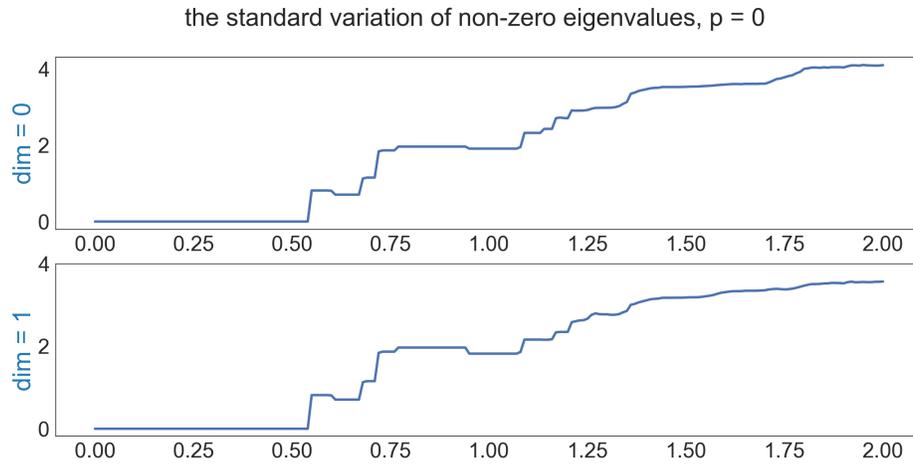


Figure A.9 The standard variation of non-zero eigenvalues when $p = 0$.

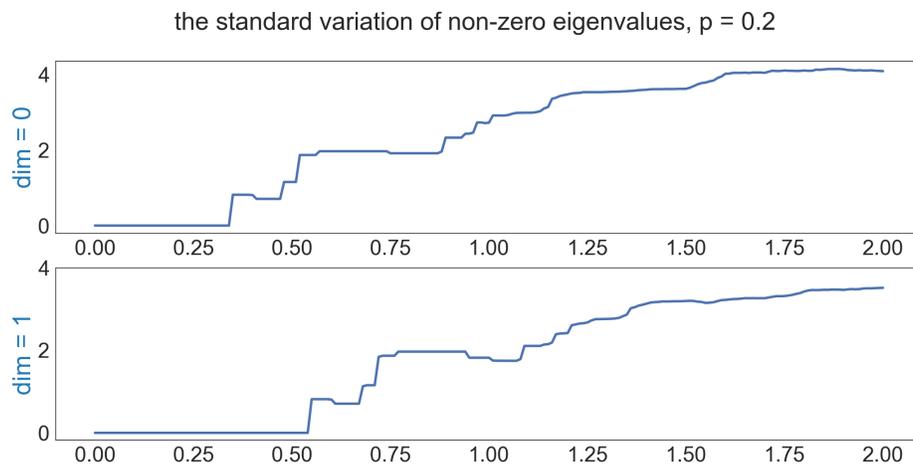


Figure A.10 The standard variation of non-zero eigenvalues when $p = 0.2$.

APPENDIX B

METHOD OF HOMOTOPY CONTINUATION

Our exposition of homotopy continuation follows [8]. More theoretical treatment of this topic can be found elsewhere [124].

Solving a system of polynomial equations f by homotopy continuation basically consists of three steps: 1) build a start system g such that g can be solved easily; 2) build a homotopy between two systems f and g ; 3) track the roots of g to the roots of f .

We first look at a simple example. Let us say we wish to solve the following polynomial in one complex variable

$$f(z) = -2z^3 - 5z^2 + 4z + 1.$$

We take a similar and simpler polynomial $g(z)$ and deform the roots of $g(z)$ to $f(z)$. For instance we may take $g(z) = z^3 + 1$ and construct a linear homotopy

$$h(z, s) = sg(z) + (1 - s)f(z)$$

where s is a complex variable. Though the second parameter of h is a complex variable, we still call h a homotopy between f and g for convenience. Then we parametrize s by a curve

$$s(t) = \frac{\gamma t}{\gamma t + (1 - t)}, t \in [0, 1], \gamma \in \mathbb{C} \setminus \mathbb{R}$$

(this is called the gamma trick [8, Section 6.1] and there are technical reasons behind such choice of parametrization). We substitute $s(t)$ in $h(z, s)$ and clear denominators, then obtain a usual homotopy

$$H(z, t) = \gamma t g(z) + (1 - t)f(z), t \in [0, 1].$$

For each $t_0 \in [0, 1]$, $H(z, t_0)$ is a polynomial. Once we know how to numerically solve $H(z, t_0 - \Delta t) = 0$ from the known roots of $H(z, t_0)$, we can pick a grid of $[0, 1]$ and track the known roots of g step by step all the way to the solutions of f . This process is called the path

tracking. Now suppose $H(z(t), t) = 0$ for any $t \in (0, 1]$ with $z(1)$ a root of g . Differentiate $H(z(t), t)$ with respect to t , we have the Davidenko differential equation

$$\frac{\partial H}{\partial t}(z(t), t) + \frac{\partial H}{\partial z}(z(t), t) \frac{dz(t)}{dt} = 0.$$

If $\frac{\partial H}{\partial z}(z(t), t)$ is nonzero, the Davidenko differential equation can be rewritten as

$$\frac{dz(t)}{dt} = - \left(\frac{\partial H}{\partial z}(z(t), t) \right)^{-1} \frac{\partial H}{\partial t}(z(t), t)$$

As we know the value of $z(t)$ at t_0 , we have indeed transformed our original problem of tracking roots to the classical initial value problem of ordinary differential equation (ODE).

One may use any ODE method to predict $z(t_0 - \Delta t)$ (The default ODE solver employed by Bertini is RKF45). For instance we can apply the simplest Euler's method and get

$$z(t_0 - \Delta t) = z(t_0) - \left(\frac{\partial H}{\partial z}(z(t_0), t_0) \right)^{-1} \frac{\partial H}{\partial t}(z(t_0), t_0) \Delta t.$$

Since we also know that $H(z(t_0 - \Delta t), t_0 - \Delta t)$ should be zero, we can apply several iterations of Newton's method to update $z(t_0 - \Delta t)$. Such combination of an ODE predictor with Newton's method is called a predictor-corrector method.

Now after the path tracking from $t = 1$ to $t = 0$, we get a sequence $\{z(t_i)\}$. If the limit $\lim_{t_i \rightarrow 0} z(t_i)$ exists and is finite, we think of $\lim_{t_i \rightarrow 0} z(t_i)$ as a solution of f .

