

ChatGPT for Computational Topology

Jian Liu ^{†1,2}, Li Shen², and Guo-Wei Wei ^{*2,3,4}

¹*Mathematical Science Research Center, Chongqing University of Technology, Chongqing 400054, China*

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

³*Department of Electrical and Computer Engineering, Michigan State University, MI 48824, USA*

⁴*Department of Biochemistry and Molecular Biology, Michigan State University, MI 48824, USA*

Abstract ChatGPT represents a significant milestone in the field of artificial intelligence (AI), finding widespread applications across diverse domains. However, its effectiveness in mathematical contexts has been somewhat constrained by its susceptibility to conceptual errors. Concurrently, topological data analysis (TDA), a relatively new discipline, has garnered substantial interest in recent years. Nonetheless, the advancement of TDA is impeded by the limited understanding of computational algorithms and coding proficiency among theoreticians. This work endeavors to bridge the gap between theoretical topological concepts and their practical implementation in computational topology through the utilization of ChatGPT. We showcase how a pure theoretician, devoid of computational experience and coding skills, can effectively transform mathematical formulations and concepts into functional codes for computational topology with the assistance of ChatGPT. Our strategy outlines a productive process wherein a mathematician trains ChatGPT on pure mathematical concepts, steers ChatGPT towards generating computational topology codes, and subsequently validates the generated codes using established examples. Our specific case studies encompass the computation of Betti numbers, Laplacian matrices, and Dirac matrices for simplicial complexes, as well as the persistence of various homologies and Laplacians. Furthermore, we explore the application of ChatGPT in computing recently developed topological theories for hypergraphs and digraphs, as well as the persistent harmonic space, which has not been computed in the literature, to the best of our knowledge. This work serves as an initial step towards effectively transforming pure mathematical theories into practical computational tools, with the ultimate goal of enabling real applications across diverse fields.

Keywords ChatGPT, computational topology, homology, Laplacians, Dirac, persistence.

[†]Corresponding author: liujian2@mail.nankai.edu.cn

^{*}Corresponding author: weig@msu.edu

¹**2020 Mathematics Subject Classification.** Primary 55N31; Secondary 55-04, 55-08.

Contents

Abstract	1
Keywords	1
1 Introduction	3
2 ChatGPT for homology, Laplacian, and Dirac	4
2.1 Simplicial complexes and Betti numbers	4
2.2 Computing Dirac and Laplacian matrices	12
3 ChatGPT for hypergraphs and directed graphs	13
3.1 Betti numbers for hypergraphs	14
3.2 Betti numbers for directed graphs	15
3.3 Computing hypergraph Laplacian	16
3.4 Computing path Laplacian	21
3.5 Computing hyperdigraph Laplacian	22
4 ChatGPT for topological persistence	22
4.1 ChatGPT for persistent homology and persistent Laplacians	22
4.2 ChatGPT for persistent harmonic space	23
5 Applications	29
6 Concluding remarks	32

1 Introduction

Mathematical theory guides the advancement of applied sciences, while the progress of applied sciences, in turn, provides direction and motivation for mathematical exploration. However, the development of these disciplines can inadvertently erect barriers between them. A theoretical mathematician often finds it challenging to divert their focus toward translating their theoretical research into practical programs that can be utilized by researchers in the applied sciences for numerical experiments. Conversely, an expert in applied sciences may struggle to grasp the intricacies of advanced mathematical theories and harness this knowledge to optimize their methods and models. This interplay between theory and application highlights the need for bridging the gap and fostering collaboration between these two seemingly disparate domains.

Recently, ChatGPT has excelled in various domains and has been driving advancements in artificial intelligence (AI). For example, ChatGPT is being used to assist in drug discovery and development [ST23, WFW23]. Harnessing ChatGPT to assist theoretical mathematicians in developing algorithms and codes, even without proficiency in computer languages, can be a game-changer. This approach addresses the critical need for collaboration between theoretical and applied scientists, as it facilitates knowledge exchange and interdisciplinary research. One promising avenue is to use natural language interfaces to interact with ChatGPT. Theoretical mathematicians can engage in conversations with ChatGPT, train ChatGPT with correct mathematical concepts or theories, and describe their computational objectives in plain language. ChatGPT can then assist them in translating these theories into algorithms and codes, without requiring an in-depth understanding of specific programming languages and algorithms by theoreticians. It can provide code snippets, algorithm frameworks, and executable scripts to help theoretical mathematicians initiate practical computational work.

Topology is one of the most profound and challenging areas in modern mathematics. In recent years, topological data analysis (TDA) has gained significant attention in applied fields such as computer science, molecular biology and biophysics, materials science, medical imaging, etc. An increasing number of scholars and researchers are dedicating themselves to develop the theories on TDA [Car09, CZCG04, ELZ02, ZC04]. The computational topology is an emerging field in mathematics [EH22, KMM04]. However, the lack of algorithm understanding and coding skills limits mathematicians' success in computational topology. In this work, we aim to leverage ChatGPT to close the gap between pure mathematical theories and practical codes for computational topology.

Some traditional perspectives argue that ChatGPT may not be suitable for tasks involving complex mathematical concepts and intricate logical reasoning due to its limited depth in mathematics. Many limitations stem from several factors. First, ChatGPT's understanding of mathematics is based on pre-existing knowledge up to its last training data, which means it may not be aware of recent developments or specialized mathematical domains. Second, while it can perform various mathematical calculations and explain some fundamental concepts, it lacks the depth of understanding and rigor that specialized mathematicians or dedicated mathematical software can provide. Additionally, ChatGPT may struggle with highly specialized or research-level problems that require nuanced domain knowledge and precise mathematical proofs. Consequently, when employing ChatGPT for coding tasks, users must supplement its capabilities with their own expertise in mathematics and employ specific strategies to supervise and guide ChatGPT in generating accurate codes.

When using ChatGPT, it is essential to keep the following considerations in mind:

- **Answers May Not Be Definitive:** ChatGPT's responses to questions are not guaranteed to be correct or entirely satisfactory. Its answers are based on patterns and information in its training data, and it may not always provide accurate or complete responses.
- **Variability in Responses:** The same question posed to ChatGPT in different instances or chat interfaces may yield different answers. This variability can occur due to the model's inherent randomness and sensitivity to the input phrasing.

- **Responses Evolve with Interaction:** In a single chat session, ChatGPT’s answers can evolve and become closer to what the user desires as the conversation progresses. It adapts to the user’s guidance and context.

During the interaction with ChatGPT, we always assume that the questioners are familiar the related mathematical theories but lacks coding experience, skill, and understanding.

