



---

*Research article*

## Multi-scale Jones polynomial and persistent Jones polynomial for knot data analysis

Ruzhi Song<sup>1,3</sup>, Fengling Li<sup>1,\*</sup>, Jie Wu<sup>2,3</sup>, Fengchun Lei<sup>1</sup> and Guo-Wei Wei<sup>4,5,6</sup>

<sup>1</sup> School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, Liaoning, China

<sup>2</sup> School of Mathematical Sciences, Hebei Normal University, Shijiazhuang 050024, Hebei, China

<sup>3</sup> Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China

<sup>4</sup> Department of Mathematics, Michigan State University, 426 Auditorium Road, East Lansing, MI 48824, USA

<sup>5</sup> Department of Biochemistry and Molecular Biology, Michigan State University, 426 Auditorium Road, East Lansing, MI 48824, USA

<sup>6</sup> Department of Electrical and Computer Engineering, Michigan State University, 426 Auditorium Road, East Lansing, MI 48824, USA

\* **Correspondence:** Email: [fenglingli@dlut.edu.cn](mailto:fenglingli@dlut.edu.cn).

**Abstract:** Many structures in science, engineering, and art can be viewed as curves in 3-space. The entanglement of these curves plays a crucial role in determining the functionality and physical properties of materials. Many concepts in knot theory provide theoretical tools to explore the complexity and entanglement of curves in 3-space. However, classical knot theory focuses on global topological properties and lacks the consideration of local structural information, which is critical in practical applications. In this work, two localized models based on the Jones polynomial were proposed, namely, the multi-scale Jones polynomial and the persistent Jones polynomial. The stability of these models, especially the insensitivity of the multi-scale and persistent Jones polynomial models to small perturbations in curve collections, was analyzed, thus ensuring their robustness for real-world applications.

**Keywords:** knot data analysis; curve data analysis; Jones polynomial; localization; stability; protein flexibility

**Mathematics Subject Classification:** 57K10, 92C10

---

## 1. Introduction

Knot theory, a branch of mathematics that focuses on the study of mathematical knots, is primarily concerned with classifying and analyzing knots based on their essential properties under ambient isotopy [1]. This approach allows mathematicians to disregard the specific manner in which knots are embedded in 3-space, emphasizing instead the invariants that remain unchanged under continuous deformations. There are several knot invariants, such as the knot crossing number, the knot group [1], the Alexander polynomial [2], the Jones polynomial [3], the knot Floer homology [4], and the Khovanov homology [5].

Knot theory has applications in many fields, including physics [6], chemistry [7], and biology [8–10]. In practical applications, however, two major challenges arise: Many structures do not form closed loops, and ambient isotopy can significantly alter local structures while preserving global knot characteristics. For example, open curves in the 3-space, such as polymers [11, 12], textiles [13], chemical compounds [14], and biological molecules [15, 16], often exhibit local entanglement that critically affects their physical properties and functions. Topological invariants, functions that remain unchanged under ambient isotopy, are essential for analysis of knots and links [17–19]. However, these invariants do not extend to open curves, since open curves can be continuously deformed without requiring cutting or rejoining, making topological equivalence inapplicable.

In recent years, methods that incorporate classical concepts from knot theory and are more applicable to practical problems have been proposed. Compared to topological data analysis (TDA), the concept of knot data analysis (KDA) was formally introduced in [20]. Panagiotou and Plaxco [21] demonstrated the utility of the Gauss link integral in protein entanglement, particularly to understand protein folding kinetics and improve future folding models. Based on this, Baldwin and Panagiotou [22] introduced a new measure of local topological and geometrical free energy based on writhing and torsion of protein chains, highlighting its critical role in the rate-limiting steps of protein folding. In addition, Baldwin et al. [23] extended these topological concepts to the study of the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein, showing how local geometric features such as writhe and torsion influence its stability and behavior. In a related effort, Shen et al. [20] introduced multi-scale Gauss link integral (mGLI), a novel method leveraging the Gauss link integral to quantify the entanglement and topological complexity of both open and closed curves at various scales. This versatile approach has broad applications in the analysis of curve structures in both physical and biological systems.

The Jones polynomial [3], a fundamental invariant in knot theory, provides a polynomial measure of entanglement that distinguishes different types of knots by smoothing their crossings. Panagiotou and Kauffman [24] extended this concept to an open curve and proposed a continuous measure of entanglement that converges to the classical Jones polynomial as the end points of an open curve approach each other. Barkataki and Panagiotou [25] further refined this by introducing the Jones polynomial for collections of curves, averaged in all projection directions. Building on these topological frameworks, Panagiotou and Kauffman [26] also used Vassiliev invariants to quantify the complexity of open and closed curves in 3-space. Furthermore, Wang and Panagiotou [27] explored the correlations between protein folding rates and topological measures, specifically writhe, average crossing number (ACN), and the second Vassiliev invariant, to understand the behavior of native protein states. In addition, Herschberg, Pifer, and Panagiotou [28] developed a computational tool

that quantifies topological complexity in systems such as polymers, proteins, and periodic structures.

Using the Jones polynomial framework for the collections of disjoint open or closed curves proposed by Barkataki and Panagiotou in [25], this manuscript introduces two novel models: The multi-scale Jones polynomial, represented by a characteristic matrix, as described in Section 3.1, and the persistent Jones polynomial, represented by weighted persistent barcodes or weighted persistent diagrams, as described in Section 3.2. The weighted barcode was introduced by Cang and Wei in [29]. Both models have the ability to capture local and global entanglement properties in open or closed curve structures in the 3-space, thus, effectively representing their topological characteristics. These models provide an improved approach to solving problems in physical, biological, and chemical environments.

The stability of the proposed models is a key feature that ensures their robust applicability to real-world scenarios. Stability in this context refers to the robustness of persistent Jones polynomial models, including the multi-scale process, against small perturbations in the data. Minor changes in the positions or configurations of the curve segments should result in correspondingly minor variations in the computed measures. This property is essential for reliable analysis in noisy environments or datasets subject to slight distortions, such as those often encountered in physical, biological, and chemical systems. The models provide a reliable framework for characterizing the topological and geometric properties of complex structures.

The proposed models are applied to the prediction of B factors and the analysis of protein  $\alpha$ -helix and  $\beta$ -sheet structures. The B factor, or Debye-Waller factor, is a critical metric in structural biology that represents the atomic displacement and flexibility within a protein structure and, thus, serves as an indicator of protein dynamics and stability. Traditional methods for predicting B factors have had limitations in capturing the topological information inherent in protein structures. To address this, we apply the multi-scale Jones polynomial model to the prediction of B factors, and achieve prediction accuracies of 0.899, 0.808, and 0.720 for small, medium, and large protein sets [30], respectively. Our results on these three datasets outperformed previous methods. In addition, the persistent Jones polynomial model is utilized to explore the structural properties of protein  $\alpha$ -helix and  $\beta$ -sheet segments, with visual representations provided by barcodes that highlight the entanglement features across these secondary structures. The proposed multi-scale Jones polynomial and persistent Jones polynomial models have potential for curve data analysis (CDA).

The article is organized as follows: In Section 2, the fundamental construction of the Jones polynomial is introduced for the collections of curves in the 3-space. In Section 3, two new models of the Jones polynomial of curves in 3-space are established, namely, the multi-scale Jones polynomial (discussed in Section 3.1) and the persistent Jones polynomial (described in Section 3.2). In Section 4, the stability of these two local models is demonstrated. Section 5 presents applications of the new models, including the prediction of B factors and the exploration of  $\alpha$ -helix and  $\beta$ -sheet structures. In Section 6, a discussion is presented on the selection of segmentation, localization, stability issues, and the comparison with classical persistent homology.

## 2. The Jones polynomial of curves in 3-space

The Jones polynomial [3] is an important invariant in classical knot theory, recognized for its ability to characterize the entanglement properties of knots and links. However, it is less applicable to practical

scenarios, which often involve open curves in 3-space rather than closed curves. To address this limitation, Barkataki and Panagiotou [25] extended the concept to the collections of disjoint open or closed curves by defining a normalized version of the bracket polynomial, averaged over all projection directions. This adaptation increases its relevance to real-world applications. Moreover, the Jones polynomial for collections of disjoint open or closed curves converges to the classical Jones polynomial as the endpoints of these curves approach each other. When projected onto a 2-dimensional plane, a collection of curves in 3-space forms a linkoid, which can be generalized to multi-component knotoids describing open-ended knot diagrams. The theory of knotoids was first introduced by Turaev [31], with further developments on linkoids elaborated in [32–35].

### 2.1. Segment cycles

Before defining the bracket polynomial of linkoids, it is essential to introduce the concept of segment cycles associated with a state. Let  $L$  be a linkoid diagram consisting of multiple components. Let  $G = \{1, 2, 3, \dots, 2n\}$  denote the set of all endpoints (heads and legs) of  $L$ . A component of a linkoid with  $n$  components is represented as  $l_{2j-1, 2j}$ , where  $j \in \{1, 2, \dots, n\}$ . The head-leg pairing forms a product of  $n$  disjoint 2-cycles, denoted by  $\hat{L} = (1, 2)(3, 4) \dots (2n-1, 2n)$ .

Let  $S$  be a state corresponding to a choice of smoothing over all crossing points in  $L$ . This state induces a pairing represented by the product of  $n$  disjoint 2-cycles,

$$\hat{S} = (s_1, s_2)(s_3, s_4) \dots (s_{2n-1}, s_{2n}),$$

where each  $s_i \in G$  and each pair  $(s_{2j-1}, s_{2j})$  for  $j \in \{1, 2, \dots, n\}$  represent the endpoints of a component in the state  $S$ .

For any endpoint  $a \in G$ , the set

$$\text{Orb}_S(a) = \{x \in G \mid x = (\hat{L} \circ \hat{S})^m(a), m \in \mathbb{Z}\}$$

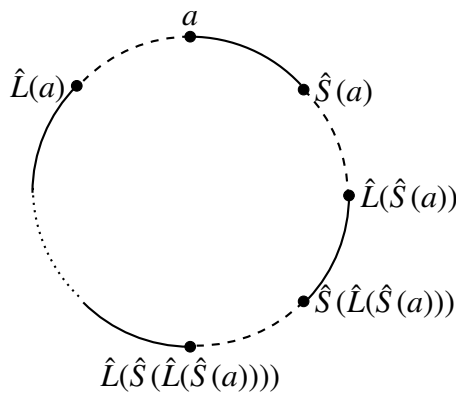
is defined as the *orbit* under the composition function  $\hat{L} \circ \hat{S}$ . The *segment cycle* of an endpoint  $a \in G$  is given by

$$\text{Seg}(a) = \text{Orb}_S(a) \sqcup \text{Orb}_S(\hat{L}(a)).$$

It is notable that for any point  $a \in G$ ,  $\hat{L}(a)$  also belongs to the same segment cycle. Thus, a segment cycle always contains an even number of elements.