Example B.0.1. The reader may wonder why we do not just use

$$H(z, t) = tg(z) + (1 - t)f(z), t \in [0, 1].$$

Consider the example $H(z, t) = t(z^2 - 1) + (1 - t)(5 - z^2)$. When $t = 1/2$, $H(z, 1/2) = 2$ has no roots. When $t = 5/6$, $H(z, 5/6) = 2/3z^2$ has a singular root 0, and the derivative of it at $z = 0$ is zero.

Example B.0.2. Though usually we are only interested in real roots of the target system, we should also track complex roots. Consider the homotopy

$$h(z, t) = z^4 - e^{2\pi i(1-t)} + 0.25 = 0, t \in [0, 1].$$

At $t = 1$, $h(z, 1)$ has two real roots $\pm\sqrt[4]{0.75}$ and two imaginary roots $\pm\sqrt[4]{0.75}i$. As t goes from 1 to 0, $e^{2\pi i(1-t)}$ travels around the unit circle in the complex plane counterclockwise; The two real roots will be deformed to the two imaginary ones and vice versa.

APPENDIX C

APPENDIX OF THE SHEAF CHAPTER



Figure C.1 A square and a trapezoid. Coordinates of vertices are shown.

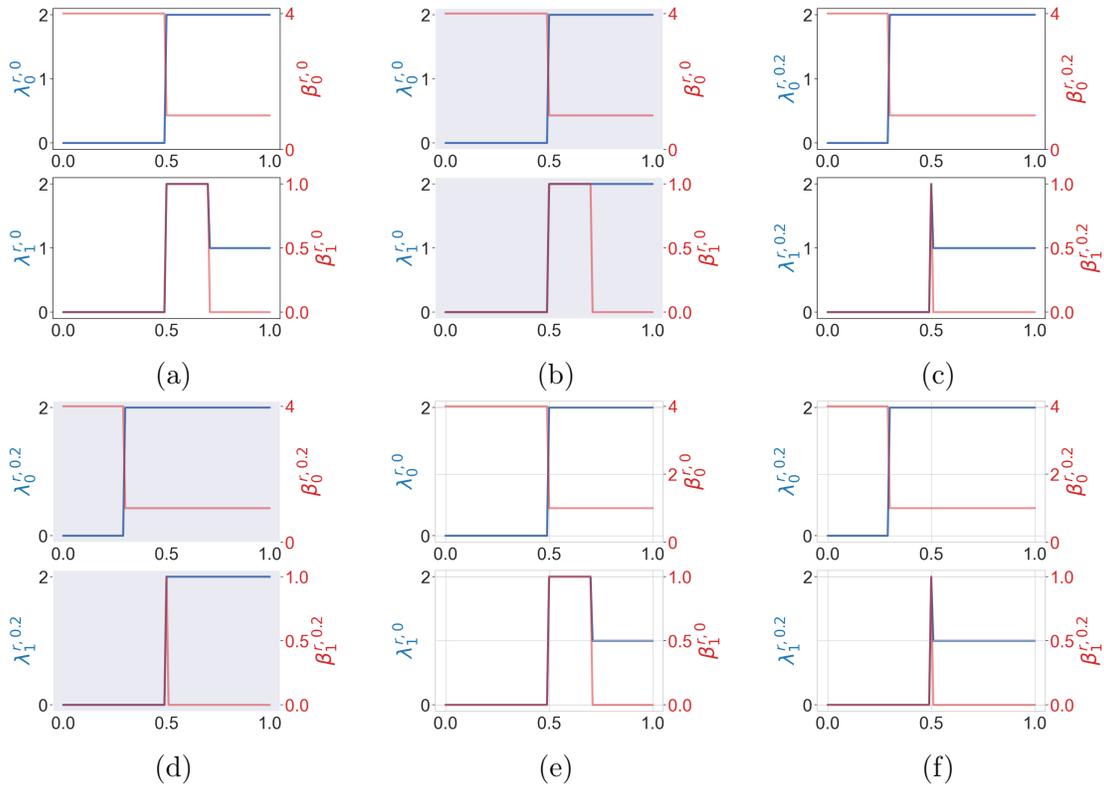


Figure C.2 The results of the square. We consider pairs (X_r, X_r) or $(X_r, X_{r+0.2})$ in a filtration. The results of the labeled point cloud $\{(0, 0), 1), ((1, 0), 1), ((1, 1), 1), ((0, 1), 1)\}$ are shown in (a)(c). The results of the point cloud $\{(0, 0), (1, 0), (1, 1), (0, 1)\}$ are shown in (b)(d). The results of the labeled point cloud $\{(0, 0), 1), ((1, 0), 1), ((1, 1), 1), ((0, 1), -1)\}$ are shown in (e)(f).