In the next section, we will recall the definitions and notations for simplicial complexes and Betti numbers. We demonstrate how to obtain practical codes for computing the Betti numbers, Laplacian, and Dirac of a simplicial complex using ChatGPT. We intend to use the Python programming language for this purpose, as Python is widely adopted in the scientific and data analysis communities. Also, Python offers a rich ecosystem of libraries and tools for scientific computing and data analysis which can significantly simplify the implementation of mathematical algorithms. In Section 3, we show how to use ChatGPT to discover the codes for computing the Betti numbers and Laplacian matrices for hypergraphs and digraphs. Section 4 provides a note on the persistence of homology, Laplacians, and harmonic space. In Section 5, we present some application examples. Finally, we summarize the main works of this paper and outline further perspectives about ChatGPT. ChatGPT version 3.5 was used in this work.

2 ChatGPT for homology, Laplacian, and Dirac

Persistent homology is one of the fundamental theoretical tools in TDA, which can describe the multi-scale information of data sets [CZCG04, ELZ02, EH⁺08, ZC04]. Persistent Laplacians, including persistent spectral graph [WNW20] and evolutionary de Rham-Hodge theory [CZTW19], were introduced in 2019. Persistent Laplacians not only provide geometric information of data sets in terms of topological invariants as persistent homology does, but also capture homotopic shape evolution that cannot be detected by persistent homology [WNW20, LLW23, MWW22]. Their applications in real-world problems show remarkable performance [CQWW22, MX21]. Furthermore, the persistent Dirac has also been introduced as a new computational tool in TDA [AMS22, WBX23]. In this section, we review the concepts of simplicial complexes, homology, and Betti numbers. Then we showcase the use of ChatGPT for computing Betti numbers from a mathematical perspective. Moreover, we instruct ChatGPT to produce the codes for computing Dirac and Laplacian operators.

Our questions to ChatGPT and the design of corresponding codes are based on our mathematical background on homology theory. In this work, the ground field is always assumed to be $\mathbb{K}$. In particular, when we consider the Laplacian, the ground field is taken to be the real number field $\mathbb{K} = \mathbb{R}$.

2.1 Simplicial complexes and Betti numbers

Computers are more adept at handling and processing discrete objects. As a result, researchers often focus their efforts on investigating simplicial complexes or their discrete representations, known as abstract simplicial complexes.

Let $\mathcal{K}$ be an abstract simplicial complex on a finite nonempty ordered set V . Let $C_p(\mathcal{K})$ be the vector space with basis given by the p -simplices of $\mathcal{K}$. We have the chain complex $C_*(\mathcal{K}) = (C_p(\mathcal{K}))_{p \geq 0}$. The p -th boundary operator $d_p : C_p(\mathcal{K}) \rightarrow C_{p-1}(\mathcal{K})$ is given by

$$d_p[v_0, \dots, v_p] = \sum_{i=0}^p (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_p].$$

Here, $[v_0, \dots, v_p]$ is a p -simplex of $\mathcal{K}$, and $\hat{v}_i$ means omission of v_i . Choose the basis of $C_p(\mathcal{K})$ as the lexicographic order of the p -simplices. Then we can obtain the representation matrix B_p of the boundary operator d_p . Then the p -th Betti number of $\mathcal{K}$ is given by

$$\beta_p = n_p - \text{rank } B_p - \text{rank } B_{p+1}. \quad (1)$$

The Betti numbers for a simplicial complex can be computed by the above formula.

In this context, our goal is to utilize ChatGPT to provide computer codes that take an abstract simplicial complex as input and calculates its Betti numbers as output. However, if we directly present this task to ChatGPT, it might offer inappropriate or impractical coding suggestions, or it might recommend using existing libraries. The former is not what we desire, and the latter does not align with our goal of modifying and adapting the code to suit our specific needs. As shown in Fig. 1 and Fig. 2, for the same question, ChatGPT provides different responses along with corresponding code snippets. In its responses, ChatGPT presents a Python code, but it references external libraries such as NetworkX and Gudhi. Importantly, referencing the Gudhi library includes functions for computing simplicial complexes and their Betti numbers, which is not the approach we want. Additionally, the method suggested in the response to Fig. 1, involving the 0-dimensional Laplacian matrix of the simplicial complex, is also incorrect as it cannot help us compute the Betti numbers of the complex.

Q: Given an abstract simplicial complex, can you give me a Python code for computing Betti numbers for the simplicial complex?

🧠 Certainly! You can compute Betti numbers for an abstract simplicial complex using various Python libraries, such as **'networkx'** and **'scipy'**. First, make sure you have these libraries installed:

```
pip install networkx scipy
```

Here's a Python code example to compute Betti numbers for a simplicial complex:

```
import networkx as nx
import scipy.linalg

def betti_numbers(simplicial_complex):
    # Create a directed graph from the simplicial complex
    G = nx.DiGraph()

    for simplex in simplicial_complex:
        for i in range(len(simplex)):
            for j in range(i + 1, len(simplex)):
                G.add_edge(simplex[i], simplex[j])

    # Compute the directed graph Laplacian matrix
    laplacian_matrix = nx.directed_laplacian_matrix(G).toarray()

    # Compute the matrix rank
    rank = scipy.linalg.matrix_rank(laplacian_matrix)
    # Compute the Betti numbers
    num_nodes = len(G.nodes())
    betti_0 = num_nodes - rank
    betti_1 = len(simplicial_complex) - num_nodes + rank

    return betti_0, betti_1

# Example usage
if __name__ == "__main__":
    # Define your simplicial complex as a list of simplices (lists of vertices)
    simplicial_complex = [[0, 1], [1, 2], [2, 3], [3, 4], [0, 2, 4]]
    betti_0, betti_1 = betti_numbers(simplicial_complex)
    print("Betti 0:", betti_0)
    print("Betti 1:", betti_1)
```

In this code:

1. We create a directed graph from the simplicial complex, where each node represents a vertex, and directed edges connect vertices if they are part of the same simplex.
 2. We compute the directed graph's Laplacian matrix using **'nx.directed_laplacian_matrix'**.
 3. We calculate the rank of the Laplacian matrix to determine the number of independent cycles in the complex.
 4. Finally, we use the rank information to compute the Betti numbers: Betti 0 and Betti 1.
- You can modify the **'simplicial_complex'** variable to represent your specific simplicial complex, and the code will compute the corresponding Betti numbers.

Figure 1: ChatGPT provides the answer for computing Betti numbers (I)

Q: Given an abstract simplicial complex, can you give me a Python code for computing Betti numbers for the simplicial complex?

🧠 Certainly! To compute the Betti numbers for an abstract simplicial complex in Python, you can use libraries such as NetworkX or Gudhi. Here, I'll show you how to compute Betti numbers using the Gudhi library, which is specifically designed for topological data analysis.