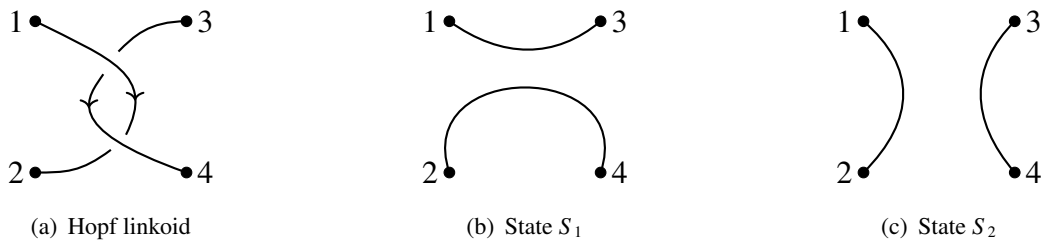
**Lemma 2.1.** [25, Proposition 3.1] *The number of segment cycles in a state  $S$ , denoted by  $|S|_{\text{cyc}}$ , satisfies  $1 \leq |S|_{\text{cyc}} \leq n$ .*

Consider a state  $S$  of  $L$  with the associated pairing  $\hat{S}$ . Let  $\text{Seg}(a)$  be a segment cycle in  $S$  with  $|\text{Seg}(a)| = 2k$ . This segment cycle can be represented by a circle marked with the  $2k$  endpoints of  $L$  (see Figure 1). Let  $a \in G$  be the starting point of the circle. The remaining  $2k-1$  endpoints are uniquely sequenced in the circle as  $\hat{S}(a)$ ,  $\hat{L}(\hat{S}(a))$ ,  $\hat{S}(\hat{L}(\hat{S}(a)))$ , etc., up to  $\hat{L}(a)$ . It is important to note that the arcs connecting adjacent points in this circular representation alternate between the functions  $\hat{S}$  and  $\hat{L}$ . The points connected by  $\hat{S}$  belong to the same component in state  $S$ , while the points connected by  $\hat{L}$  belong to the same component in  $L$ .

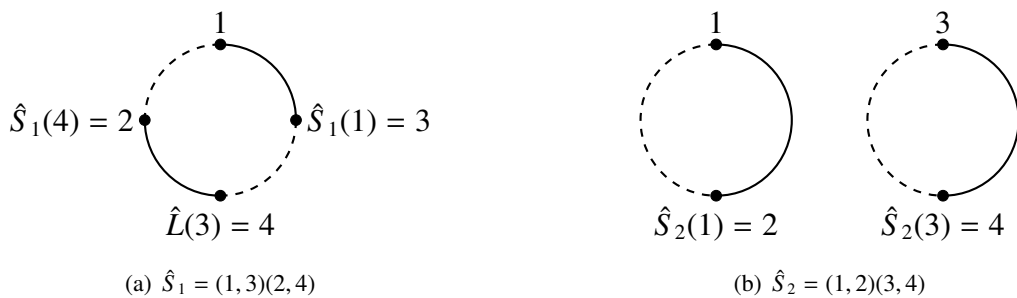


**Figure 1.** Representation of the segment cycle of  $a \in G$ .

**Example 1.** Consider the linkoid diagram  $L$  as shown in Figure 2(a). The set  $G$  of all endpoints is  $\{1, 2, 3, 4\}$ . There are two states of  $L$ ,  $S_1$ , and  $S_2$ , as shown in Figures 2(b) and (c). The associated pairings  $\hat{S}_1$  and  $\hat{S}_2$  are represented by the permutations  $(1, 3)(2, 4)$  and  $(1, 2)(3, 4)$ , respectively. The segment cycles of states  $S_1$  and  $S_2$  are shown in Figure 3. Then,  $|S_1|_{\text{cyc}} = 1$  and  $|S_2|_{\text{cyc}} = 2$ .



**Figure 2.** Hopf linkoid and its states.



**Figure 3.** Segment cycles of two states.

### 2.2. Jones polynomial

The bracket polynomial of linkoids in  $S^2$  or  $\mathbb{R}^2$  is defined through an extension of the bracket polynomial of links. The following initial conditions and diagrammatic relations are sufficient for the skein computation of the bracket polynomial of linkoids.

**Definition 2.1.** Let  $L$  be a linkoid diagram with  $n$  components. The bracket polynomial of the linkoid is uniquely determined by the following skein relation and initial conditions:

$$\langle L \cup \bigcirc \rangle = (-A^2 - A^{-2})\langle L \rangle, \quad \langle \diagdown \diagup \rangle = A \langle \text{smoothing} \rangle + A^{-1} \langle \text{other smoothing} \rangle, \quad \langle || \dots || \rangle = (-A^2 - A^{-2})^{|\text{cyc}|},$$

where  $|\text{cyc}|$  denotes the number of distinct segment cycles.

The bracket polynomial of  $L$  can be expressed as the following state sum expression:

$$\langle L \rangle = \sum_S A^{\sigma(S)} d^{|S|_{\text{circ}}-1} d^{|S|_{\text{cyc}}},$$

where  $S$  is a state corresponding to a choice of smoothing over all crossing points in  $L$ ;  $\sigma(S)$  is the algebraic sum of the smoothing labels of  $S$ ;  $|S|_{\text{circ}}$  is the number of disjoint circles in  $S$ ,  $|S|_{\text{cyc}}$  is the number of distinct segment cycles in  $S$ ; and  $d = (-A^2 - A^{-2})$ .

The *normalized bracket polynomial* is defined as follows:

$$f_L = (-A^{-3})^{-Wr(L)} \langle L \rangle,$$

where  $Wr(L)$  is the writhe of the linkoid diagram  $L$ .

Now, consider curves in 3-space. A regular projection of curves fixed in 3-space can result in different linkoid diagrams depending on the projection direction chosen. Barkataki and Panagiotou in [25] define the bracket polynomial of curves in 3-space as the average of the bracket polynomial of a projection of the curve over all possible projection directions. This definition is made precise as follows:

**Definition 2.2.** [25, Definition 4.1.] Let  $L$  be a collection of disjoint open or closed curves in the 3-space. Let  $(L)_\xi$  denote the projection of  $L$  on a plane with normal vector  $\xi$ . The *normalized bracket polynomial of  $L$*  is defined as follows:

$$f_L = \frac{1}{4\pi} \int_{\xi \in S^2} (-A^3)^{-Wr((L)_\xi)} \langle (L)_\xi \rangle dS,$$

where each  $(L)_\xi$  is a linkoid diagram, and its bracket polynomial can be calculated using Definition 2.1. Note that the integral is taken over all vectors  $\xi \in S^2$ , excluding a set of measure zeros (corresponding to the irregular projections). This gives the Jones polynomial of a collection of disjoint open or closed curves in the 3-space with substitution  $A = t^{-\frac{1}{4}}$ .

**Proposition 2.1.** [25, Proposition 4.1.]

- For open curves, the Jones polynomial has real coefficients and is a continuous function of the curve coordinates.
- As the endpoints of the open curves tend to coincide in 3-space, the Jones polynomial tends to that of the corresponding link.

### 3. Localized Jones polynomials

The Jones polynomial of a collection of disjoint open or closed curves in 3-space describes the entanglement of the curves within the entire collection. However, in many applications, it is desirable to extract the local structural information of the curves. Two methods for localizing the Jones polynomial are proposed to capture the entanglement of a collection of curves and to meet the needs of practical applications.

3.1. Multi-scale Jones polynomial

Let  $L$  be a collection of disjoint open or closed curves in the 3-space. Given a segmentation  $P_n = \{l_1, l_2, \dots, l_n\}$  of  $L$ , where each  $l_i$  represents a finite curve segment of  $L$ , the segments  $l_i$ ,  $1 \leq i \leq n$  can be connected sequentially to reconstruct  $L$ .

To investigate the entanglement properties between each curve segment and its neighboring segments, multi-scale analysis is a suitable approach. Multi-scale analysis requires the definition of a distance metric between the curve segments. The distance between segments, denoted  $d(l_i, l_j)$ , can be specified by different metrics, depending on the application context. In this study, for simplicity, we define the distance  $d(l_i, l_j)$  as the upper bound of the Eulerian distances between a point of one curve segment and another curve segment.

For any curve segment  $l_i$ , consider the set of segments within  $P_n$  whose distances from  $l_i$  fall within the range  $[r, R)$ , where  $r \leq d(l_i, l_j) < R$ . This set, which includes  $l_i$  itself, is denoted as

$$P_{r,R}^i = \{l_j \in P_n \mid r \leq d(l_i, l_j) < R\} \cup \{l_i\}.$$

The Jones polynomial of the set of curve segments  $P_{r,R}^i$ , denoted by  $JP_{r,R}^i$ , quantifies the entanglement of the curve segment  $l_i$  with other segments as  $r$  and  $R$  vary. By selecting two sets of distance parameters,  $\{r_1, r_2, \dots, r_m\}$  and  $\{R_1, R_2, \dots, R_m\}$ , with  $r_i < R_i$  for  $1 \leq i \leq m$ , we obtain a set of characteristic polynomials for  $l_i$ :

$$\{JP_{r_1,R_1}^i, JP_{r_2,R_2}^i, \dots, JP_{r_m,R_m}^i\}.$$

Applying this procedure to all curve segments in  $P_n$ , we obtain an  $n \times m$  matrix:

$$\begin{pmatrix} JP_{r_1,R_1}^1 & JP_{r_2,R_2}^1 & \cdots & JP_{r_m,R_m}^1 \\ JP_{r_1,R_1}^2 & JP_{r_2,R_2}^2 & \cdots & JP_{r_m,R_m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ JP_{r_1,R_1}^n & JP_{r_2,R_2}^n & \cdots & JP_{r_m,R_m}^n \end{pmatrix},$$

capturing both local and global entanglement properties for the collection of disjoint open or closed curves  $L$ .

Each matrix entry is a polynomial. For practical applications, the Jones polynomial can be evaluated at a specific parametrization, such as  $t = 10$ , resulting in an  $n \times m$  characteristic matrix for the segmentation  $P_n$  of  $L$ :

$$mJ(P_n) = \begin{pmatrix} JP_{r_1,R_1}^1(10) & JP_{r_2,R_2}^1(10) & \cdots & JP_{r_m,R_m}^1(10) \\ JP_{r_1,R_1}^2(10) & JP_{r_2,R_2}^2(10) & \cdots & JP_{r_m,R_m}^2(10) \\ \vdots & \vdots & \ddots & \vdots \\ JP_{r_1,R_1}^n(10) & JP_{r_2,R_2}^n(10) & \cdots & JP_{r_m,R_m}^n(10) \end{pmatrix},$$

where each entry is a real number.

**Remark 3.1.** For specific application contexts involving research objects that can be represented as a collection of disjoint open or closed curves  $L$ , the choice of segmentation  $P_n = \{l_1, l_2, \dots, l_n\}$  plays a critical role. The choice of an appropriate segmentation, tailored to the requirements of the application,

allows for a more precise capture of the entanglement properties inherent to the research objects, and, thus, a more accurate reflection of their characteristics. Similarly, the choice of parameters  $\{r_1, r_2, \dots, r_m\}$  and  $\{R_1, R_2, \dots, R_m\}$  influences the robustness and precision of the final assessment of the entanglement features of the objects.

### 3.2. Persistent Jones polynomial

To address different application scenarios and to better capture the information about entanglement of curves in the 3-space, a second localization of the Jones polynomial has been proposed. This adaptation of the Jones polynomial for a collection of disjoint open or closed curves, aimed at quantifying the complexity of entanglement, serves as an extension of the classical Jones polynomial [25].