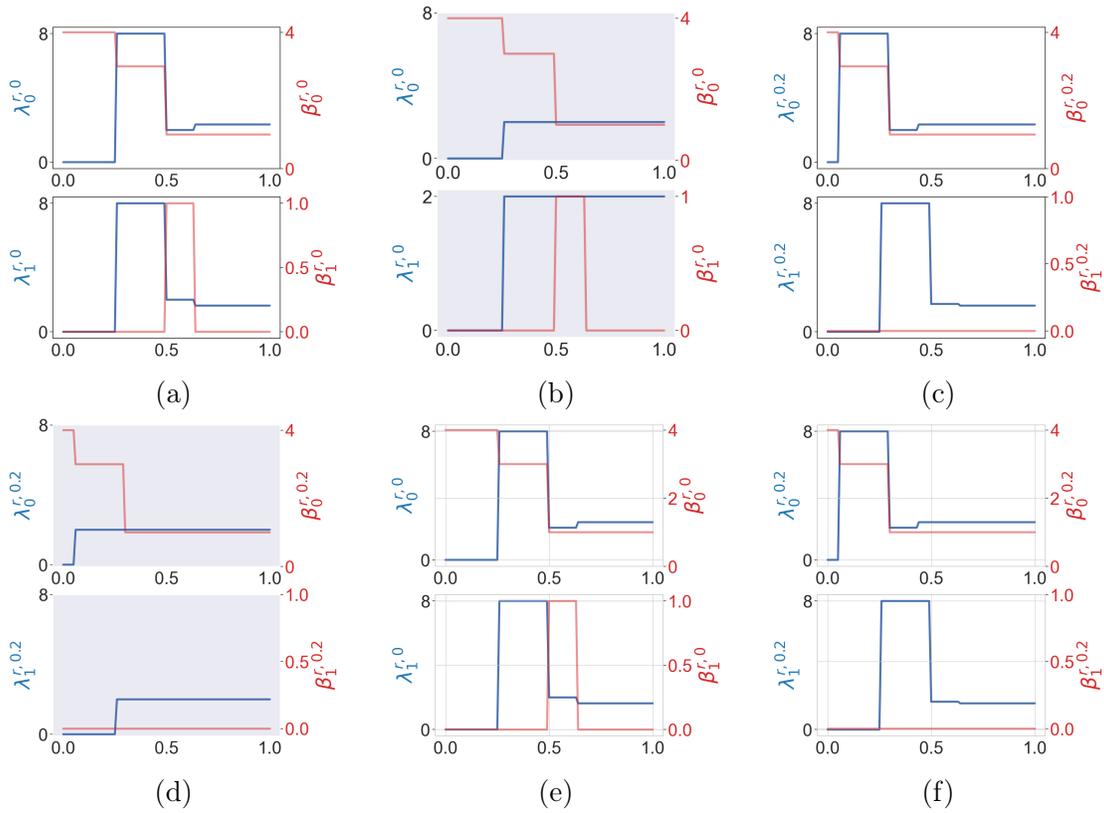


Figure C.3 The results of the trapezoid. We consider pairs (X_r, X_r) or $(X_r, X_{r+0.2})$ in a filtration. The results of the labeled point cloud $\{((0, 0), 1), ((1, 0), 1), ((3/4, \sqrt{15})/4, 1), ((0, 1), 1)\}$ are shown in (a)(c). The results of the point cloud $\{(0, 0), (1, 0), (3/4, \sqrt{15}/4), (1/4, \sqrt{15}/4)\}$ are shown in (b)(d). The results of the labeled point cloud $\{((0, 0), 1), ((1, 0), 1), ((3/4, \sqrt{15})/4, 1), ((1/4, \sqrt{15})/4, -1)\}$ are shown in (e)(f).

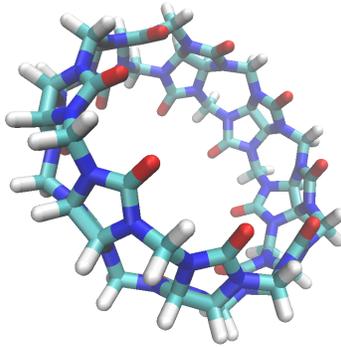


Figure C.4 The structure of CB8.

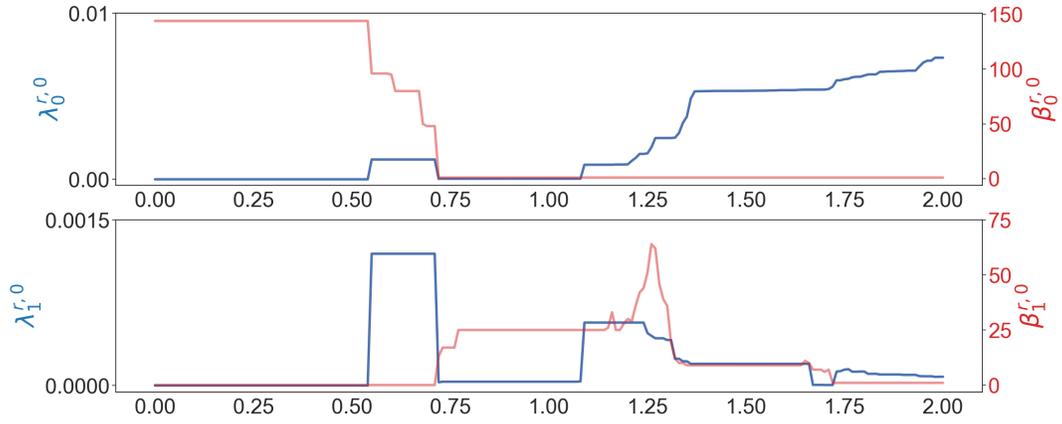


Figure C.5 The results of the CB8 when $p = 0$.

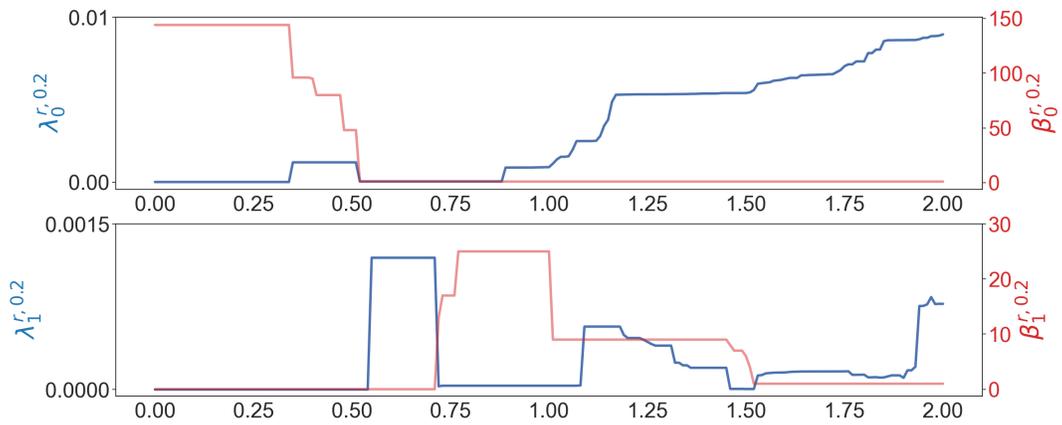


Figure C.6 The results of the CB8 when $p = 0.2$.

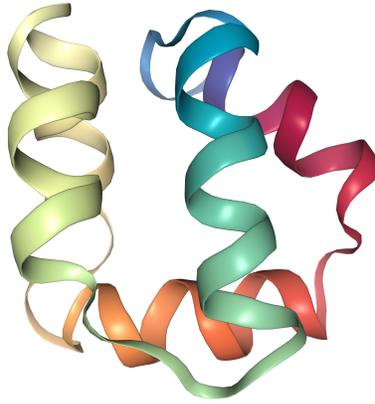


Figure C.7 Illustration of the structure of bacteriocin AS-48.