First, you'll need to install Gudhi if you haven't already. You can install it using pip:

```
pip install gudhi
```

Now, let's compute the Betti numbers for your simplicial complex:

```
import gudhi
# Create a simplicial complex
simplicial_complex = gudhi.SimplexTree()

# Add simplices to the complex
# For example, adding a triangle
simplicial_complex.insert([0, 1, 2])
# Add more simplices as needed

# Compute Betti numbers
betty_numbers = simplicial_complex.betti_numbers()

# Print the Betti numbers
print("Betti numbers:", betty_numbers)
```

Replace the 'insert' calls with the simplices of your simplicial complex. The code above will compute the Betti numbers for the complex and print them.

Here's a brief explanation of the code:

1. We import the 'gudhi' library.
2. We create a 'SimplexTree' object to represent the simplicial complex.
3. We add simplices to the complex using the 'insert' method. You should replace this with the simplices from your abstract simplicial complex.
4. We compute the Betti numbers using the 'betti_numbers' method.
5. Finally, we print the Betti numbers.

This code should work for computing Betti numbers of a simplicial complex using Gudhi. Make sure to replace the simplices with your actual data.

Figure 2: ChatGPT provides the answer for computing Betti numbers (II)

The example mentioned above highlights the necessity of utilizing our own mathematical understanding and expertise to guide and supervise ChatGPT in writing mathematical codes. This involves two main aspects. First, we need to train ChatGPT to "understand" the concepts and notations we provide, such as simplicial complexes, homology, and Betti numbers. Second, we need to design test problems for which we know the answer and guide ChatGPT to come up with codes to compute correct results.

With regard to this matter, we intended to train ChatGPT to use the formula in Eq. (1) to calculate the Betti numbers of a simplicial complex. During this process, we needed to make sure ChatGPT understood what an abstract simplicial complex is. To achieve this, we began by asking ChatGPT for the definition of an

abstract simplicial complex. We then discovered that ChatGPT can comprehend the concept of an abstract simplicial complex. However, we had noticed that the representation of simplicial complexes may vary. For example, the representation of $\Delta[1]$ could sometimes be given as $\{\{0\}, \{1\}, \{0, 1\}\}$, or $[[0], [1], [0, 1]]$, or $[(0), (1), (0, 1)]$. And the order of simplices might not always follow lexicographical order. These variations in representation might be due to differences in computer data types or mathematical expressions. To address this issue, we had decided to standardize the representation of simplicial complexes to a format similar to $[[0], [1], [0, 1]]$, with simplices ordered in lexicographical order. We did not dwell on the specific computer data types and storage mechanisms.

After confirming that ChatGPT can correctly understand abstract simplicial complexes, we decided to guide it in generating abstract simplicial complexes in a specific manner. More precisely, by providing a finite number of points in a Euclidean space and a real-value distance as input, ChatGPT needed to generate a Rips complex as output. This process may not succeed in a single attempt, but through iterative questioning and some debugging, ChatGPT successfully provided the desired code. The sample results provided by ChatGPT's code is shown in Fig. 3. In this example, we take the discrete point set $\{(0, 0), (2, 1), (2, 1), (3, 3), (3, 4)\}$ and the distance $d = 3.0$ as input.

```
Abstract Simplicial Complex:
Vertices: [0, 1, 2, 3, 4]
Simplices: [[0], [1], [2], [3], [4], [0, 1], [0, 2], [1, 2], [1, 3], [2, 3], [3, 4], [0, 1, 2], [1, 2, 3]]
Highest Dimension: 2
```

Figure 3: An example for generating a Rips complex

In order to enable ChatGPT to generate codes for computing Betti numbers correctly, we had decided to have it first performing the computation of the boundary matrix. As before, we first confirmed whether ChatGPT understood what the boundary operator on the chain complex associated with a simplicial complex is. Fig. 4 shows how we approached ChatGPT with this task.

Q: Given an abstract simplicial complex $[[2], [5], [4], [1], [3], [2, 4], [1, 2], [3, 4], [2, 3], [2, 3, 4]]$, can you show how to compute the boundary matrix by a Python code?

 To compute the boundary matrix of an abstract simplicial complex, you can use Python. The boundary matrix represents the boundary operators that map simplices to their faces. Here's a Python code to compute the boundary matrix for the given abstract simplicial complex:

```
import numpy as np

# Define the abstract simplicial complex as a list of simplices
complex = [[2], [5], [4], [1], [3], [2, 4], [1, 2], [3, 4], [2, 3], [2, 3, 4]]

# Extract the unique vertices from the simplicial complex
vertices = list(set([v for simplex in complex for v in simplex]))
vertices.sort() # Sort vertices for consistent ordering

# Create a dictionary to map vertices to their indices
vertex_to_index = {v: i for i, v in enumerate(vertices)}

# Determine the number of vertices and simplices
num_vertices = len(vertices)
num_simplices = len(complex)

# Initialize the boundary matrix with zeros
boundary_matrix = np.zeros((num_simplices, num_vertices), dtype=int)

# Fill in the boundary matrix
for i, simplex in enumerate(complex):
    for j, vertex in enumerate(simplex):
        # Remove the j-th vertex from the simplex to get the (j-1)-face
        face = simplex[:j] + simplex[j + 1:]

        # Find the index of the face in the vertex list
        face_index = vertex_to_index[face[0]]

        # Add 1 to the (i, face_index) entry of the boundary matrix
        boundary_matrix[i, face_index] = 1 if j % 2 == 0 else -1

# Print the boundary matrix
print("Boundary Matrix:")
print(boundary_matrix)
```

This code first extracts the unique vertices and assigns indices to them, initializes the boundary matrix with zeros, and then fills in the matrix according to the boundary operator rules. The result is the boundary matrix for the given abstract simplicial complex.

Figure 4: We ask ChatGPT how to compute the boundary matrix.

ChatGPT provided us with various responses to the task we were given. As object-oriented individuals, we chose our preferences to proceed with the question. The result code from ChatGPT's responses is shown in Fig. 4. We had placed and executed the code into Python, however, obtaining an error as shown in Fig. 5.

```
File "C:\Users\Lenovo\spyder-py3\ChatGPT for topology\boundary.2.py", line 35, in <module>
face_index = vertex_to_index[face[0]]
IndexError: list index out of range
```

Figure 5: Feedback information on the error.

For the debugging process, we directly input the error message to ChatGPT for feedback. After several attempts, ChatGPT offered the code without errors by adding the code in Fig. 6 before “# Find the index of the face in the vertex list” in Fig. 4.

```
if len(face) == 0:
    continue
```

Figure 6: The code added after debugging.