Let  $L$  be a collection of disjoint open or closed curves with a segmentation denoted by  $P_n = \{l_1, l_2, \dots, l_n\}$ , where  $l_i$  is a segment of  $L$ . The segments  $l_i$ ,  $1 \leq i \leq n$ , can be connected sequentially to reconstruct  $L$ . To effectively represent these multiple segments, both the Čech complex and the Vietoris-Rips complex, constructed here from the distance matrix,

$$d(P_n) = \begin{pmatrix} 0 & d(l_1, l_2) & \cdots & d(l_1, l_n) \\ d(l_2, l_1) & 0 & \cdots & d(l_2, l_n) \\ \vdots & \vdots & \ddots & \vdots \\ d(l_n, l_1) & d(l_n, l_2) & \cdots & 0 \end{pmatrix}$$

from  $P_n = \{l_1, l_2, \dots, l_n\}$ , are suitable methods. Given the similarity between the Čech complex and the Vietoris-Rips complex, we will focus on the Vietoris-Rips complex in the following discussion. Let  $r$  denote the variable parameter of the Vietoris-Rips complex.

**Definition 3.1.** *A critical value of the Vietoris-Rips complex is a real number  $r$  such that, for any sufficiently small  $\varepsilon > 0$ , the map  $K_{r-\varepsilon} \hookrightarrow K_{r+\varepsilon}$  is an inclusion but not an isomorphism, where  $K_r$  denotes the complex at the Vietoris-Rips parameter  $r$ .*

**Remark 3.2.** *For a segmentation  $P_n$  of  $L$  into finite curve segments, the Vietoris-Rips complex has a finite number of critical values.*

Let  $r_0 < r_1 < r_2 < \dots < r_m$  represent the critical values of the Vietoris-Rips complex for a segmentation  $P_n$  of  $L$ . This generates a sequence of complexes from  $P_n$  that form a filtration  $\mathcal{F}(P_n)$ :

$$K_{r_0} \subsetneq K_{r_1} \subsetneq K_{r_2} \subsetneq \cdots \subsetneq K_{r_m},$$

where the final complex  $K_{r_m}$  is an  $(n-1)$ -simplex. For any  $x < y$ , let  $V_x^y : K_x \hookrightarrow K_y$  denote the inclusion map.

**Lemma 3.1.** [36, Critical Value Lemma] *If a closed interval  $[x, y]$  does not contain a critical value of the Vietoris-Rips complex, then  $V_x^y : K_x \rightarrow K_y$  is an isomorphism.*

Within the filtration  $\mathcal{F}(P_n)$ , consider a complex  $K_r$ . Each vertex  $v_a \in K_r$  corresponds to a segment  $l_a$  of the curve segments in  $P_n$ . An edge  $\{v_a, v_b\} \in K_r$  indicates that the distance between segments  $l_a$  and  $l_b$  is less than the Vietoris-Rips parameter  $r$ . A simplex  $\Delta = \{v_a, v_b, \dots, v_t\} \in K_r$  means that the pair-wise distances among the corresponding segments  $\{l_a, l_b, \dots, l_t\}$ , collectively denoted by  $\Delta(P_n)$ , are all less than  $r$ .



**Definition 3.2.** Let  $K$  be a simplicial complex. The maximal faces of  $K$  with respect to inclusion are named the facets of  $K$ . The simplicial complex  $K$ , characterized by the facets  $F_1, \dots, F_q$ , is denoted by

$$K = \langle F_1, \dots, F_q \rangle,$$

and the set  $\{F_1, \dots, F_q\}$  is called the set of facets of  $K$ .

Each complex within the filtration  $\mathcal{F}(P_n)$  can be described by its facets. Thus, the filtration  $\mathcal{F}(P_n)$  can be expressed as:

$$\langle F_1^{r_0}, \dots, F_{q_0}^{r_0} \rangle \subsetneq \langle F_1^{r_1}, \dots, F_{q_1}^{r_1} \rangle \subsetneq \dots \subsetneq \langle F_1^{r_m}, \dots, F_{q_m}^{r_m} \rangle.$$

**Definition 3.3.** The birth of a facet  $F$  in the filtration  $\mathcal{F}(P_n)$  is the smallest index  $r_b$  such that  $F$  appears as a facet in the complex  $K_{r_b}$  but not in  $K_{r_b-\varepsilon}$  for any sufficiently small  $\varepsilon > 0$ .

The death of a facet  $F$  in the filtration  $\mathcal{F}(P_n)$  is the largest index  $r_d$  such that  $F$  is a facet in  $K_{r_d}$  but not in  $K_{r_d+\varepsilon}$  for any sufficiently small  $\varepsilon > 0$ .

The life-span of a facet  $F$  in the filtration  $\mathcal{F}(P_n)$  is the interval  $[r_b, r_d]$ .

Therefore,  $\mathcal{F}(P_n)$  can be represented as a sequence of facets, each associated with a birth-and-death interval,

$$\{(F_1, [r_b^1, r_d^1]), (F_2, [r_b^2, r_d^2]), \dots, (F_t, [r_b^t, r_d^t])\}.$$

Similar to persistent barcodes in persistent homology, a barcode can represent the facets within a filtration. For any given dimension, each bar corresponds to a facet, with the start and end points of the bar indicating the birth and death of the associated facet, respectively.

**Definition 3.4.** The barcode  $B(P_n)$  for the filtration of facets  $\mathcal{F}(P_n)$  consists of horizontal line segments  $[r_i, r_j]$ , where  $r_i \leq r_j$ , representing the birth and death times of the associated facet.

A barcode provides a visual representation of a filtration as a collection of horizontal line segments on a plane, where the horizontal axis corresponds to the parameter, and the vertical axis represents an ordering of the facets.

**Definition 3.5.** For each facet  $F = \{v_a, v_b, \dots, v_t\}$  in the filtration  $\mathcal{F}(P_n)$ , there exists a corresponding subset of segments  $F(P_n) = \{l_a, l_b, \dots, l_t\}$  in the segmentation  $P_n$ . The Jones polynomial of  $F(P_n)$ , denoted by  $JF(P_n)$ , is defined as the weight of the facet  $F$ . Consequently, the Jones polynomial can be treated as a weighting function for filtration  $\mathcal{F}(P_n)$ . The resulting weighted filtration is denoted by  $J\mathcal{F}(P_n)$ ,

$$\{(F_1, [r_b^1, r_d^1], JF_1(P_n)), (F_2, [r_b^2, r_d^2], JF_2(P_n)), \dots, (F_t, [r_b^t, r_d^t], JF_t(P_n))\}$$

referred to as the persistent Jones polynomial of the segmentation  $P_n$  of  $L$ .

Since filtration  $\mathcal{F}(P_n)$  can be expressed by a barcode of facets, the persistent Jones polynomial of the segmentation  $P_n$  of  $L$  can also be expressed by a barcode, with the Jones polynomials of the associated facets as weights. The weights in the persistent Jones polynomial of  $P_n$  are polynomials. To improve applicability in specific scenarios, setting the Jones polynomial variable  $t = 10$  converts these weights to real numbers, producing a real-number weighted barcode  $BJ(P_n)(10)$ .

**Remark 3.3.** *The persistent Jones polynomial and the classical persistent homology perform a multi-scale analysis of the data through filtration using Vietoris-Rips complexes or other types of complexes. However, they focus on different aspects of data analysis. The persistent Jones polynomial uses tools from geometric topology to capture the entanglement complexity of curves in 3-space, whereas classical persistent homology uses methods from algebraic topology to investigate the behavior of connected components, one-dimensional loops, two-dimensional cavities, and high-dimensional cavities (i.e., generators of homology groups) within a point cloud data.*

#### 4. Stability

The stability of a model is characterized by the property that small perturbations in the collection of disjoint open or closed curves  $L$  result in only minor variations in the localized measures of the multi-scale Jones polynomial and the persistent Jones polynomial.

Let  $f : L \rightarrow f(L)$  be a continuous function acting on a collection of disjoint open or closed curves  $L$  in the 3-space. The difference between  $f(L)$  and  $L$  is measured using the supremum norm  $\|f(L) - L\|_\infty$ , defined as:

$$\|f(L) - L\|_\infty = \sup_{x \in L} |f(x) - x|.$$

For a given segmentation  $P_n = \{l_1, l_2, \dots, l_n\}$  of  $L$ , where each  $l_i$  ( $1 \leq i \leq n$ ) is a finite curve segment and  $L$  can be reconstructed by connecting these segments end-to-end, consider a continuous mapping  $f : L \rightarrow f(L)$  such that  $\|f(L) - L\|_\infty < \varepsilon$  for a sufficiently small  $\varepsilon > 0$ . This induces a corresponding segmentation  $f(P_n) = \{f(l_1), f(l_2), \dots, f(l_n)\}$  of  $f(L)$ .

The characteristic matrices of the multi-scale Jones polynomial for  $P_n$  and  $f(P_n)$ , denoted  $mJ(P_n)$  and  $mJ(f(P_n))$ , respectively, exhibit only minor differences at the corresponding positions. Furthermore, the weighted Bottleneck distance between the weighted persistence diagrams of persistent Jones polynomials for  $P_n$  and  $f(P_n)$  is also minimal.

**Remark 4.1.** *The Jones polynomial evaluated at  $t = 10$  can be viewed as a function on collections of curves in 3-space. Let  $L$  be a collection of curves in  $\mathbb{R}^3$ . According to Proposition 2.1, let  $f : L \rightarrow f(L)$  be a continuous function. If  $\|f(L) - L\|_\infty < \varepsilon$  for all sufficiently small  $\varepsilon > 0$ , then  $|J(L)(10) - J(f(L))(10)| < \varepsilon_J$  for some sufficiently small  $\varepsilon_J > 0$ .*

##### 4.1. Stability of multi-scale Jones polynomial

Let  $L$  be a collection of disjoint open or closed curves in the 3-space, and let  $P_n = \{l_1, l_2, \dots, l_n\}$  denote a segmentation of  $L$  into  $n$  segments. Consider a continuous function  $f : L \rightarrow f(L)$ , which induces a corresponding segmentation of  $f(L)$ , represented by  $f(P_n) = \{f(l_1), f(l_2), \dots, f(l_n)\}$ .

**Proposition 4.1.** *Suppose  $f : L \rightarrow f(L)$  is a continuous function such that  $\|f(L) - L\|_\infty < \varepsilon$  for all sufficiently small  $\varepsilon > 0$ . Then, the two sets of curve segments  $f(P_{r,R}^i)$  and  $f(P_{r,R}^i)$  are equal:*

$$f(P_{r,R}^i) = f(P_{r,R}^i),$$

where:

- $f(P_{r,R}^i)$  is the image of the set  $P_{r,R}^i$  under the function  $f : L \rightarrow f(L)$ ;

- $f(P)_{r,R}^i$  represents the set of curve segments in the segmentation  $f(P_n)$  such that their distance from  $f(l_i)$  is within  $[r, R)$ , including of  $f(l_i)$  itself.