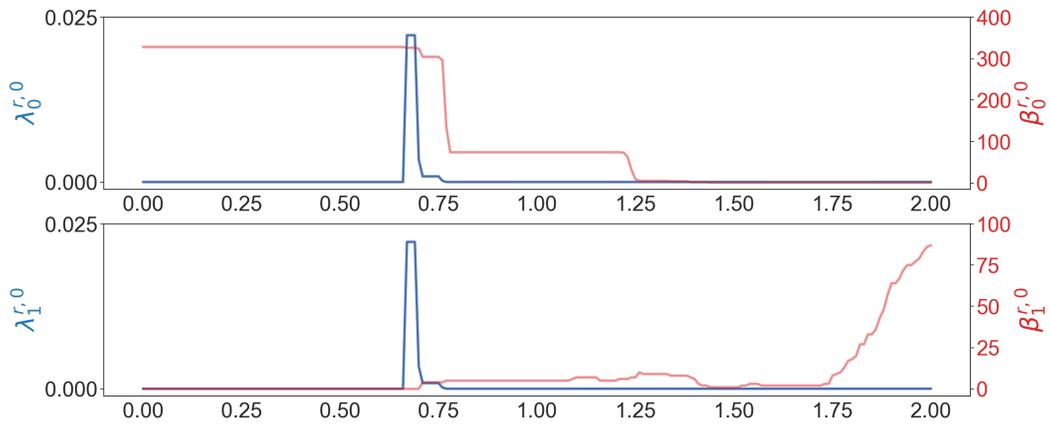


Figure C.8 The results of AS-48 when $p = 0$.

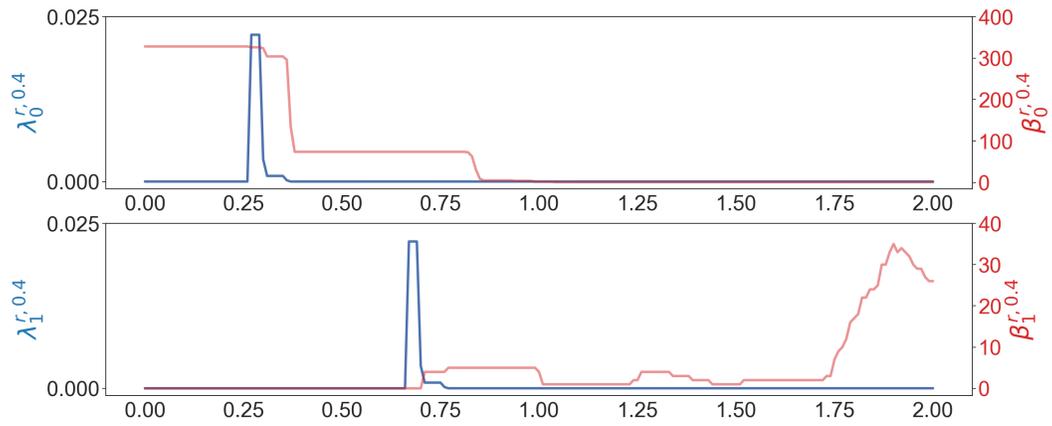


Figure C.9 Results of AS-48 when $p = 0.4$.

APPENDIX D

APPENDIX OF THE SARS-COV-2 CHAPTER

Here we provide the information of PDB structures we used and additional topological analysis using persistent Laplacian and persistent sheaf Laplacian. The information of PDB structures is given in Table D.1. PTL results are given in Figures D.1, D.2, D.3, D.4, D.5, D.6, D.7, D.8, D.9, D.10, D.11, and D.12.

Specifically, Figures D.1, and D.2 are element-specific analysis of the wide type, Alpha, Beta, Gamma, BA.1, and BA.2 using LP and PSL, respectively. Figure D.3 presents carbon specific analysis of the wide type, Alpha, Beta, Gamma, BA.1, and BA.2. Figures D.4, D.5, D.6, and D.7 demonstrate the PL analysis of Alpha, Beta, BA.2, and BA.2, respectively. The spectral analysis of three major types of elements, namely carbon atoms, nitrogen atoms, and oxygen atoms, is presented in these figures. Finally, D.8, D.9, D.10, D.11, and D.12 illustrate the PSL analysis of the wide type, Alpha, Beta, BA.2, and BA.2, respectively. These figures display the spectral analysis of three major types of elements, namely carbon atoms, nitrogen atoms, and oxygen atoms.

PDB ID	Method	Resolution (unit: Å)	Description
6M0J[80]	X-ray diffraction	2.45	wild type RBD-ACE2
8DLK[93]	Electron microscopy	3.04	Alpha RBD-ACE2
8DLN[93]	Electron microscopy	3.04	Beta RBD-ACE2
8DLQ[93]	Electron microscopy	2.77	Gamma RBD-ACE2
7T9L[94]	Electron microscopy	2.66	BA.1 RBD-ACE2
7XB0[82]	X-ray diffraction	2.90	BA.2 RBD-ACE2
7DF3[148]	Electron microscopy	2.70	wild type spike
7LWS[51]	Electron microscopy	3.22	Alpha spike
7LYM[51]	Electron microscopy	3.57	Beta spike
7TF8[50]	Electron microscopy	3.36	BA.1 spike
7XIX[21]	Electron microscopy	3.25	BA.2 spike

Table D.1 Information of PDB 3D structures used in this work.

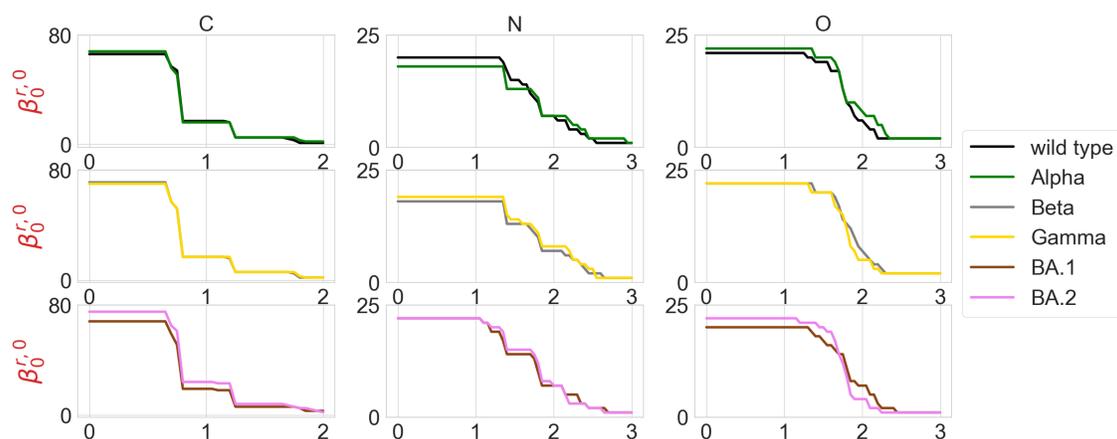


Figure D.1 Illustration of persistent Betti numbers of element specific persistent Laplacian of the residue 501 mutation site at different filtration values, i.e., radii (unit: Å). The wild type (PDB ID: 6M0J) and Alpha (PDB ID: 8DLK) are given in the first row. The Beta (PDB ID: 8DLN) and Gamma (PDB ID: 8DLQ) are given in the second row. The BA.1 (PDB ID: 7T9L) and BA.2 (PDB ID: 7XB0) are given the third row.