We validated the code through several test examples and found that the computed boundary matrices were incorrect. At this point, we need to review the code and understand what caused the error. Fortunately, ChatGPT provides comments for its codes, which will help us read and understand it better. Additionally, through interaction with ChatGPT, we can inquire about the parts of the code that are difficult to understand, and ChatGPT will respond accordingly. This process can not only help us quickly pinpoint the reasons for code errors but also evaluate whether ChatGPT lacks the understanding of mathematical algorithms. ChatGPT plays the role of providing a general code framework and interpreting the code in the above process. In the end, we obtained the code as shown in Fig. 7, which can correctly compute Betti numbers of a given abstract simplicial complex.

```

import numpy as np

def compute_betti(complex):
    num_simplices = len(complex)
    m = len(complex[num_simplices - 1]) - 1
    b = [0, 0]
    k = 1

    for i in range(num_simplices):
        if len(complex[i]) == k:
            b[k] += 1
        else:
            k += 1
            b.append(b[k - 1] + 1)

    boundary_matrix = np.zeros((num_simplices, num_simplices), dtype=int)

    for i, simplex in enumerate(complex):
        for j, vertex in enumerate(simplex):
            face = simplex[:j] + simplex[j + 1:]
            face_index = complex.index(face)
            boundary_matrix[i, face_index] = 1 if j % 2 == 0 else -1

    r = [0] * (m + 1)
    for i in range(m):
        r[i + 1] = np.linalg.matrix_rank(boundary_matrix[b[i + 1]:b[i + 2], b[i]:b[i + 1]])

    betti = [0] * (m + 1)
    if m == 0:
        betti[0] = b[1] - b[0]
    elif m == 1:
        betti[0] = b[1] - b[0] - r[0] - r[1]
        betti[m] = b[2] - b[1] - r[1]
    else:
        for i in range(m):
            betti[i] = b[i + 1] - b[i] - r[i] - r[i + 1]

    return betti

# Define the abstract simplicial complex as a list of simplices
complex = [[1], [2], [3], [4], [5], [1, 2], [2, 3], [2, 4], [3, 4], [2, 3, 4]]

# Calculate Betti numbers using the function
betti_numbers = compute_betti(complex)

# Print the Betti numbers
print("Betti Numbers:")
print(betti_numbers)

```

Figure 7: The final code for computing Betti numbers.

2.2 Computing Dirac and Laplacian matrices

In this section, we further guide ChatGPT to generate the codes for computing Dirac and Laplacian operators of a simplicial complex.

Let $\mathcal{K}$ be a simplicial complex, and let $C_*(\mathcal{K})$ be chain complex of $\mathcal{K}$. One can endow $C_*(\mathcal{K})$ with an inner product structure given by

$$\langle \sigma, \tau \rangle = \begin{cases} 1, & \sigma = \tau; \\ 0, & \text{otherwise.} \end{cases}$$

Then we have the adjoint operator $d^* : C_*(\mathcal{K}) \rightarrow C_*(\mathcal{K})$ of d with respect to the above inner product. The *Dirac operator* $\delta : C_*(\mathcal{K}) \rightarrow C_*(\mathcal{K})$ on $\mathcal{K}$ is defined by $\delta = d + d^*$. The *Laplacian operator* $\Delta : C_*(\mathcal{K}) \rightarrow C_*(\mathcal{K})$ on $\mathcal{K}$ is defined by $\Delta = \delta^2 = d^*d + dd^*$. As we know, the eigenvalues of Laplacian operators are crucial features for capturing the shape of a data-set. The number of zero eigenvalues is equal to the Betti number, and the second smallest eigenvalue, called the Fiedler vector, corresponds to the connectivity of the simplicial complex. The smallest positive eigenvalue, denoted by $\lambda(1)$ is the spectral gap, which is closely related the Cheeger constant in differential geometry. The spectral gap $\lambda(1)$ is also one of the most commonly used features in applications.

Let B_p be the representation matrix of $d_p : C_p(\mathcal{K}) \rightarrow C_{p-1}(\mathcal{K})$. More precisely, we write

$$d_p \begin{pmatrix} e_p^1 \\ \vdots \\ e_p^{k_p} \end{pmatrix} = B_p \begin{pmatrix} e_{p-1}^1 \\ \vdots \\ e_{p-1}^{k_{p-1}} \end{pmatrix}$$

Here, $e_p^1, \dots, e_p^{k_p}$ is the basis of $C_p(\mathcal{K})$. Then the representation matrix of the p -dimensional Laplacian operator $\Delta_p : C_p(\mathcal{K}) \rightarrow C_p(\mathcal{K})$ can be written as

$$L_p = B_p B_p^T + B_{p+1}^T B_{p+1}, \quad p \geq 1.$$

In particular, $L_0 = B_1 B_1^T$. It is worth noting that if the representation matrix B_p for d_p is given by right multiplication, the corresponding Laplacian matrix is $L_p = B_p^T B_p + B_{p+1} B_{p+1}^T$.

After we provided the matrix representation of Dirac and Laplacian operators to ChatGPT, it immediately generates the desire code for computing the corresponding Dirac and Laplacian matrices. The interaction with ChatGPT is shown in Fig 8 and Fig. 9. Although ChatGPT has been shown to perform relatively well in complex mathematical computations, its understanding of mathematical formulas is satisfactory.

Q: The Dirac matrix is given by the sum of the boundary matrix and its transpose matrix. Can you provide the function for computing the Dirac matrix in coding?

🌀 Certainly! You can compute the Dirac matrix by summing the boundary matrix and its transpose. Here's a Python function to do that:

```
.....  
def compute_dirac_matrix(complex):  
    # Compute the boundary matrix  
    boundary_matrix = compute_boundary_matrix(complex)  
    # Compute the Dirac matrix by adding the boundary matrix and its transpose  
    dirac_matrix = boundary_matrix + boundary_matrix.T  
    return dirac_matrix  
.....
```

This code defines a `'compute_dirac_matrix'` function that takes your abstract simplicial complex as input, computes the boundary matrix using the previously defined `'compute_boundary_matrix'` function, and then computes the Dirac matrix by adding the boundary matrix to its transpose.

Figure 8: The added code for Dirac matrix.

Q: The Laplacian matrix is given by the Dirac matrix multiplies its transpose. Could you please provide the Laplacian matrix in coding?

🌀 To compute the Laplacian matrix, you can multiply the Dirac matrix by its transpose. Here's a Python function to do that:

```
.....  
def compute_laplacian_matrix(complex):  
    # Compute the Dirac matrix  
    dirac_matrix = compute_dirac_matrix(complex)  
    # Compute the Laplacian matrix by multiplying the Dirac matrix by its  
    transpose  
    laplacian_matrix = np.dot(dirac_matrix, dirac_matrix.T)  
    return laplacian_matrix  
.....
```

This code defines a `'compute_laplacian_matrix'` function that takes your abstract simplicial complex as input, computes the Dirac matrix using the previously defined `'compute_dirac_matrix'` function, and then computes the Laplacian matrix by multiplying the Dirac matrix by its transpose.