*Proof.* By definition, we have

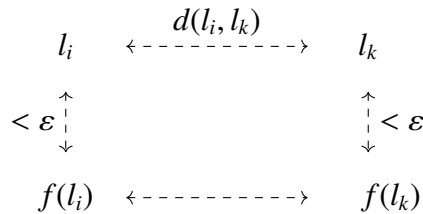
$$P_{r,R}^i = \{l_j \in P_n \mid r \leq d(l_i, l_j) < R\} \cup \{l_i\},$$

$$f(P_{r,R}^i) = \{f(l_j) \in P_n \mid r \leq d(l_i, l_j) < R\} \cup \{f(l_i)\},$$

and

$$f(P)_{r,R}^i = \{f(l_j) \in f(P_n) \mid r \leq d(f(l_i), f(l_j)) < R\} \cup \{f(l_i)\},$$

for any  $l_k \in P_{r,R}^i$ , and it holds that  $r \leq d(l_i, l_k) < R$ . Since  $\|f(L) - L\|_\infty < \varepsilon$ , the difference between each segment of the curve and its image under  $f$  is less than  $\varepsilon$ :  $\|f(l_i) - l_i\|_\infty < \varepsilon$  and  $\|f(l_k) - l_k\|_\infty < \varepsilon$ , and, thus, the distances satisfy  $d(f(l_i), l_i) < \varepsilon$  and  $d(f(l_k), l_k) < \varepsilon$ .



Then, we have:

$$d(l_i, l_k) - 2\varepsilon < d(f(l_i), f(l_k)) < d(l_i, l_k) + 2\varepsilon.$$

Therefore,

$$r - 2\varepsilon < d(f(l_i), f(l_k)) < R + 2\varepsilon,$$

which implies  $f(l_k) \in f(P)_{r-2\varepsilon, R+2\varepsilon}^i$ . Thus,  $f(P_{r,R}^i) \subseteq f(P)_{r-2\varepsilon, R+2\varepsilon}^i$ .

Similarly, we can show that:

$$f(P)_{r-2\varepsilon, R+2\varepsilon}^i \subseteq f(P_{r-4\varepsilon, R+4\varepsilon}^i).$$

Consequently, we obtain:

$$f(P_{r,R}^i) \subseteq f(P)_{r-2\varepsilon, R+2\varepsilon}^i \subseteq f(P_{r-4\varepsilon, R+4\varepsilon}^i).$$

Since  $\varepsilon > 0$  is sufficiently small, we conclude that  $P_{r-4\varepsilon, R+4\varepsilon}^i = P_{r,R}^i$ ,  $f(P)_{r-2\varepsilon, R+2\varepsilon}^i = f(P)_{r,R}^i$ , and  $f(P_{r,R}^i) = f(P)_{r,R}^i$ . □

**Theorem 4.1.** Suppose  $f : L \rightarrow f(L)$  is a continuous function such that  $\|f(L) - L\|_\infty < \varepsilon$  for all sufficiently small  $\varepsilon > 0$ . Consider two sets of distances  $\{r_1, r_2, \dots, r_m\}$  and  $\{R_1, R_2, \dots, R_m\}$ . There exist two characteristic matrices for the segmentation  $P_n$  of  $L$  and for the segmentation  $f(P_n)$  of  $f(L)$ , given by

$$mJ(P_n) = \begin{pmatrix} JP_{r_1, R_1}^1(10) & JP_{r_2, R_2}^1(10) & \cdots & JP_{r_m, R_m}^1(10) \\ JP_{r_1, R_1}^2(10) & JP_{r_2, R_2}^2(10) & \cdots & JP_{r_m, R_m}^2(10) \\ \vdots & \vdots & \ddots & \vdots \\ JP_{r_1, R_1}^n(10) & JP_{r_2, R_2}^n(10) & \cdots & JP_{r_m, R_m}^n(10) \end{pmatrix},$$

$$mJ(f(P_n)) = \begin{pmatrix} Jf(P)_{r_1, R_1}^1(10) & Jf(P)_{r_2, R_2}^1(10) & \cdots & Jf(P)_{r_m, R_m}^1(10) \\ Jf(P)_{r_1, R_1}^2(10) & Jf(P)_{r_2, R_2}^2(10) & \cdots & Jf(P)_{r_m, R_m}^2(10) \\ \vdots & \vdots & \ddots & \vdots \\ Jf(P)_{r_1, R_1}^n(10) & Jf(P)_{r_2, R_2}^n(10) & \cdots & Jf(P)_{r_m, R_m}^n(10) \end{pmatrix}.$$

Then, the difference between the corresponding entries in these two matrices is less than  $\varepsilon_J$ , where  $\varepsilon_J$  is sufficiently small.

*Proof.* We have

$$\|P_{r_j, R_j}^i - f(P_{r_j, R_j}^i)\|_\infty \leq \|L - f(L)\|_\infty < \varepsilon.$$

From Remark 4.1, it follows that

$$|JP_{r_j, R_j}^i(10) - Jf(P_{r_j, R_j}^i)(10)| < \varepsilon_J,$$

where  $\varepsilon_J$  is sufficiently small. By Proposition 4.1, we know that  $f(P_{r, R}^i) = f(P)_{r, R}^i$ . Therefore,

$$|JP_{r_j, R_j}^i(10) - Jf(P)_{r_j, R_j}^i(10)| < \varepsilon_J,$$

where  $\varepsilon_J$  is sufficiently small. □

**Remark 4.2.** *The stability of the method used in [20] can be proved in a manner similar to that of Theorem 4.1. Suppose  $f : L \rightarrow f(L)$  is a continuous function such that  $\|f(L) - L\|_\infty < \varepsilon$  for all sufficiently small  $\varepsilon > 0$ .*

For the Gauss linking integral, there exist two Gauss linking integral matrices for the segmentation  $P_n = \{l_1, l_2, \dots, l_n\}$  of a collection of disjoint open or closed curves  $L$  in 3-space, as well as for  $f(P_n)$  of  $f(L)$ . These matrices are given by:

$$GL(P_n) = \begin{pmatrix} g(l_1, l_1) & g(l_1, l_2) & \cdots & g(l_1, l_n) \\ g(l_2, l_1) & g(l_2, l_2) & \cdots & g(l_2, l_n) \\ \vdots & \vdots & \ddots & \vdots \\ g(l_n, l_1) & g(l_n, l_2) & \cdots & g(l_n, l_n) \end{pmatrix},$$

$$GL(f(P_n)) = \begin{pmatrix} g(f(l_1), f(l_1)) & g(f(l_1), f(l_2)) & \cdots & g(f(l_1), f(l_n)) \\ g(f(l_2), f(l_1)) & g(f(l_2), f(l_2)) & \cdots & g(f(l_2), f(l_n)) \\ \vdots & \vdots & \ddots & \vdots \\ g(f(l_n), f(l_1)) & g(f(l_n), f(l_2)) & \cdots & g(f(l_n), f(l_n)) \end{pmatrix},$$

where

$$g(l_i, l_j) = \begin{cases} GL(l_i, l_j), & \text{if } l_i \cap l_j = \emptyset; \\ 0, & \text{otherwise.} \end{cases}$$

Here,  $GL(l_i, l_j)$  denotes the Gauss linking integral of the curve segments  $l_i$  and  $l_j$ .

As stated in [24, page 3], we can treat the Gauss linking integral of a curve as a continuous function of the coordinates of the curve. Similar to Theorem 4.1, we also have

$$|g(l_i, l_j) - g(f(l_i), f(l_j))| < \varepsilon_{GL}$$

for sufficiently small  $\varepsilon_{GL}$ .

4.2. Stability of the persistence Jones polynomial

In this section, we state and prove the stability of the persistent Jones polynomial, which asserts that small changes in the collection of disjoint open or closed curves  $L$  lead to only small changes in the persistent Jones polynomial. As discussed in Section 3.2, the persistent Jones polynomial is represented by a weighted Jones polynomial filtration, which can be expressed as a Jones polynomial weight barcode of facets. For a given barcode, there is a corresponding diagram where each bar in the barcode can be mapped to a point in the diagram. The  $x$ -coordinate of this point represents the birth time of the corresponding bar, while the  $y$ -coordinate represents its death time.

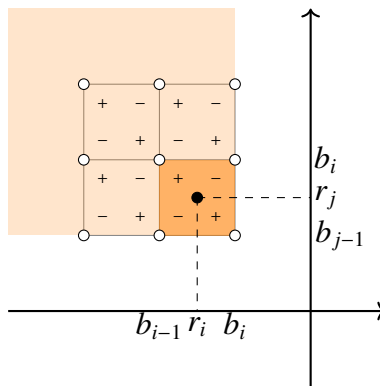
Let  $L$  be a collection of disjoint open or closed curves in the 3-space, and let  $P_n = \{l_1, l_2, \dots, l_n\}$  denote a segmentation of  $L$  into  $n$  segments. Consider a continuous function  $f : L \rightarrow f(L)$ , which induces a corresponding segmentation of  $f(L)$ , represented by  $f(P_n) = \{f(l_1), f(l_2), \dots, f(l_n)\}$ . Let  $r_1 < r_2 < \dots < r_m$  be the critical values of the Vietoris-Rips complex of  $P_n$ . We denote an interleaved sequence  $(b_i)_{i=0,1,\dots,m}$  such that  $b_{i-1} < r_i < b_i$  for all  $i$ . We set  $b_{-1} = r_0 = -\infty$  and  $b_{m+1} = r_{m+1} = +\infty$ .

For two integers  $0 \leq i < j \leq m + 1$  and a fixed integer  $k$ , we define the *multiplicity* of the pair  $(r_i, r_j)$  as

$$\mu_i^j = \beta_{b_{i-1}}^{b_j} - \beta_{b_i}^{b_j} + \beta_{b_{i-1}}^{b_{j-1}} - \beta_{b_i}^{b_{j-1}},$$

where  $\beta_x^y$  is the number of  $k$ -facets contained in  $K_x$  that remain in  $K_y$  for all  $-\infty \leq x \leq y \leq +\infty$ . To visualize this definition, consider  $\beta_x^y$  as the value of a function  $\beta$  at the point  $(r_i, r_j) \in \bar{\mathbb{R}}^2$ , where  $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ . Thus,  $\mu_i^j$  is the alternating sum of  $\beta$  in the corners of the box  $[b_{i-1}, b_i] \times [b_{j-1}, b_j]$ , as depicted in Figure 4.

Note that if  $x$  and  $x'$  are in the open interval  $(r_i, r_{i+1})$ , and  $y$  and  $y'$  are in  $(r_{j-1}, r_j)$ , then  $\beta_x^y = \beta_{x'}^{y'}$ . Therefore, the multiplicities  $\mu_i^j$  are well-defined and always nonnegative.



**Figure 4.** The multiplicity of the point  $(r_i, r_j)$  is the alternating sum at the corners of the lower right square. When other multiplicities are added, cancellations between plus and minus signs occur.

**Definition 4.1.** The diagram  $D(P_n) \subset \bar{\mathbb{R}}^2$  of the persistent Jones polynomial for  $P_n$  consists of points  $(r_i, r_j)$  with Jones polynomial weights, counted with multiplicity  $\mu_i^j$  for  $0 \leq i < j \leq m + 1$ , along with all points on the diagonal, which are counted with infinite multiplicity.

Each off-diagonal point in the diagram represents the lifespan of a  $k$ -facet in the filtration. Similar to the weighted persistent barcode, the Jones polynomial corresponding to the set of curve segments

for the  $k$ -facet can be used as the weight. This approach results in a weighted persistent diagram for the persistent Jones polynomial.