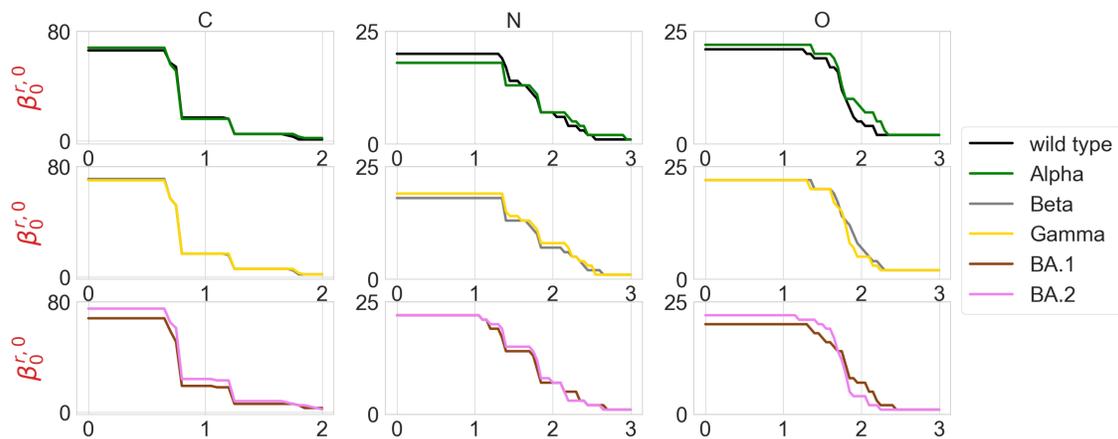


Figure D.2 Illustration of persistent Betti numbers of element specific persistent sheaf Laplacian of the residue 501 mutation site at different filtration values, i.e., radii (unit: Å). The wild type (PDB ID: 6M0J) and Alpha (PDB ID: 8DLK) are given in the first row. The Beta (PDB ID: 8DLN) and Gamma (PDB ID: 8DLQ) are given in the second row. The BA.1 (PDB ID: 7T9L) and BA.2 (PDB ID: 7XB0) are given the third row.

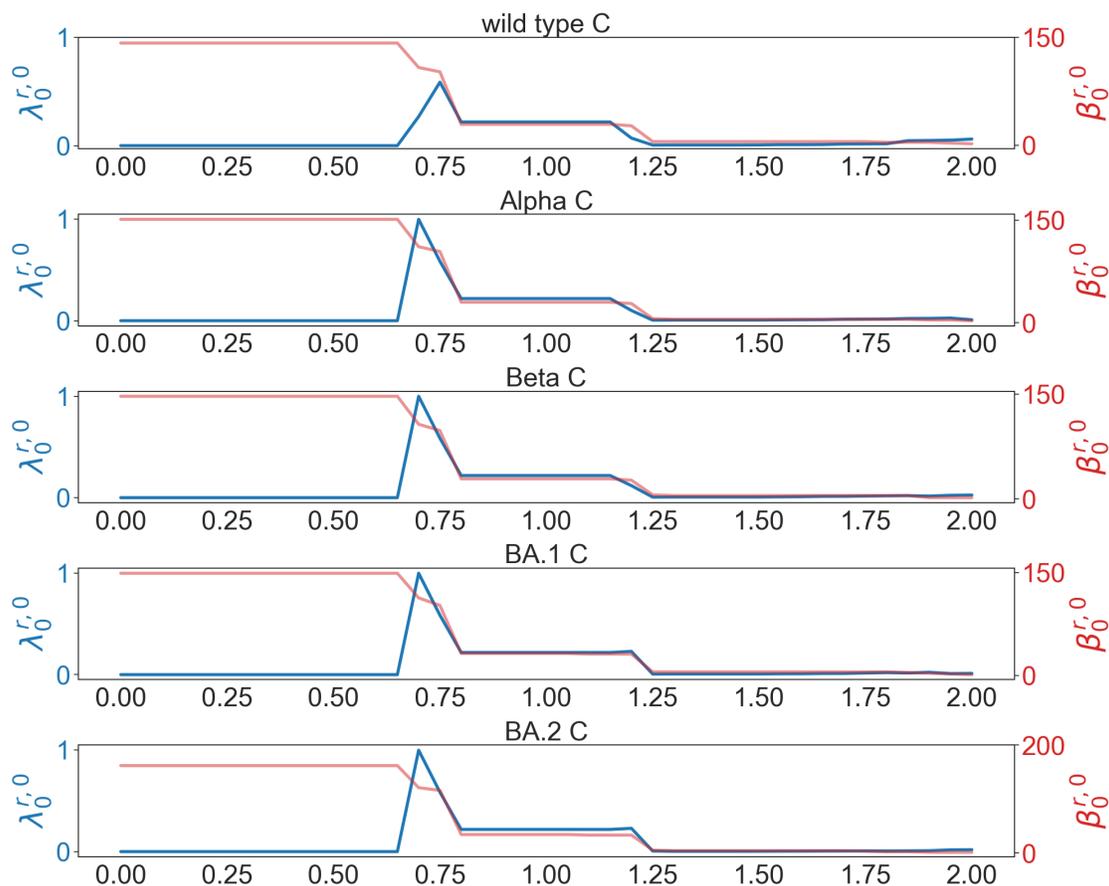


Figure D.3 Illustration of persistent Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent Laplacian of carbon atoms of the RBD binding site in RBD-ACE2 complex of wild type (PDB ID: 6M0J), Alpha (PDB ID: 8DLK), Beta (PDB ID: 8DLN), BA.1 (PDB ID: 7T9L), and BA.2 (PDB ID: 7XB0) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of the wild type, Alpha, Beta, BA.1, and BA.2 variants, respectively.