Figure 9: The added code for Laplacian matrix.

3 ChatGPT for hypergraphs and directed graphs

Topological hypergraph and digraph (directed graph) are relatively new mathematical models for studying the complex systems in applications. From a combinatorial perspective, hypergraphs are an extension

of graphs where edges can be of higher dimension. From a geometric standpoint, hypergraphs are a generalization of abstract simplicial complexes, relaxing the condition of face operator closure that is inherent in simplicial complexes. In [BLRW19], the authors introduced the embedded homology hypergraphs, which are exactly a generalization of the simplicial homology of simplicial complexes. In this section, the topology invariants for hypergraphs is based on the embedded homology. The GLMY theory was introduced to study the topology of directed graphs [GLMY12, GLMY14, GLMY15]. Specifically, simplicial complexes are essential path complexes on digraphs, and the path homology can also be viewed as a generalization of simplicial homology. The embedded homology and GLMY theory provide us with a broader perspective and general tools to address practical problems.

3.1 Betti numbers for hypergraphs

Let us recall the embedded homology of hypergraphs. Let V be a finite and ordered nonempty set. Let $\mathbf{P}(V)$ be the power set of V . A *hypergraph* on V is a pair $(V, \mathcal{H})$ such that $\mathcal{H}$ is a nonempty subset of $\mathbf{P}(V)$. The *simplicial closure* is defined by $\Delta\mathcal{H} = \{\sigma | \sigma \subseteq \tau \in \mathcal{H}\}$, which can be viewed as the minimal simplicial complex contain $\mathcal{H}$.

Let $\mathbb{K}$ be the ground field. Let $C_*(\Delta\mathcal{H}; \mathbb{K})$ be the chain complex of the simplicial complex $\Delta\mathcal{H}$. Let $D_*(\mathcal{H}; \mathbb{K})$ be the graded subspace of $C_*(\Delta\mathcal{H}; \mathbb{K})$ generated by the hyperedges in $\mathcal{H}$. Then one has the *infimum chain complex*

$$\text{Inf}_*(\mathcal{H}; \mathbb{K}) = \sum \{C'_* \subseteq D_*(\mathcal{H}; \mathbb{K}) | C'_* \text{ is a subchain complex of } C_*(\Delta\mathcal{H}; \mathbb{K})\}$$

and the *supremum chain complex*

$$\text{Sup}_*(\mathcal{H}; \mathbb{K}) = \bigcap \{C'_* \supseteq D_*(\mathcal{H}; \mathbb{K}) | C'_* \text{ is a subchain complex of } C_*(\Delta\mathcal{H}; \mathbb{K})\}$$

of $\mathcal{H}$. It is proved that the homology of the infimum chain complex is isomorphic to the supremum chain complex of $\mathcal{H}$. The *embedded homology* is defined to be

$$H_p(\mathcal{H}; \mathbb{K}) := H_p(\text{Inf}_*(\mathcal{H}; \mathbb{K})), \quad p \geq 0.$$

Now, we will sketch the algorithm for computing the Betti numbers for hypergraphs. By [BLRW19, Proposition 2.1], one has that $\text{Inf}_*(\mathcal{H}; \mathbb{K}) = D_*(\mathcal{H}; \mathbb{K}) \cap d^{-1}D_*(\mathcal{H}; \mathbb{K})$. Here, d is the differential on $C_*(\Delta\mathcal{H}; \mathbb{K})$. In this work, the calculation of embedded homology is based on the chain complex $D_*(\mathcal{H}; \mathbb{K}) \cap d^{-1}D_*(\mathcal{H}; \mathbb{K})$. By [LCLW22, Theorem 2.3], we have

$$\beta_p(\mathcal{H}) = \dim D_p(\mathcal{H}; \mathbb{K}) \cup dD_{p+1}(\mathcal{H}; \mathbb{K}) - \dim dD_{p+1}(\mathcal{H}; \mathbb{K}) - \dim dD_p(\mathcal{H}; \mathbb{K}).$$

Let B_p be the representation matrix of the differential $d_p : D_p(\mathcal{H}; \mathbb{K}) \rightarrow C_{p-1}(\Delta\mathcal{H}; \mathbb{K})$. Then the Betti numbers can be computed by

$$\beta_p(\mathcal{H}) = \text{rank} \begin{pmatrix} E_p \\ B_p \end{pmatrix} - \text{rank } B_{p+1} - \text{rank } B_p. \quad (2)$$

Here, E_p is the matrix of p -hyperedges. We need to compute the boundary matrix B_p and the hyperedge matrix E_p . The algorithm for computing the boundary matrix of a hypergraph is almost identical to the algorithm for computing the boundary matrix of a simplicial complex. The key difference lies in the induces of hyperedges within the simplicial closure of the hypergraph. More precisely, we should find the index of each hyperedge in the list of the simplices of the simplicial complex $\Delta\mathcal{H}$. Besides, we also need the index of the boundary of each hyperedge. Based on the reasons mentioned above, we posed the question to ChatGPT in Fig. 10.

Q: Find the indices of each element in a subset within another ordered set's subset and provide Python code for it.

 To find the indices of each element in a subset within another ordered set's subset in Python, you can use list comprehensions and the `'enumerate()'` function. Here's a Python code example to achieve this:

```
def find_indices_of_subset(main_set, subset):
    indices = [i for i, x in enumerate(main_set) if x in subset]
    return indices

# Example usage:
main_set = [1, 2, 3, 4, 5, 6, 7, 8, 9]
subset = [3, 5, 7]

indices = find_indices_of_subset(main_set, subset)
print("Indices of subset elements in the main set:", indices)
```

In this code:

1. We define a function `'find_indices_of_subset'` that takes two lists as input, `'main_set'` and `'subset'`.
2. We use a list comprehension to iterate over the elements in the `'main_set'` along with their indices using the `'enumerate()'` function.
3. For each element `'x'` in `'main_set'`, we check if it exists in the `'subset'`. If it does, we append its index `'i'` to the `'indices'` list.
4. Finally, we return the `'indices'` list, which contains the indices of elements in the `'subset'` that appear in the `'main_set'`.

Figure 10: ChatGPT helps us find the index of a subset within an ordered set.

As shown Fig. 8 and Fig. 9, we can obtain the code for computing the Betti numbers for hypergraphs based on Eq. (2). In the progress, we need to validate whether the examples computed by the algorithm are correct. We examined the computed results at various locations in the code and determined which step of the calculation went wrong. We also tried to understand what caused these errors. While ChatGPT plays a significant role in writing and explaining code, our primary expertise lies in comprehending mathematical principles, and our responsibility is to oversee ChatGPT rather than engage in coding.