#### 4.2.1. Bottleneck distance

The *Bottleneck distance* is a classical measure used to quantify the difference between two persistent diagrams. It naturally extends to the comparison of weighted persistent diagrams, allowing for a precise demonstration of the variations between two persistent Jones polynomials.

To better capture the differences between persistent Jones polynomials, we apply a slight modification to the traditional definition of the Bottleneck distance. Let  $\mathcal{C}$  and  $\mathcal{D}$  be two multi-sets of pairs  $(\langle a, b \rangle, w)$ , where  $\langle a, b \rangle$  denotes an interval that can be any well-defined member of the set  $\{[a, b], [a, b), (a, b], (a, b)\}$ , and  $w \in \mathbb{R}$  represents the weight of the interval  $\langle a, b \rangle$ .

A *matching* between the sets  $\mathcal{C}$  and  $\mathcal{D}$  is defined as a collection of pairs  $\chi = \{(I, J) \in \mathcal{C} \times \mathcal{D}\}$ , where each element  $I \in \mathcal{C}$  and each element  $J \in \mathcal{D}$  appears in at most one pair within  $\chi$ . A matching forms a bijection between a subset of  $\mathcal{C}$  and a subset of  $\mathcal{D}$ . If a pair  $(I, J) \in \chi$ , we say that  $I$  is *matched* with  $J$ . Conversely, if an element  $I$  does not appear in any pair, it is considered *unmatched*.

The *cost*  $c(I, J)$  of the matching elements  $I = (\langle a, b \rangle, w_1)$  and  $J = (\langle c, d \rangle, w_2)$  is defined as follows:

$$c(I, J) = \max \{|c - a|, |d - b|, |w_1 - w_2|\}.$$

Similarly, the cost  $c(I)$  of leaving an unmatched element  $I$  is defined as:

$$c(I) = \frac{b - a}{2}.$$

Finally, the cost of a matching  $\chi$  is given by:

$$c(\chi) = \max \left( \sup_{(I, J) \in \chi} c(I, J), \sup_{\text{unmatched } I \in \mathcal{C} \cup \mathcal{D}} c(I) \right).$$

**Definition 4.2.** *The weighted Bottleneck distance between  $\mathcal{C}$  and  $\mathcal{D}$  is defined as*

$$d_B(\mathcal{C}, \mathcal{D}) = \inf \{c(\chi) \mid \chi \text{ is a matching between } \mathcal{C} \text{ and } \mathcal{D}\}.$$

The modified Bottleneck distance increases the emphasis on weight factors compared to the classical Bottleneck distance.

#### 4.2.2. Stability

Let  $L$  be a collection of disjoint open or closed curves in the 3-space, and let  $P_n = \{l_1, l_2, \dots, l_n\}$  denote a segmentation of  $L$  into  $n$  segments. Consider a continuous function  $f : L \rightarrow f(L)$ , which induces a segmentation of  $f(L)$ , denoted by  $f(P_n) = \{f(l_1), f(l_2), \dots, f(l_n)\}$ .

There are two persistent Jones polynomials based on  $P_n$  and  $f(P_n)$ , represented as  $J\mathcal{F}(P_n)$  and  $J\mathcal{F}(f(P_n))$ , respectively. The facet weights in these persistent Jones polynomials are polynomials. By setting the Jones polynomial variable  $t = 10$ , the weight of each facet is converted to a real number. Thus, the converted persistent Jones polynomials can be denoted by  $J\mathcal{F}(P_n)(10)$  and  $J\mathcal{F}(f(P_n))(10)$ . These can be expressed using weighted persistent diagrams, denoted by  $D(P_n)$  and  $D(f(P_n))$ .

Suppose  $\|f(L) - L\| < \varepsilon$  for all sufficiently small  $\varepsilon > 0$ . Then, for all  $l_{p_i} \in P_n(p)$ , we have  $\|l_{p_i} - f(l_{p_i})\|_\infty < \varepsilon$ . Let  $l_{p_i}, l_{p_j}$  be any two curve segments in  $P_n(p)$ . Then, there exists

$$\begin{array}{ccc}
 & d(l_{p_i}, l_{p_j}) & \\
 l_{p_i} & \leftarrow \text{-----} \rightarrow & l_{p_j} \\
 \uparrow & & \uparrow \\
 < \varepsilon & & < \varepsilon \\
 \downarrow & & \downarrow \\
 f(l_{p_i}) & \leftarrow \text{-----} \rightarrow & f(l_{p_j})
 \end{array}$$

Thus,

$$d(l_{p_i}, l_{p_j}) - 2\varepsilon \leq d(f(l_{p_i}), f(l_{p_j})) \leq d(l_{p_i}, l_{p_j}) + 2\varepsilon.$$

There is an important lemma, proved in [36].

**Lemma 4.1.** [36, Box Lemma] For  $a < b < c < d$ , let  $R = [a, b] \times [c, d]$  be a box in  $\mathbb{R}^2$ , and let  $R_{2\varepsilon} = [a + 2\varepsilon, b - 2\varepsilon] \times [c + 2\varepsilon, d - 2\varepsilon]$  be the box obtained by shrinking  $R$  on all sides by  $2\varepsilon$ . It follows that:

$$\#(D(P_n) \cap R_{2\varepsilon}) \leq \#(D(f(P_n)) \cap R).$$

**Theorem 4.2.** Let  $L$  be a collection of disjoint open or closed curves, and let  $P_n = \{l_1, l_2, \dots, l_n\}$  denote a segmentation of  $L$  into  $n$  segments. Suppose  $f : L \rightarrow f(L)$  is a continuous function such that  $\|f(L) - L\|_\infty < \varepsilon$  for all sufficiently small  $\varepsilon > 0$ . Then, the weighted Bottleneck distance between the weighted persistent diagrams of the persistent Jones polynomials,  $d_B(D(P_n), D(f(P_n)))$ , is sufficiently small.

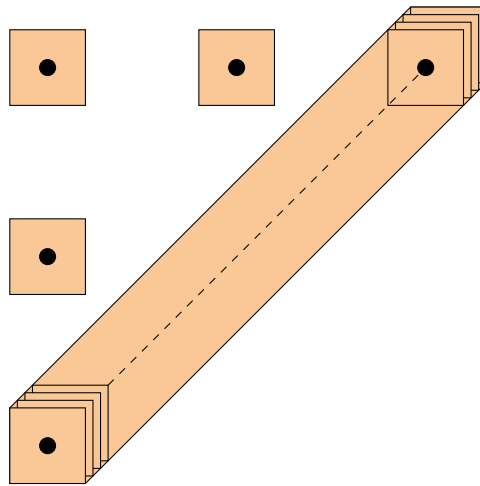
*Proof.* Let  $L$  be a collection of disjoint open or closed curves in 3-space, and let  $P_n = \{l_1, l_2, \dots, l_n\}$  denote a segmentation of  $L$  into  $n$  segments. Consider a continuous function  $f : L \rightarrow f(L)$ , which induces a segmentation of  $f(L)$ , denoted by  $f(P_n) = \{f(l_1), f(l_2), \dots, f(l_n)\}$ . There are two persistent Jones polynomial diagrams for  $P_n$  and  $f(P_n)$ , denoted by  $D(P_n)$  and  $D(f(P_n))$ .

Consider the minimum distance between two distinct off-diagonal points or between an off-diagonal point and the diagonal:

$$\delta_L = \min\{\|p - q\|_\infty \mid p \neq q \in D(P_n) - \Delta\}.$$

Assuming that  $\varepsilon > 0$  is sufficiently small, we take  $\varepsilon < \delta_L/4$ .

By drawing cubes of radius  $2\varepsilon$  around points in  $D(P_n)$ , we obtain a thickened diagonal plane along with a finite set of disjoint cubes that are also disjoint from the thickened diagonal, as shown in Figure 5.



**Figure 5.** The shaded squares are centered at the black points of  $D(P_n)$ .

Let  $\mu$  denote the multiplicity of a point  $p$  in  $D(P_n) \setminus \Delta$ , and let  $\square_{2\varepsilon}$  represent the cube centered at  $p$  with radius  $2\varepsilon$ . According to Lemma 4.1,

$$\mu \leq \#(D(f(P_n)) \cap \square_{2\varepsilon}) \leq \#(D(P_n) \cap \square_{4\varepsilon}).$$

Since  $4\varepsilon < \delta_L$ ,  $p$  is the only point in  $D(P_n)$  within  $\square_{4\varepsilon}$ , which implies that  $\#(D(f(P_n)) \cap \square_{2\varepsilon}) = \mu$ .

Now, let  $p = (x, y)$  be an off-diagonal point in the diagram  $D(P_n)$  with multiplicity  $\mu = 1$ . A corresponding collection of curve segments  $P_n(p) = \{l_{p_1}, l_{p_2}, \dots, l_{p_t}\}$  exists in  $P_n$ . Applying the continuous function  $f$  to these segments yields the transformed collection  $f(P_n(p)) = \{f(l_{p_1}), f(l_{p_2}), \dots, f(l_{p_t})\}$ .

According to the definition of the persistent Jones polynomial, the point  $p = (x, y)$  indicates that  $P_n(p) = \{l_{p_1}, l_{p_2}, \dots, l_{p_t}\}$  forms a facet with birth  $x$  and death  $y$  in the filtration  $\mathcal{F}(P_n)$ . This implies that any two segments in  $P_n(p)$  are at a distance less than  $x$ , and any segment  $l \in P_n \setminus P_n(p)$  is at least a distance  $y$  from all  $l_{p_i} \in P_n(p)$ , for  $1 \leq i \leq t$ .

Since

$$d(l_{p_i}, l_{p_j}) - 2\varepsilon \leq d(f(l_{p_i}), f(l_{p_j})) \leq d(l_{p_i}, l_{p_j}) + 2\varepsilon,$$

then any two curve segments in  $f(P_n(p))$  are within a distance of  $x + 2\varepsilon$ . Furthermore, for  $f(l) \in f(P_n \setminus P_n(p))$  and  $f(l_{p_i}) \in f(P_n(p))$ , the distance is at least  $y - 2\varepsilon$ .

Thus, the segment collection  $f(P_n(p))$  forms a facet within  $\mathcal{F}(f(P_n))$ , corresponding to a point  $f(p)$  in  $D(f(P_n))$ , within the region  $[x, x + 2\varepsilon] \times [y - 2\varepsilon, y]$ , indicating that  $\|f(p) - p\|_\infty < 2\varepsilon$ . Moreover, we have  $\#(D(f(P_n)) \cap \square_{2\varepsilon}) = 1$ ; hence,  $f(p)$  is the only point in  $D(f(P_n)) \cap \square_{2\varepsilon}$ .

For off-diagonal points  $p^1 = p^2 = \dots = p^\mu = (x, y)$  with  $\mu > 1$ , this indicates  $\mu$  distinct facets in  $\mathcal{F}(P_n)$  with the same birth  $x$  and death  $y$ , but distinct curve segment collections  $P_n(p^1), P_n(p^2), \dots, P_n(p^\mu)$ . Analogous to the case when  $\mu = 1$ , the images  $f(p^1), f(p^2), \dots, f(p^\mu)$  lie within  $[x, x + 2\varepsilon] \times [y - 2\varepsilon, y]$ , which gives  $\|f(p^i) - p^i\|_\infty < 2\varepsilon$  for  $1 \leq i \leq \mu$ . Moreover, we have that  $\#(D(f(P_n)) \cap \square_{2\varepsilon}) = \mu$ ; hence,  $f(p^1), f(p^2), \dots, f(p^\mu)$  are the  $\mu$  points in  $D(f(P_n)) \cap \square_{2\varepsilon}$ .