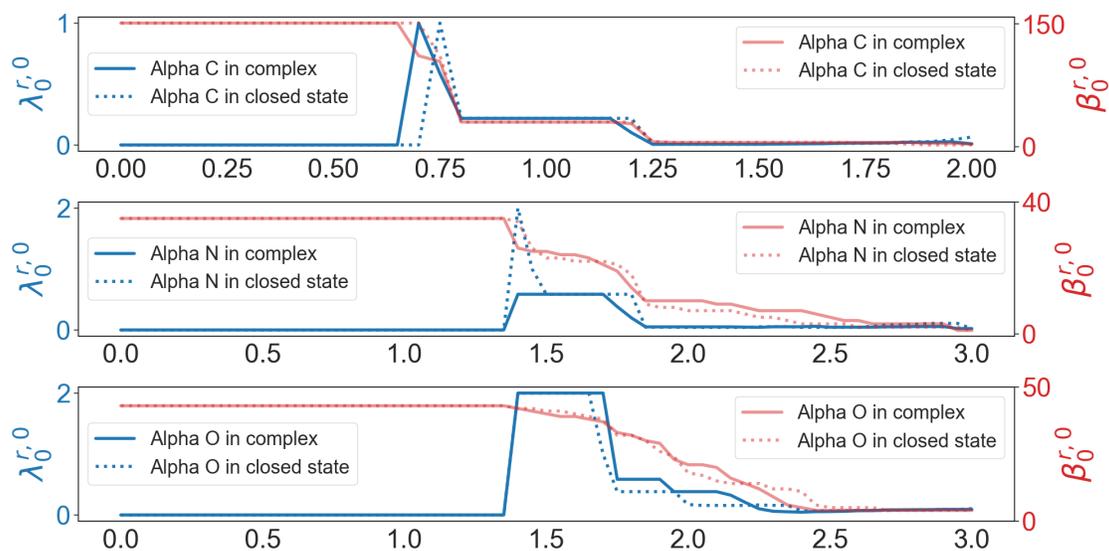


Figure D.4 Illustration of persistent Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent Laplacian of the RBD binding site of Alpha RBD-ACE2 complex (PDB ID: 8DLK) and closed state spike protein (PDB ID: 7LWS, Chain ID: A) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of carbon atoms, nitrogen atoms, and oxygen atoms, respectively.

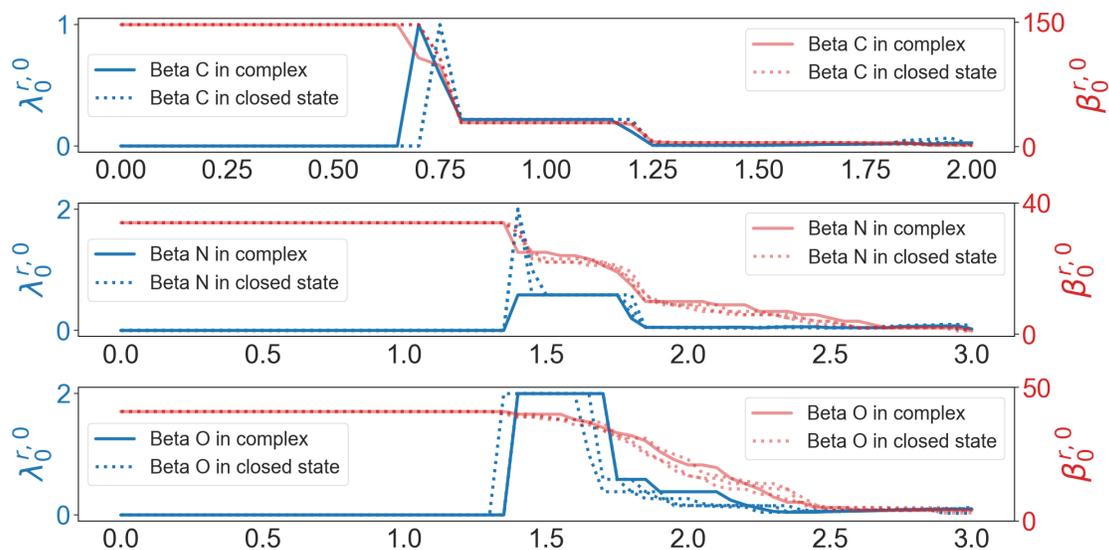


Figure D.5 Illustration of persistent Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent Laplacian of the RBD binding site of Beta RBD-ACE2 complex (PDB ID: 8DLN) and closed state spike protein (PDB ID: 7LYM, Chain ID: A, B, C) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of carbon atoms, nitrogen atoms, and oxygen atoms, respectively.

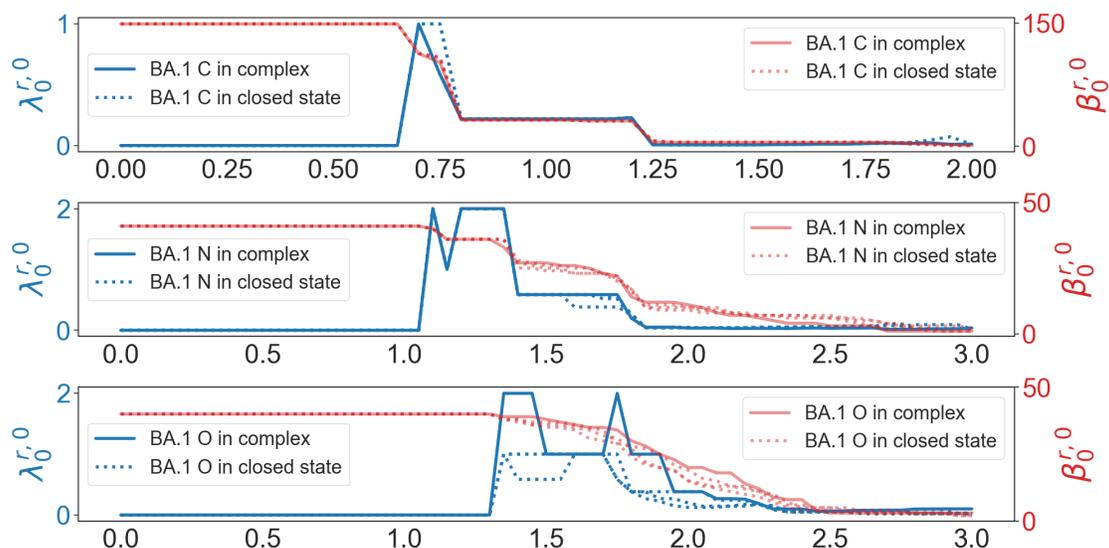


Figure D.6 Illustration of persistent Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent Laplacian of the RBD binding site of BA.1 RBD-ACE2 complex (PDB ID: 7T9L) and closed state spike protein (PDB ID: 7TF8, Chain ID: A, B, C) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of carbon atoms, nitrogen atoms, and oxygen atoms, respectively.