3.2 Betti numbers for directed graphs

Directed graphs (digraphs) naturally possess a 1-complex structure but lack higher-dimensional structure. However, the GLMY theory can detect the higher-dimensional structure of digraphs and compute their higher-order homology, referred to as path homology. Persistent path homology has been applied in various field [CLW⁺23, CHY22, CM18].

From now on, the Betti numbers for digraphs always refer to the Betti numbers based on the path homology. The algorithm for computing the Betti numbers for digraphs is quite similar. Recall that a digraph $G = (V, E)$ consists of a finite nonempty set V , called the vertex set, and the edge set $E \subseteq V \times V$. From now on, all the digraphs considered are assumed to be simple, that is, there is no loop or multi-edge in the same direction. An *elementary p -path* on V is a sequence $i_0 i_1 \cdots i_p$ of vertex in V . Let $\Lambda_p(V; \mathbb{K})$ be the K -linear space generated by all the elementary p -paths on V . With special attention, the basis for $\Lambda(V; \mathbb{K})$ is denoted by $\{e_{i_0 i_1 \cdots i_p}\}$, where $i_0 i_1 \cdots i_p$ runs across all the elementary p -paths on V . It follows

that $\Lambda_*(V; \mathbb{K})$ is a chain complex with the differential $d_p : \Lambda_p(V; \mathbb{K}) \rightarrow \Lambda_{p-1}(V; \mathbb{K})$ given by

$$d_p e_{i_0 i_1 \dots i_p} = \sum_{t=0}^p (-1)^t e_{i_0 \dots \hat{i}_t \dots i_p}, \quad p \geq 1$$

and $d_0 e_{i_0} = 0$. Here, $\hat{i}_t$ means omission of the index i_t . An *allowed p -path* on a digraph G is an elementary path $i_0 i_1 \dots i_p$ such that $(i_{t-1}, i_t) \in E$ for $t = 1, \dots, p$. Let $\mathcal{A}_p(G; \mathbb{K})$ be the $\mathbb{K}$ -linear space generated by all the allowed p -paths on G . Then $\mathcal{A}_p(G; \mathbb{K})$ is a subspace of $\Lambda_p(V; \mathbb{K})$. Let

$$\Omega_p(G; \mathbb{K}) = \{x \in \mathcal{A}_p(G; \mathbb{K}) \mid d_p x \in \mathcal{A}_{p-1}(G; \mathbb{K})\}.$$

Then $\Omega_*(G; \mathbb{K})$ is a chain complex. The *path homology* of G is defined by

$$H_p(G; \mathbb{K}) := H_p(\Omega_*(G; \mathbb{K})), \quad p \geq 0.$$

It is worth noting that the chain complex $\Omega_*(G; \mathbb{K})$ is exactly the infimum chain complex of $\mathcal{A}_*(G; \mathbb{K})$ in $\Lambda_*(V; \mathbb{K})$. The construction of $\Omega_*(G; \mathbb{K})$ for digraphs is quite similar as the construction of $\text{Inf}_*(\mathcal{H}; \mathbb{K})$ for hypergraphs. This similarity determines that we can use algorithms for calculating Betti numbers of hypergraphs to compute the Betti numbers of directed graphs in GLMY theory. The main difference in computing the Betti numbers of a directed graph is that we need to find all paths and represent these paths as lists, which is consistent with the representation of simplices in simplicial complexes or hyperedges in hypergraphs. It is worth noting that there may be simplified algorithms for applying the boundary operator to paths in view of the specific characteristics of paths.

Here, let us briefly outline the logic of how we will approach by asking ChatGPT to demonstrate how it can gradually help us write the code we need: Given a directed graph, provide a list of paths.

- (1) Question: Can you give me a Python code for finding all the directed path on a digraph?
ChatGPT gives us the all the paths for a given starting node and ending node.
- (2) Question: I need to obtain all paths, regardless of the starting and ending nodes.
When we run the code, it does not show the 0-dimensional paths.
- (3) Question: Please add vertices as path of length 0 into the 'paths'.
Then ChatGPT provides the satisfactory code.

Besides, we also request a code for finding the path with the length not larger than 2 or an integer number. Indeed, a finite digraph, for example, $V = \{0, 1\}, E = \{(0, 1), (1, 0)\}$, there are infinite many paths which behave as the binary numbers.

3.3 Computing hypergraph Laplacian

In this section, we will study the algorithms for computing the Dirac and Laplacian matrices of on hypergraphs and digraphs. We will present the mathematical progress for hypergraphs first. The algorithm for digraphs is much similar.

Let $\mathcal{H}$ be a hypergraph. Recall that we have the chain complex $C_*(\Delta\mathcal{H}; \mathbb{K})$ and its subspace $D_*(\mathcal{H}; \mathbb{K})$. Let $e_p^1, \dots, e_p^{l_p}$ and $\varepsilon_p^1, \dots, \varepsilon_p^{k_p}$ be the basis of $C_*(\Delta\mathcal{H}; \mathbb{K})$ and $D_*(\mathcal{H}; \mathbb{K})$, respectively. We write the differential $d_p : D_p(\mathcal{H}; \mathbb{K}) \rightarrow C_{p-1}(\Delta\mathcal{H}; \mathbb{K})$ by

$$d_p \begin{pmatrix} \varepsilon_p^1 \\ \vdots \\ \varepsilon_p^{k_p} \end{pmatrix} = B_p \begin{pmatrix} e_p^1 \\ \vdots \\ e_p^{l_p} \end{pmatrix}.$$

Let $\pi_p : C_p(\Delta\mathcal{H}; \mathbb{K}) \rightarrow C_p(\Delta\mathcal{H}; \mathbb{K})/D_p(\mathcal{H}; \mathbb{K})$ and $\rho_p : C_p(\Delta\mathcal{H}; \mathbb{K}) \rightarrow D_p(\mathcal{H}; \mathbb{K})$ be projections of $\mathbb{K}$ -linear spaces. Then we have $\text{Inf}_p(\mathcal{H}; \mathbb{K}) = \ker(\pi \circ d_p)$. Let $\tilde{B}_p$ be the representation matrix of $\pi \circ d_p$. Let

$A_p \begin{pmatrix} \varepsilon_p^1 \\ \vdots \\ \varepsilon_p^{k_p} \end{pmatrix}$ be an orthogonal basis of $\text{Inf}_p(\mathcal{H}; \mathbb{K})$. Then we have

$$\pi \circ d_p \left(A_p \begin{pmatrix} \varepsilon_p^1 \\ \vdots \\ \varepsilon_p^{k_p} \end{pmatrix} \right) = A_p \left(\pi \circ d_p \begin{pmatrix} \varepsilon_p^1 \\ \vdots \\ \varepsilon_p^{k_p} \end{pmatrix} \right) = A_p \tilde{B}_p \begin{pmatrix} e_p^1 \\ \vdots \\ e_p^{l_p} \end{pmatrix}.$$