The weights of  $p$  and  $f(p)$  are  $JP_n(p)(10)$  and  $Jf(P_n(p))(10)$ , respectively. Since

$$\|f(P_n(p)) - P_n(p)\|_\infty \leq \|f(L) - L\|_\infty < \varepsilon,$$



by Remark 4.1, the weight difference satisfies  $|Jf(P_n(p))(10) - JP_n(p)(10)| < \varepsilon_J$ , where  $\varepsilon_J > 0$  is sufficiently small.

After examining all the off-diagonal points in  $D(P_n)$ , the only points in  $D(f(P_n))$  that are not images  $f(p)$  for some  $p \in (D(P_n) \setminus \Delta)$  lie beyond  $2\varepsilon$  from  $D(P_n) \setminus \Delta$ .

Let  $q \in D(f(P_n))$  be a point for which there is no corresponding point  $p \in D(P_n)$  such that  $f(p) = q$ . Suppose the distance from  $q$  to  $\Delta$  is greater than  $2\varepsilon$ . Then, there exists a square  $\square_{2\varepsilon}^q$  centered at  $q$  with a radius of  $2\varepsilon$  such that  $D(P_n) \cap \square_{2\varepsilon}^q = \emptyset$ . This contradicts Lemma 4.1, which states that  $1 \leq \#(D(P_n) \cap \square_{2\varepsilon}^q) \neq 0$ . Therefore, the distance from  $q$  to  $\Delta$  must be less than  $2\varepsilon$ .

There exists a natural matching between  $D(P_n)$  and  $D(f(P_n))$ , represented as  $\chi = \{(p, f(p)) \mid p \in D(P_n) \setminus \Delta\}$ , with unmatched points in  $D(f(P_n))$  regarded as not corresponding to any point in  $D(P_n)$ . Therefore, by the definition of the weighted Bottleneck distance,

$$d_B(D(P_n), D(f(P_n))) < \max\{2\varepsilon, \varepsilon_J\},$$

where  $\varepsilon$  and  $\varepsilon_J$  are sufficiently small, thus, completing the proof.  $\square$

In other words, the weighted persistent diagrams of persistent Jones polynomials are stable under small-amplitude or possibly irregular perturbations.

## 5. Applications

The utility of the proposed multi-scale and persistent Jones polynomial models is demonstrated through their application to real-world problems. These models provide robust frameworks for analyzing both local and global structural properties of curves in the 3-space. By leveraging their capacity to encode entanglement complexity features, these models provide new perspectives and tools for understanding protein flexibility, stability, and entanglement.

In this section, we illustrate two important applications of the proposed models. The first application focuses on predicting the B-factor of protein residues, a critical indicator of protein flexibility. This application demonstrates the practical utility of the multi-scale Jones polynomial in processing curve structural data. The second application uses persistent Jones polynomial to analyze the topological features of protein secondary structures, specifically  $\alpha$ -helix and  $\beta$ -sheets. These applications demonstrate the versatility and potential of the proposed models to advance computational structural biology.

### 5.1. Multi-scale Jones polynomial for B-factor prediction

B-factors, also known as Debye-Waller factors, measure atomic displacements within protein structures and provide insight into molecular flexibility and stability. Analysis of B-factors enables a deeper understanding of protein dynamics and aids in predicting regions with high structural mobility, which is crucial for understanding protein function and interactions.

To eliminate the influence of irrelevant atomic information and better capture the geometric and topological properties of the protein structure, each amino acid is represented by its  $C_\alpha$  atom. These  $C_\alpha$  atoms are sequentially connected to form a  $C_\alpha$  chain  $L$ . Let  $C = \{c_0, c_1, \dots, c_n\}$  denote the set of  $C_\alpha$  atoms arranged in the sequence of the protein. The  $C_\alpha$  chain of the protein is considered a disjoint open curve. Segmentation of the  $C_\alpha$  chain is achieved by cutting the midpoint between each  $C_\alpha$  atom

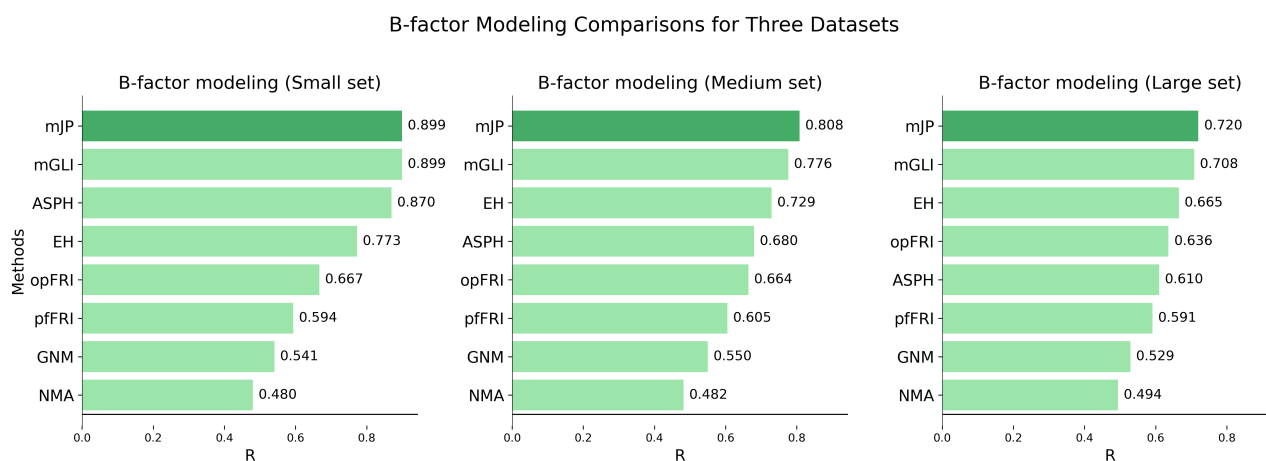
and its adjacent  $C_\alpha$  atom, denoted by  $P_n = \{l_0, l_1, \dots, l_n\}$ . The distance between two curve segments is defined as the distance between the  $C_\alpha$  atoms contained in the segments, that is,  $d(l_i, l_j) = d(c_i, c_j)$ .

In this study, we select the radius  $r$  to range from  $4\text{\AA}$  to  $15\text{\AA}$  with a step size of  $0.25\text{\AA}$ . We set  $R = (r + 1)\text{\AA}$ , so the radius  $R$  ranges from  $5\text{\AA}$  to  $16\text{\AA}$ . In total, the interception range is from  $4\text{\AA}$  to  $16\text{\AA}$ . Thus, there is a characteristic matrix for  $P_n$  of the protein  $C_\alpha$  chain  $L$ ,

$$\begin{pmatrix} JP_{4\text{\AA},5\text{\AA}}^1(10) & JP_{4.25\text{\AA},5.25\text{\AA}}^1(10) & \cdots & JP_{15\text{\AA},16\text{\AA}}^1(10) \\ JP_{4\text{\AA},5\text{\AA}}^2(10) & JP_{4.25\text{\AA},5.25\text{\AA}}^2(10) & \cdots & JP_{15\text{\AA},16\text{\AA}}^2(10) \\ \vdots & \vdots & \ddots & \vdots \\ JP_{4\text{\AA},5\text{\AA}}^n(10) & JP_{4.25\text{\AA},5.25\text{\AA}}^n(10) & \cdots & JP_{15\text{\AA},16\text{\AA}}^n(10) \end{pmatrix}.$$

This choice is motivated by the fact that the average distance between  $C_\alpha$  atoms is approximately  $3.8\text{\AA}$ . The selected radius scheme results in a powerful feature extraction method that provides rich representations of local protein structures. To minimize the influence of overly complex machine learning models, and to emphasize the effectiveness of the multi-scale Jones polynomial and avoid overfitting, we chose to use a Lasso regression model with parameter 0.16 for B-factor prediction.

To validate the effectiveness of the multi-scale Jones polynomial in predicting  $C_\alpha$  atom B factors across proteins of varying sizes, we compared our method with several previous approaches, including mGLI [20], evolutionary homology (EH) [37], atom-specific persistent homology (ASPH) [38], optimal flexibility-rigidity index (opFRI) [39], parameter free flexibility-rigidity index (pfFRI) [39], Gaussian network model (GNM) [30], and normal mode analysis (NMA) [30]. The comparison was performed on three sets of proteins from [30], as shown in Figure 6.

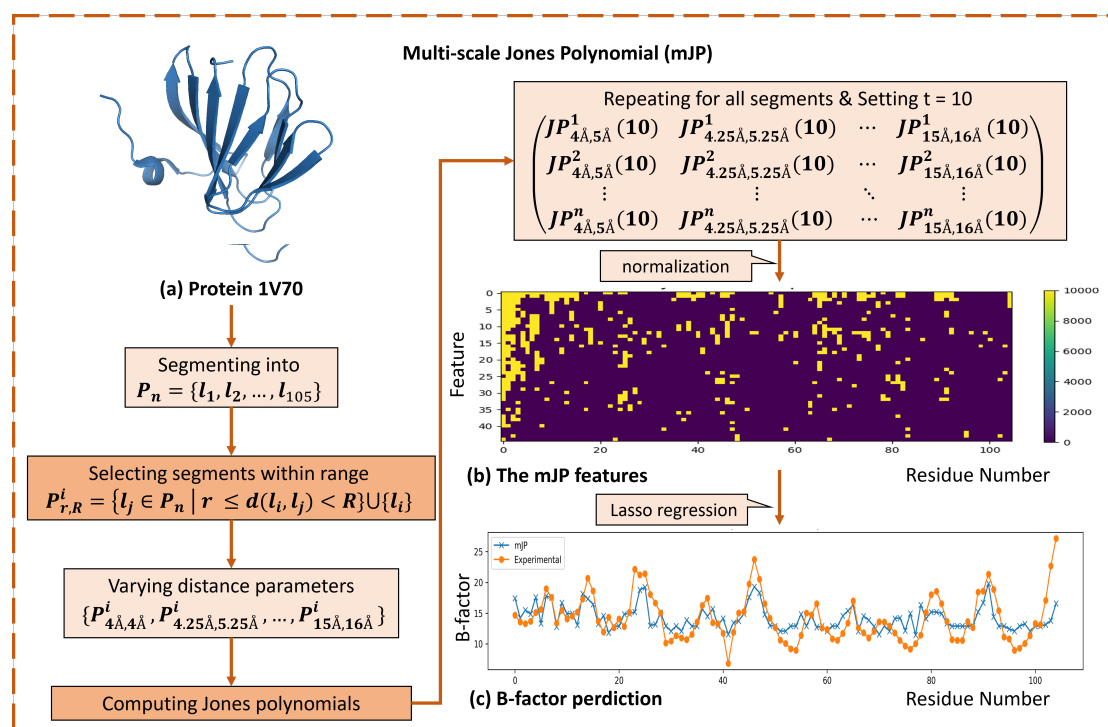


**Figure 6.** Comparison of B-factor predictions on three protein datasets between our multi-scale Jones polynomial method and other approaches from the literature.