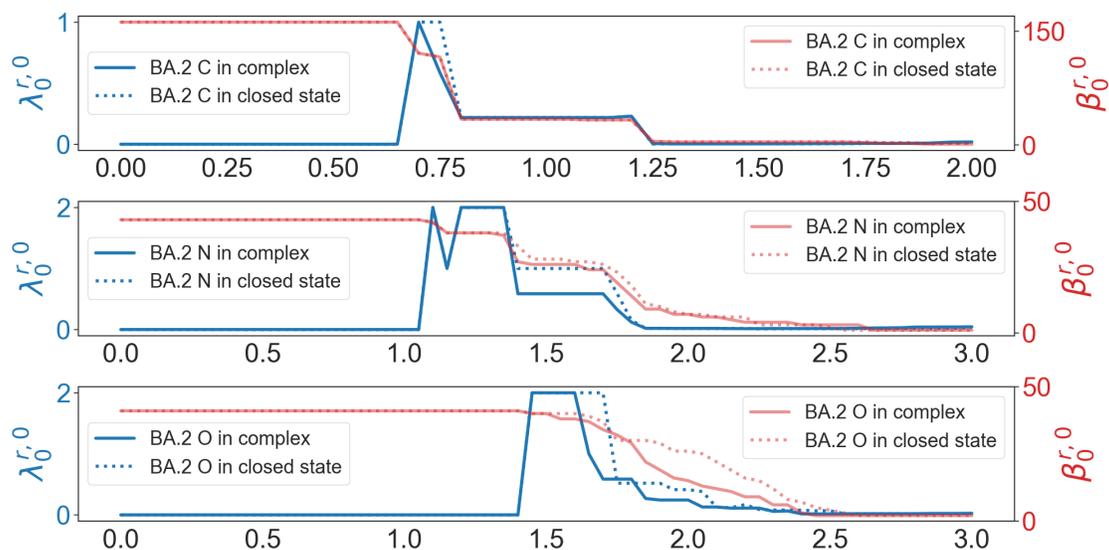


Figure D.7 Illustration of persistent Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent Laplacian of the RBD binding site of BA.2 RBD-ACE2 complex (PDB ID: 7XB0) and closed state spike protein (PDB ID: 7XIX, Chain ID: A) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of carbon atoms, nitrogen atoms, and oxygen atoms, respectively.

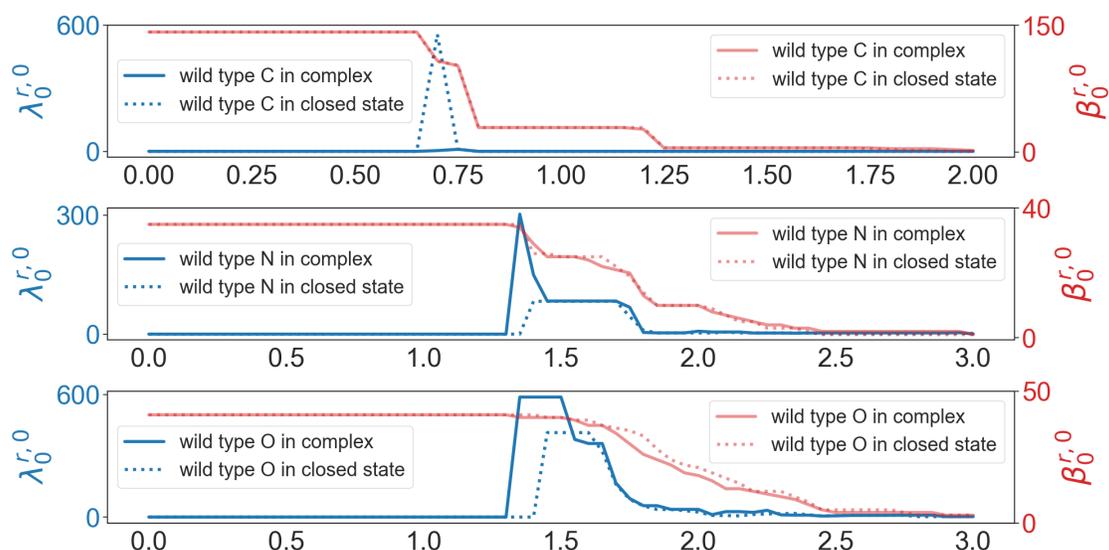


Figure D.8 Illustration of persistent sheaf Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent sheaf Laplacian of the RBD binding site of wild type RBD-ACE2 complex (PDB ID: 6M0J) and closed state spike protein (PDB ID: 7DF3, Chain ID: A) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of carbon atoms, nitrogen atoms, and oxygen atoms respectively.

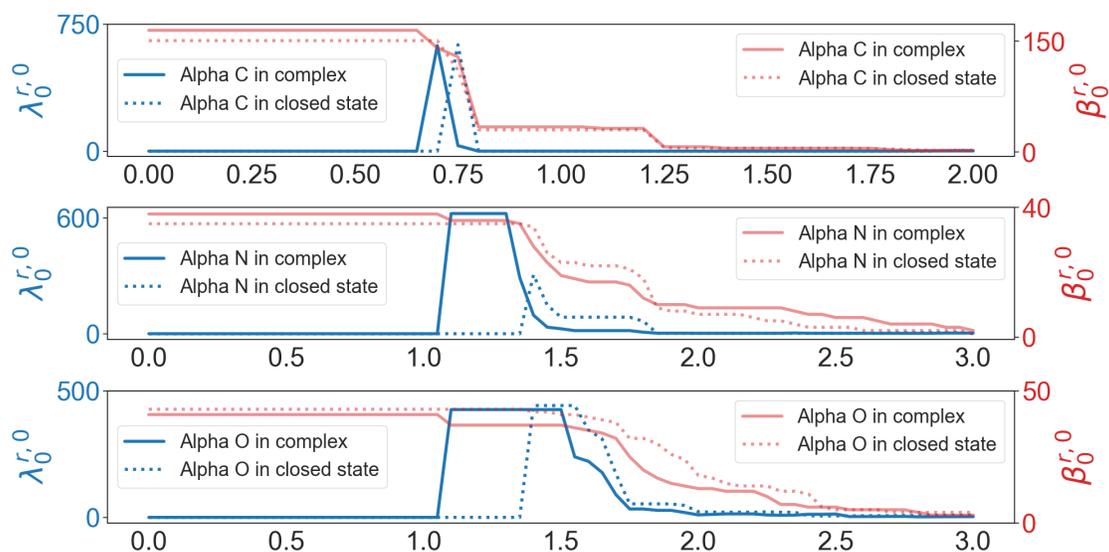


Figure D.9 Illustration of persistent sheaf Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent sheaf Laplacian of the RBD binding site of Alpha RBD-ACE2 complex (PDB ID: 8DLK) and closed state spike protein (PDB ID: 7LWS, Chain ID: A) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of carbon atoms, nitrogen atoms, and oxygen atoms, respectively.