Note that A_p is an $r_p \times k_p$ matrix. Here, k_p is the dimension of $\text{Inf}_p(\mathcal{H}; \mathbb{K})$. The algorithm to obtain the matrix A_p is to calculate the left zero matrix of $\tilde{B}_p$. Let $\tilde{B}_p$ be the representation matrix of $\rho \circ d_p$. Based on the discussion above, we have

$$d_p \left(A_p \begin{pmatrix} \varepsilon_p^1 \\ \vdots \\ \varepsilon_p^{k_p} \end{pmatrix} \right) = \rho \circ d_p \left(A_p \begin{pmatrix} \varepsilon_p^1 \\ \vdots \\ \varepsilon_p^{k_p} \end{pmatrix} \right) = A_p \left(\tilde{B}_p \begin{pmatrix} \varepsilon_p^1 \\ \vdots \\ \varepsilon_p^{k_p} \end{pmatrix} \right) = A_p \tilde{B}_p A_{p-1}^{-1} \left(A_{p-1} \begin{pmatrix} \varepsilon_p^1 \\ \vdots \\ \varepsilon_p^{k_p} \end{pmatrix} \right).$$

Here, A_p^{-1} is the generalized inverse or Moore-Penrose matrix of A_p , i.e., $A_p^{-1} A_p = I_{r_p}$. Then $A_p \tilde{B}_p A_{p-1}^{-1}$ is the desired boundary matrix of the infimum complex. Furthermore, the Laplacian matrix can be computed by

$$L_p = A_p \tilde{B}_p A_{p-1}^{-1} (A_p \tilde{B}_p A_{p-1}^{-1})^T + (A_{p+1} \tilde{B}_{p+1} A_p^{-1})^T A_{p+1} \tilde{B}_{p+1} A_p^{-1}, \quad p \geq 1.$$

In particular, $L_0 = (A_1 \tilde{B}_1 A_0^{-1})^T A_1 \tilde{B}_1 A_0^{-1}$. The eigenvalues of Laplacian matrices are independent of the choices of the orthogonal for the infimum chain complex.

Next, we guide ChatGPT to generate the code for computing the boundary matrix for the infimum chain complex $\text{Inf}_p(\mathcal{H}; \mathbb{K})$. The algorithm is based on the code for computing Betti numbers. Recall that we have obtained the code for computing the boundary matrix B_p in Section 3.1.

Let I be the list of the index of the hyperedges of $\mathcal{H}$ in the set of simplices of $\Delta\mathcal{H}$. The matrix $\tilde{B}$ is obtained by removing the columns indexed by I , while the matrix $\tilde{B}$ consists of the columns indexed by I .

Q: With a given a matrix and index list, please give me a Python code to obtain two new matrices by removing the columns indexed by the list and not indexed by the list.

 To obtain two new matrices from an input matrix by removing columns based on an index list, you can use Python and NumPy. Here's a Python code snippet to achieve this:

```
import numpy as np

def split_matrix(matrix, index_list):
    # Create an array of boolean values indicating which columns to keep
    keep_columns = np.zeros(matrix.shape[1], dtype=bool)
    keep_columns[index_list] = True

    # Use boolean indexing to split the matrix
    matrix_with_indexed_columns = matrix[:, keep_columns]
    matrix_without_indexed_columns = matrix[:, ~keep_columns]

    return matrix_with_indexed_columns, matrix_without_indexed_columns

# Example usage:
# Create a sample matrix and index list for testing
sample_matrix = np.array([[1, 2, 3, 4], [5, 6, 7, 8], [9, 10, 11, 12]])

index_list = [0, 2] # Specify the indices of columns to keep

# Call the function to split the matrix
result_with_indexed, result_without_indexed = split_matrix(sample_matrix,
index_list)

# Print the results
print("Matrix with indexed columns:")
print(result_with_indexed)

print("\nMatrix without indexed columns:")
print(result_without_indexed)
```

In this code, 'split_matrix' takes an input matrix and an index list as arguments and returns two new matrices: one with columns indexed by the list and another with columns not indexed by the list. It uses NumPy to perform the column selection efficiently.

Figure 11: ChatGPT helps us split a matrix into two matrices.

Then there are three tasks left for us:

- (1) Compute the orthogonal left zero matrix A_p of $\bar{B}_p$;
- (2) Compute the generalized inverse matrix A_p^{-1} of A_p ;
- (3) Compute the Laplacian matrices for hypergraphs.

Fig. 12 gives the code for computing orthogonal left zero matrix, while the generalized inverse matrix is provided in Fig. 13. We combine all the functions together and obtain the final code for computing the Laplacians for hypergraphs.

Q: Given a matrix B , please find a matrix A such that $AB = 0$ and the rows of A are orthogonal to each other. Please give me a Python code.

..... (Not desired content is omitted here.)

Q: Suppose B is an $m \times n$ matrix, where $m \leq n$. Please find a $r \times m$ matrix A as above. Here, $r = m - \text{rank}(B)$.

If you want to find a matrix A with r rows (where $r = m - \text{rank}(B)$) such that $AB = 0$ and the rows of A are orthogonal to each other, you can use the concept of the left null space of matrix B . The left null space contains vectors that are orthogonal to the rows of B . Here's a Python code to do that:

```
import numpy as np

# Define matrix B
B = np.array([[1, 2, 3],[4, 5, 6],[7, 8, 9]])

# Calculate the rank of matrix B
rank_B = np.linalg.matrix_rank(B)

# Calculate r (number of rows in A)
r = B.shape[0] - rank_B

# Perform SVD on B to get the left singular vectors (V matrix in SVD)
U, S, VT = np.linalg.svd(B)

# Extract the left null space vectors from VT
left_null_space_vectors = VT[rank_B:].T

# Create matrix A with orthogonal rows from the left null space vectors
A = left_null_space_vectors[:r]

# Verify that A is orthogonal and AB = 0
result = np.dot(A, B)
print("Matrix A:")
print(A)
print("\nMatrix AB:")
print(result)
```

In this code:

1. We define matrix B as given.
2. We calculate the rank of matrix B using '`numpy.linalg.matrix_rank`'.
3. We calculate r , which is $m - \text{rank}(B)$.
4. We perform Singular Value Decomposition (SVD) on B to get the left singular vectors (V matrix in SVD).
5. We extract the left null space vectors from VT (transpose of V).
6. We create matrix A with r rows from the left null space vectors.
7. We verify that A is orthogonal by calculating '`A * A.T`', which should be the identity matrix if A is orthogonal.
8. We also calculate AB to ensure that $AB = 0$. 20

Make sure you have NumPy installed ('`pip install numpy`') to run this code.