The multi-scale Jones polynomial method achieved average correlation coefficients of 0.899, 0.808, and 0.720 for small, medium, and large protein sets, respectively. Our results on these three datasets outperformed previous methods.

To further illustrate the performance of the multi-scale Jones polynomial (mJP) analysis, we present a case study involving a potential antibiotic synthesis protein (PDBID: 1V70) containing 105 residues,

as shown in Figure 7(a). After processing with the mJP model, as shown in Figure 7, a characteristic matrix is generated. The normalized characteristic matrix is then used as the input to the Lasso regression model, as illustrated in Figure 7(b). The Lasso regression model is used to predict the B-factor of each residue and compare the predicted values with the experimentally determined values, as shown in Figure 7(c).



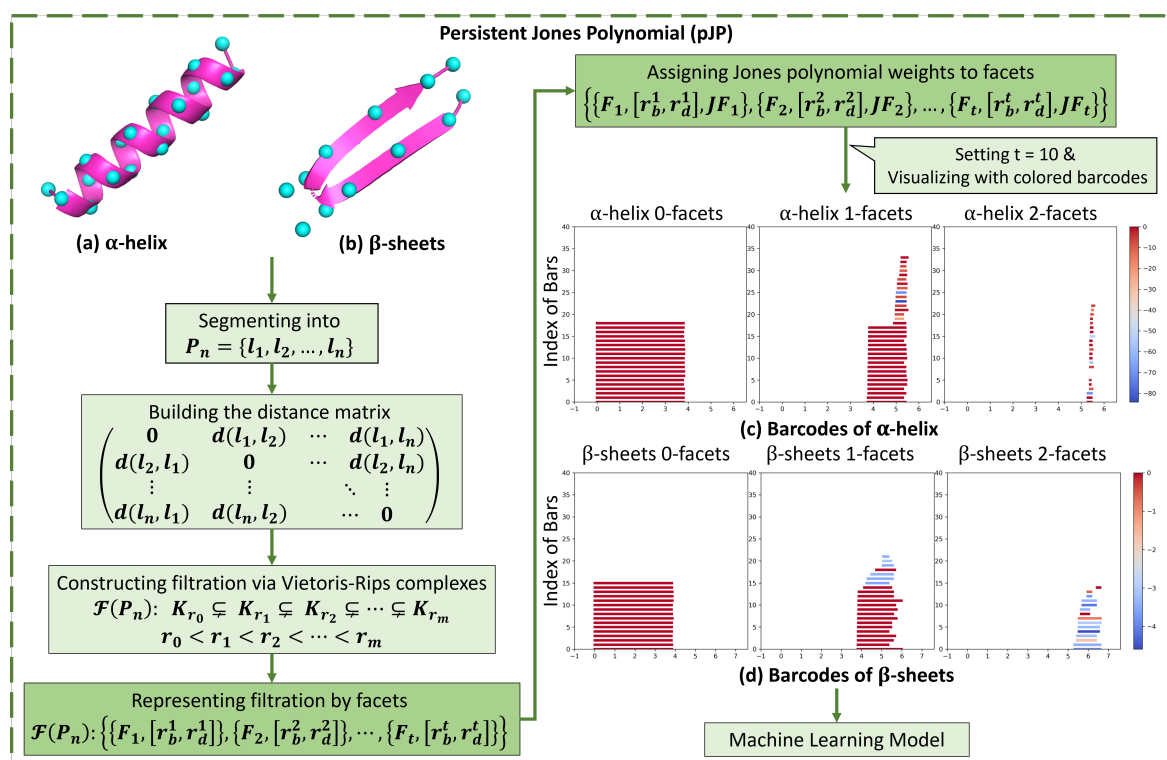
**Figure 7.** The process of mJP analysis for protein B-factor prediction. (a) The 3D structure of the protein 1V70; (b) The normalized characteristic matrix; (c) A comparison of the predicted B-factors with experimentally determined values.

Compared to traditional B-factor analysis methods, which focus on individual atoms, their spatial positions in the 3-space, and the thermal motion and disorder of atoms within the protein structure, our approach effectively captures the torsional entanglement of the peptide chain at each  $C_\alpha$  atomic position by incorporating the Jones polynomial. This torsional entanglement of the peptide chain significantly influences the observed B-factor values.

The torsional entanglement of protein peptide chains, captured through the multi-scale Jones polynomial, provides critical insight into the structural and functional dynamics of proteins. By analyzing torsional entanglement, this approach reveals patterns of molecular flexibility and rigidity that allow for an in-depth understanding of protein stability and function. Furthermore, this method improves our ability to model and predict regions of structural mobility, offering potential applications for protein engineering and drug discovery.

## 5.2. Barcodes of persistent Jones polynomial of $\alpha$ -helix and $\beta$ -sheets

In molecular biology,  $\alpha$ -helices and  $\beta$ -sheets are fundamental secondary structures in proteins, stabilized by hydrogen bonding patterns that contribute to the overall stability and function of the protein. Additionally  $\alpha$ -helices are typically more rigid than  $\beta$ -sheets. To explore the local structural complexity and stability of these structures, we employed topological analysis using the barcode representation of the persistent Jones polynomial (pJP). Using protein data from the Protein Data Bank (PDB), we demonstrated this approach with examples, including the analysis of an  $\alpha$ -helix chain consisting of 19 residues of the protein with PDB ID 1C26. Additionally, we extracted two parallel  $\beta$ -sheets consisting of 16 residues from the protein 2JOX to explore their barcodes representations of the persistent Jones polynomial.



**Figure 8.** The process of the pJP analysis for the  $\alpha$ -helix and  $\beta$ -sheets. (a) (b) The 3D structures of the  $\alpha$ -helix and  $\beta$ -sheets. (c) The colored barcodes visualizing the  $\alpha$ -helix. (d) The colored barcodes visualizing the  $\beta$ -sheets. The colored barcodes obtained through the pJP model can be applied to machine learning for protein structure analysis.

To eliminate the influence of irrelevant atomic information and better capture the geometric and topological properties of the protein structure, each amino acid is represented by its  $C_\alpha$  atom, as shown in Figure 8(a) and (b). These  $C_\alpha$  atoms are sequentially connected to form a  $C_\alpha$  chain, which is treated as a disjoint open curve. Segmentation of the  $C_\alpha$  chain is achieved by cutting the midpoint between each  $C_\alpha$  atom and its adjacent  $C_\alpha$  atom. The distance between two curve segments is defined as the distance between the  $C_\alpha$  atoms contained within these segments.

The process of the pJP model is illustrated in Figure 8, using colored barcodes to represent the

$\alpha$ -helix and  $\beta$ -sheets structures.

Figure 8(c) represents the barcodes corresponding to the  $\alpha$ -helix. In the 0-facets panel, there are 19 bars with Jones polynomial weights of 0. Each bar has a length of approximately  $3.8\text{\AA}$ , which is the average distance between two  $C_\alpha$  atoms. Additionally, the 1-facets panel contains 18 bars with similar birth times and life-spans, starting around  $3.8\text{\AA}$  and persisting until approximately  $5.4\text{\AA}$ , each with a Jones polynomial weight of 0. These bars correspond to facets formed by two adjacent  $C_\alpha$  atoms. Moreover, 16 short-lived bars represent facets formed by two nonadjacent  $C_\alpha$  atoms. As shown in the 2-facets panel, the bars capture the persistence of facets formed by three  $C_\alpha$  atoms, along with the complexity of entanglement of the corresponding polyline system.

Figure 8(d) represents the barcodes corresponding to the  $\beta$ -sheets. Similar to the  $\alpha$ -helix case, the segmentation and distance definitions are consistent. The 0-facets panel includes 16 0-facets bars, indicating the presence of 16  $C_\alpha$  atoms. In the 1-facets panel, there are 14 bars for facets formed by two adjacent  $C_\alpha$  atoms and 8 for facets formed by nonadjacent  $C_\alpha$  atoms. The bars in the 2-facets panel represent facets formed by three  $C_\alpha$  atoms and provide information about the complexity of entanglement. The longer lifespans of the bars in the 2-facets panel of the  $\beta$ -sheets compared to the  $\alpha$ -helix suggest that the  $C_\alpha$  atoms in the  $\beta$ -sheets are more spatially dispersed.

The color of the barcodes reflects the value of the Jones polynomial weight, which indicates the difference of the torsional entanglement of the set of curve segments in the system. It is important to note that the color gradient, whether it tends to red or blue, does not imply a higher or lower degree of entanglement complexity in the represented set of curve segments. Rather, a greater color difference between two bars indicates a greater difference between the sets of curve segments they represent.

From the color bars in Figures 8 (c) and (d), it can be observed that the Jones polynomial weight for the  $\alpha$ -helix ranges from  $-86$  to  $0$ , while for the  $\beta$ -sheets it ranges from  $-6$  to  $0$ . This indicates a greater variability in the sets of curve segments represented by the facets within the  $\alpha$ -helix compared to the  $\beta$ -sheets, suggesting that the  $\alpha$ -helix exhibits more complex entanglement, as observed.

The persistent Jones polynomial effectively captures the torsional entanglement of secondary structures, such as  $\alpha$ -helices and  $\beta$ -sheets, within protein peptide chains. This torsional entanglement plays a critical role in the analysis of protein structure and function as it provides insight beyond atomic positions alone. By incorporating the Jones polynomial, our approach reveals important topological characteristics that are important for protein stability and functionality.

## 6. Concluding remarks

In this study, the selection of the segmentation  $P_n = \{l_1, l_2, \dots, l_n\}$  for a collection of curves  $L$  is crucial to capture the topological and geometric characteristics of the curve structure  $L$ . The segmentation serves as the basis for defining and calculating both the multi-scale analysis of the Jones polynomial and the persistent Jones polynomial. First, the outcomes of these two models depend not only on the spatial positions of the segments but also on their relative lengths in relation to the entire curve. As the segment length approaches zero, the results of the models tend toward triviality. Similarly, when the segment extends to cover the entire curve, the models recover global information. In both of these cases, the models cannot extract meaningful local information for spatial data. Thus, the choice of segmentation depends on the specific application.

Knot theory has traditionally focused on global invariants, but real-world applications often require

local structural insights. Classical knot invariants mainly reflect global topology and fail to capture crucial local structural details in applications like molecular biology and highway crossing design. To address this gap, localized versions of invariants such as the multi-scale Jones polynomials and persistent Jones polynomials have been developed. These localized models decompose global invariants for analyzing local topology in the context of the entire structure, promoting the application of KDA or CDA in systems where both global and local structures matter.

The stability of the model is crucial for practical applications. In real-world data, noise and minor perturbations pose challenges. Stability ensures that minor input changes do not cause disproportionate changes in the calculated invariants. It is critical in biological or physical contexts. For multi-scale Jones polynomial and persistent Jones polynomial models, we demonstrate stability under small perturbations. Minor adjustments in the collection  $L$  result in slight modifications to characteristic matrices and barcodes or diagrams. This stability makes the models reliable for structural topology and is applicable in KDA or CDA for real-world data.

The torsional entanglement of protein peptide chains, captured through both multi-scale and persistent Jones polynomials, provides crucial insights into the structure and function of proteins. By analyzing torsional entanglement, this approach reveals patterns of molecular flexibility and rigidity that allow a detailed understanding of protein structure and function. Furthermore, this method improves our capacity to model and predict regions of structural mobility and reactivity, offering valuable implications for enzyme kinetics and protein engineering.