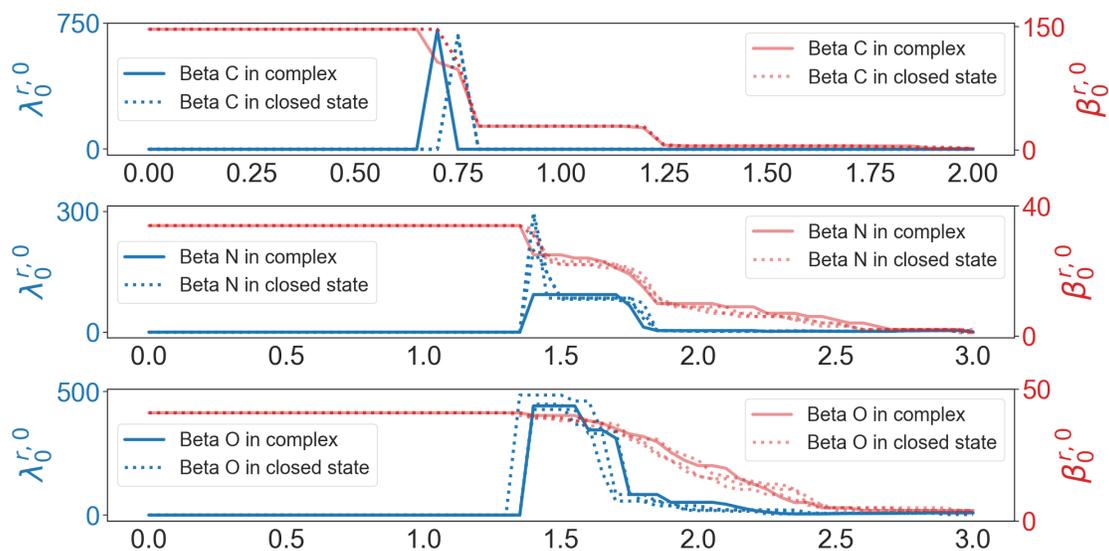


Figure D.10 Illustration of persistent sheaf Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent sheaf Laplacian of the RBD binding site of Beta RBD-ACE2 complex (PDB ID: 8DLN) and closed state spike protein (PDB ID: 7LYM, Chain ID: A, B, C) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of carbon atoms, nitrogen atoms, and oxygen atoms, respectively.

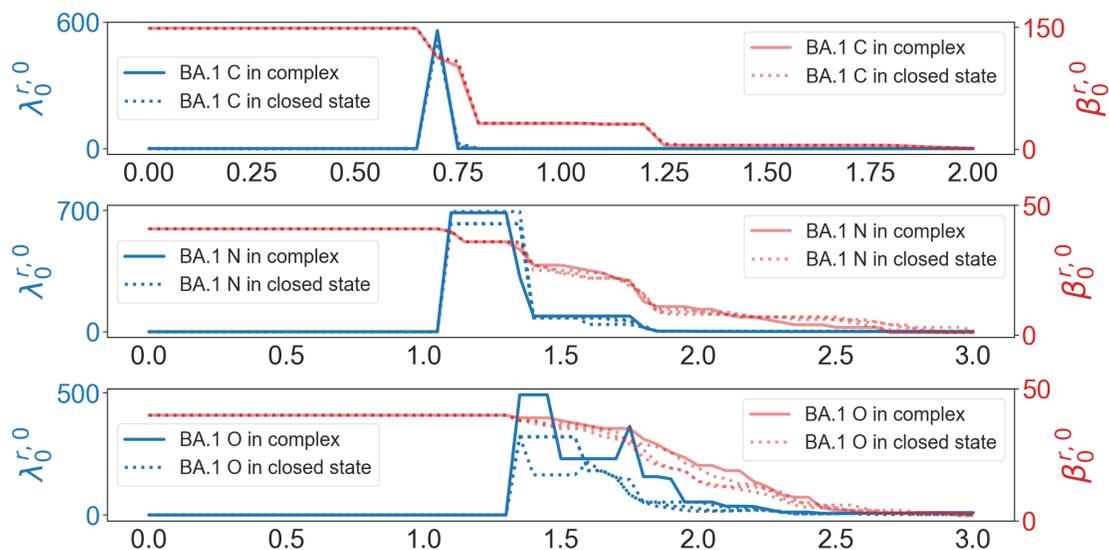


Figure D.11 Illustration of persistent sheaf Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent sheaf Laplacian of the RBD binding site of BA.1 RBD-ACE2 complex (PDB ID: 7T9L) and closed state spike protein (PDB ID: 7TF8, Chain ID: A, B, C) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of carbon atoms, nitrogen atoms, and oxygen atoms, respectively.

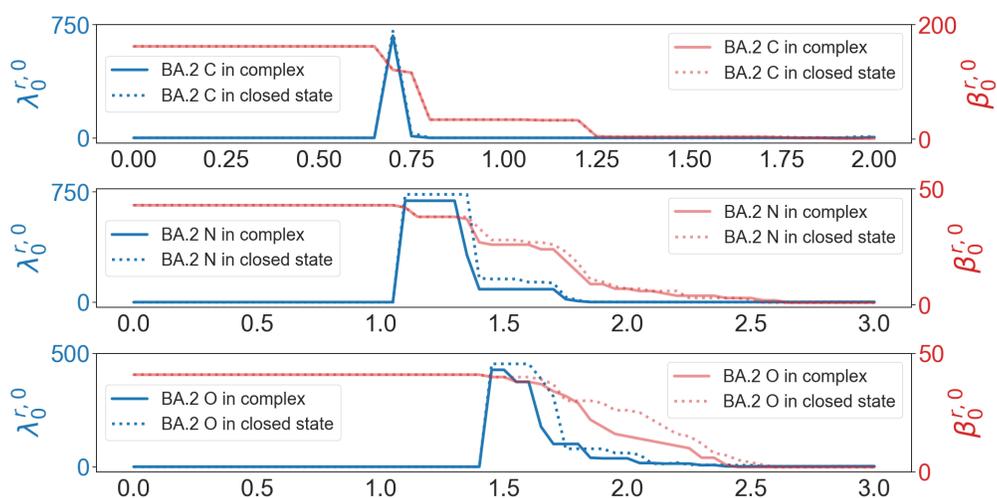


Figure D.12 Illustration of persistent sheaf Betti numbers (red line) and the first nonzero eigenvalues (blue line) of persistent sheaf Laplacian of the RBD binding site of BA.2 RBD-ACE2 complex (PDB ID: 7XB0) and closed state spike protein (PDB ID: 7XIX, Chain ID: A) at different filtration values, i.e., radii (unit: Å). The graphs from top to bottom represent the results of carbon atoms, nitrogen atoms, and oxygen atoms, respectively.