In Section 5.2 of this manuscript, the barcodes of the persistent Jones polynomial represent the birth, death, and lifespan of the facets in the complexes under filtration. In contrast, persistent homology barcodes capture the changes in the homology classes of the complexes during filtration, i.e., the changes in the generators of the homology groups of the complexes (the number of generators corresponds to the Betti number). Despite these differences, there are important similarities between the two concepts. Both are based on filtration and provide insight into data characteristics by examining the evolution of the complexes during filtration. In addition, the bars in the barcodes of the persistent Jones polynomial represent facets formed by subsets of curve segments from the segmentation  $P_n = \{l_1, l_2, \dots, l_n\}$  of the collection of curves  $L$ . Therefore, these bars are constructed based on the distance conditions between the curve segments in the segmentation  $P_n$ . Similarly, in the case of point cloud data, persistent homology barcodes are also constructed according to the distance conditions of the points in the point cloud data.

The multi-scale Gauss link integral model [20] and the present multi-scale Jones polynomial and persistent Jones polynomial models represent solid advances in computational geometric topology. These approaches have great potential for real-world applications when combined with machine learning and artificial intelligence.

## Author contributions

Ruzhi Song: Writing-original draft, Writing-review & editing; Fengling Li: Methodology, Writing-original draft, Writing-review & editing; Jie Wu: Conceptualization; Fengchun Lei: Conceptualization, Methodology; Guo-Wei Wei: Conceptualization, Writing-review & editing. All authors have read and approved the final version of the manuscript for publication.

## Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Acknowledgments

This work was supported in part by grant (No. 12331003) of National Natural Science Foundation of China, State Key Laboratory of Structural Analysis, Optimization and CAE Software for Industrial Equipment, and Beijing Institute of Mathematical Sciences and Applications. The work of GWW was supported in part by NIH grants R01GM126189, R01AI164266, and R35GM148196, NSF grants DMS-2052983, DMS-1761320, DMS-2245903, and IIS-1900473, MSU Research Foundation, and Bristol-Myers Squibb 65109. The work of JW was supported in part by High-level Scientific Research Foundation of Hebei Province.

## Conflict of interest

The authors declare no competing interest.

## References

1. R. H. Crowell, R. H. Fox, *Introduction to knot theory*, New York: Springer, 1963. <https://dx.doi.org/10.1007/978-1-4612-9935-6>
2. J. W. Alexander, Topological invariants of knots and links, *Trans. Amer. Math. Soc.*, **30** (1928), 275–306. <https://dx.doi.org/10.1090/S0002-9947-1928-1501429-1>
3. V. F. R. Jones, A polynomial invariant for knots via von Neumann algebras, *Bull. Amer. Math. Soc.*, **12** (1985), 103–111. <https://dx.doi.org/10.1090/s0273-0979-1985-15304-2>
4. C. Manolescu, An introduction to knot Floer homology, 2014, arXiv:1401.7107.
5. M. Khovanov, A categorification of the Jones polynomial, *Duke Math. J.*, **101** (2000), 359–426. <https://dx.doi.org/10.1215/S0012-7094-00-10131-7>
6. T. Ohtsuki, *Quantum invariants: A study of knots, 3-manifolds, and their sets*, Singapore: World Scientific, 2001. <https://dx.doi.org/10.1142/4746>
7. C. Z. Liang, K. Mislow, Knots in proteins, *J. Am. Chem. Soc.*, **116** (1994), 11189–11190. <https://dx.doi.org/10.1021/ja00103a057>
8. D. W. Sumners, *The role of knot theory in DNA research*, Boca Raton: CRC Press, 1986.
9. T. Schlick, Q. Y. Zhu, A. Dey, S. Jain, S. T. Yan, A. Laederach, To knot or not to knot: multiple conformations of the SARS-CoV-2 frameshifting RNA element, *J. Am. Chem. Soc.*, **143** (2021), 11404–11422. <https://dx.doi.org/10.1021/jacs.1c03003>
10. K. C. Millett, E. J. Rawdon, A. Stasiak, J. I. Sułkowska, Identifying knots in proteins, *Biochem. Soc. Trans.*, **41** (2013), 533–537. <https://dx.doi.org/10.1042/bst20120339>
11. J. Qin, S. T. Milner, Counting polymer knots to find the entanglement length, *Soft Matter*, **7** (2011), 10676–10693. <https://dx.doi.org/10.1039/c1sm05972f>

12. Y. Z. Liu, M. O’Keeffe, M. M. J. Treacy, O. M. Yaghi, The geometry of periodic knots, polycatenanes and weaving from a chemical perspective: a library for reticular chemistry, *Chem. Soc. Rev.*, **47** (2018), 4642–4664. <https://dx.doi.org/10.1039/c7cs00695k>
13. R. L. Ricca, Topology bounds energy of knots and links, *Proc. R. Soc. A.*, **464** (2008), 293–300. <https://dx.doi.org/10.1098/rspa.2007.0174>
14. E. Panagiotou, K. C. Millett, P. J. Atzberger, Topological methods for polymeric materials: characterizing the relationship between polymer entanglement and viscoelasticity, *Polymers*, **11** (2019), 437. <https://dx.doi.org/10.3390/polym11030437>
15. J. Arsuaga, M. Vazquez, P. McGuirk, S. Trigueros, D. W. Sumners, J. Roca, DNA knots reveal a chiral organization of DNA in phage capsids, *Proc. Natl. Acad. Sci. U.S.A.*, **102** (2005), 9165–9169. <https://dx.doi.org/10.1073/pnas.0409323102>
16. J. I. Sulkowska, E. J. Rawdon, K. C. Millett, J. N. Onuchic, A. Stasiak, Conservation of complex knotting and slipknotting patterns in proteins, *Proc. Natl. Acad. Sci. U.S.A.*, **109** (2012), E1715–E1723. <https://dx.doi.org/10.1073/pnas.1205918109>
17. P. Freyd, D. Yetter, J. Hoste, W. B. R. Lickorish, K. Millett, A. Ocneanu, A new polynomial invariant of knots and links, *Bull. Amer. Math. Soc.*, **12** (1985), 239–246. <https://dx.doi.org/10.1090/s0273-0979-1985-15361-3>
18. L. H. Kauffman, An invariant of regular isotopy, *Trans. Amer. Math. Soc.*, **318** (1990), 417–471. <https://dx.doi.org/10.1090/S0002-9947-1990-0958895-7>
19. J. H. Przytycki, P. Traczyk, Conway algebras and skein equivalence of links, *Proc. Amer. Math. Soc.*, **100** (1987), 744–748. <https://dx.doi.org/10.1090/S0002-9939-1987-0894448-2>
20. L. Shen, H. S. Feng, F. L. Li, F. C. Lei, J. Wu, G.-W. Wei, Knot data analysis using multiscale Gauss link integral, *Proc. Natl. Acad. Sci. U.S.A.*, **121** (2024), e2408431121. <https://dx.doi.org/10.1073/pnas.2408431121>
21. E. Panagiotou, K. W. Plaxco, A topological study of protein folding kinetics, 2018, arXiv:1812.08721.
22. Q. Baldwin, E. Panagiotou, The local topological free energy of proteins, *J. Theor. Biol.*, **529** (2021), 110854. <https://dx.doi.org/10.1016/j.jtbi.2021.110854>
23. Q. Baldwin, B. Sumpter, E. Panagiotou, The local topological free energy of the SARS-CoV-2 Spike protein, *Polymers*, **14** (2022), 3014. <https://dx.doi.org/10.3390/polym14153014>
24. E. Panagiotou, L. H. Kauffman, Knot polynomials of open and closed curves, *Proc. R. Soc. A.*, **476** (2020), 20200124. <https://dx.doi.org/10.1098/rspa.2020.0124>
25. K. Barkataki, E. Panagiotou, The Jones polynomial of collections of open curves in 3-space, *Proc. R. Soc. A.*, **478** (2022), 20220302. <https://dx.doi.org/10.1098/rspa.2022.0302>
26. E. Panagiotou, L. H. Kauffman, Vassiliev measures of complexity of open and closed curves in 3-space, *Proc. R. Soc. A.*, **477** (2021), 20210440. <https://dx.doi.org/10.1098/rspa.2021.0440>
27. J. Wang, E. Panagiotou, The protein folding rate and the geometry and topology of the native state, *Sci. Rep.*, **12** (2022), 6384. <https://dx.doi.org/10.1038/s41598-022-09924-0>



28. T. Herschberg, K. Pifer, E. Panagiotou, A computational package for measuring Topological Entanglement in Polymers, Proteins and Periodic systems (TEPPP), *Comput. Phys. Commun.*, **286** (2023), 108639. <https://dx.doi.org/10.1016/j.cpc.2022.108639>
29. Z. X. Cang, G.-W. Wei, Persistent cohomology for data with multicomponent heterogeneous information, *SIAM J. Math. Data Sci.*, **2** (2020), 396–418. <https://dx.doi.org/10.1137/19m1272226>
30. J. K. Park, R. Jernigan, Z. J. Wu, Coarse grained normal mode analysis vs. refined gaussian network model for protein residue-level structural fluctuations, *Bull. Math. Biol.*, **75** (2013), 124–160. <https://dx.doi.org/10.1007/s11538-012-9797-y>
31. V. Turaev, Knotoids, *Osaka J. Math.*, **49** (2012), 195–223. <https://dx.doi.org/10.18910/10080>
32. N. Gügümcü, L. H. Kauffman, New invariants of knotoids, *Eur. J. Combin.*, **65** (2017), 186–229. <https://dx.doi.org/10.1016/j.ejc.2017.06.004>
33. N. Gügümcü, S. Lambropoulou, Knotoids, braidoids and applications, *Symmetry*, **9** (2017), 315. <https://dx.doi.org/10.3390/sym9120315>
34. N. Gügümcü, L. Kauffman, Parity in knotoids, 2019, arXiv:1905.04089.
35. M. Manouras, S. Lambropoulou, L. H. Kauffman, Finite type invariants for knotoids, *Eur. J. Combin.*, **98** (2021), 103402. <https://dx.doi.org/10.1016/j.ejc.2021.103402>
36. D. Cohen-Steiner, H. Edelsbrunner, J. Harer, Stability of persistence diagrams, In: *Proceedings of the twenty-first annual symposium on computational geometry*, New York: Association for Computing Machinery, 2005, 263–271. <https://dx.doi.org/10.1145/1064092.1064133>
37. Z. X. Cang, E. Munch, G.-W. Wei, Evolutionary homology on coupled dynamical systems with applications to protein flexibility analysis, *J. Appl. and Comput. Topology*, **4** (2020), 481–507. <https://dx.doi.org/10.1007/s41468-020-00057-9>
38. D. Bramer, G.-W. Wei, Atom-specific persistent homology and its application to protein flexibility analysis, *Comput. Math. Biophys.*, **8** (2020), 1–35. <https://dx.doi.org/10.1515/cmb-2020-0001>
39. K. Opron, K. L. Xia, G.-W. Wei, Fast and anisotropic flexibility-rigidity index for protein flexibility and fluctuation analysis, *J. Chem. Phys.*, **140** (2014), 234105. <https://dx.doi.org/10.1063/1.4882258>



AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)