Figure 12: ChatGPT helps us find the orthogonal left zero matrix of a given matrix. Note that the code cannot run smoothly before debugging.

Q: Please give me a Python code for computing the right generalized inverse matrix of a given matrix.

 To compute the right generalized inverse (also known as the Moore-Penrose inverse) of a given matrix in Python, you can use the numpy library, which provides a convenient function for this purpose. Here's how you can do it:

```
import numpy as np

# Define your matrix
A = np.array([[1, 2, 3],[4, 5, 6]])

# Compute the right generalized inverse (Moore-Penrose inverse)
A_inv = np.linalg.pinv(A)

# Print the result
print("Right Generalized Inverse:")
print(A_inv)
```

In this code:

1. We import the numpy library as **'np'**.
2. We define your matrix **'A'**. You can replace this matrix with the one you want to compute the right generalized inverse for.
3. We use the **'np.linalg.pinv()'** function to compute the right generalized inverse of **'A'**. This function automatically calculates the Moore-Penrose inverse.
4. We print the result, which is the right generalized inverse of the input matrix **'A'**.

Make sure to have numpy installed in your Python environment. You can install it using pip if it's not already installed:

```
pip install numpy
```

Replace the **'A'** matrix with your desired matrix to compute its right generalized inverse.

Figure 13: ChatGPT can provide codes for computing the generalized inverse matrix.

3.4 Computing path Laplacian

Path Laplacian and persistent path Laplacian have also been introduced to reveal crucial structured characteristic that cannot be detected by (persistent) path homology [WW23]. Like in Section 3.2, the code for computing the hypergraph Laplacian is equally applicable for calculating the path Laplacian of a digraph.

Let $G = (V, E)$ be a digraph. We have a chain complex $\Omega_*(G; \mathbb{K})$ with the differential $d_p : \Omega_p(G; \mathbb{K}) \rightarrow \Omega_{p-1}(G; \mathbb{K})$. The *path Laplacian* $\Delta_p : \Omega_p(G; \mathbb{K}) \rightarrow \Omega_p(G; \mathbb{K})$ is defined by

$$\Delta_p = d_p^* \circ d_p + d_{p+1} \circ d_{p+1}^*, \quad p \geq 1.$$

In particular, $\Delta_0 = d_1 \circ d_1^*$. The number of zero eigenvalues of Δ_p , known as the harmonic information, is equal to the Betti number β_p from the path homology. The non-harmonic information is expressed as the positive eigenvalues of Δ_p . Note that the smallest positive eigenvalue of Δ_p is one of the most used feature in illustrating the difference between path homology and path Laplacian.

To compute the representation matrix of Laplacian operator Δ_p , it suffices to find the representation matrix of the boundary d_p . This can be done similar as in Section 3.3. We have also obtained a functional

code for computing path Laplacians.

3.5 Computing hyperdigraph Laplacian

Topological hypergraphs are highly versatile models capable of characterizing high-dimensional data structures. Topological hyperdigraphs can be viewed as assigning additional directions to each hyperedge, thereby encompassing more information and considered a more generalized mathematical model. Recently, the persistent hyperdigraph Laplacians have been introduced to characterize the topological information of complex systems [CLWW23].

Let V be a finite and ordered nonempty set. Let $\mathbf{S}(V)$ be the set of all sequence of distinct elements in V . A *hyperdigraph* on V is a pair $(V, \vec{\mathcal{H}})$ such that $\vec{\mathcal{H}}$ is a nonempty subset of $\mathbf{S}(V)$. A sequence in $\vec{\mathcal{H}}$ is called *directed hyperedge*. Hyperdigraph can be viewed as a kind of generalization of hypergraphs or digraphs.

Let $S_p(V; \mathbb{K})$ be the $\mathbb{K}$ -linear space generated by the $(p+1)$ -element sets in $\mathbf{S}(V)$. Then $S_*(V; \mathbb{K})$ is a chain complex with the boundary operator $d_p : S_p(V; \mathbb{K}) \rightarrow S_{p-1}(V; \mathbb{K})$ given by

$$d_p(v_0, \dots, v_p) = \sum_{i=1}^p (-1)^i (v_0, \dots, \hat{v}_i, \dots, v_p), \quad p \geq 0.$$

Here, $\hat{v}_i$ means the omission of the term v_i . In particular, $d_0 = 0$ on $S_0(V; \mathbb{K})$. Let $F_p(\vec{\mathcal{H}}; \mathbb{K})$ be the $\mathbb{K}$ -linear space generated all the directed hyperedges in $\vec{\mathcal{H}}$. Let

$$\Omega_p(\vec{\mathcal{H}}; \mathbb{K}) = \{x \in F_p(\vec{\mathcal{H}}; \mathbb{K}) \mid d_p x \in F_{p-1}(\vec{\mathcal{H}}; \mathbb{K})\}.$$

Then $\Omega_*(\vec{\mathcal{H}}; \mathbb{K})$ is a chain complex. The homology of $\vec{\mathcal{H}}$ is defined by

$$H_p(\vec{\mathcal{H}}; \mathbb{K}) := H_p(\Omega_p(\vec{\mathcal{H}}; \mathbb{K})), \quad p \geq 0.$$

We endow the directed hyperedges with the standard inner product structure. The Laplacian $\Delta_p : \Omega_p(\vec{\mathcal{H}}; \mathbb{K}) \rightarrow \Omega_p(\vec{\mathcal{H}}; \mathbb{K})$ is given by

$$\Delta_p = d_p^* \circ d_p + d_{p+1} \circ d_{p+1}^*.$$

The representation of a hyperdigraph in the code is provided as a collection of lists, with each list representing a directed hyperedge. The Betti numbers and Laplacian matrices for hyperdigraphs can be computed in a similar manner as for hypergraphs. The ChatGPT code obtained with our supervision can also be used to compute the hyperdigraph Betti numbers and hyperdigraph Laplacians.

4 ChatGPT for topological persistence

The development of applied topology is profoundly benefited by topological persistence. Topological persistence can characterize multi-scale topological information in data sets. Without persistence, homology theory by itself would have little application value. Nowadays, persistent homology and persistent Laplacian are some of the most popular TDA tools. The data set considered is always a family of discrete points in a metric space. Given a data set, one can compute the distance between any two of the points with a specific metric, and the persistence is based on the distance. The Vietoris-Rips complex or Rips complex is the standard construction of the persistence from a data set. There have been numerous studies on persistent homology [CZCG04, ELZ02, EH⁺08, ZC04], persistent Laplacians [CZTW19, LLW23, MWW22, WNW20], and persistent Dirac operators [AMS22, WBX23].

4.1 ChatGPT for persistent homology and persistent Laplacians

Consider the simplicial complex $\mathcal{K}_\varepsilon$ parametrized by a real number ε such that $\mathcal{K}_\varepsilon \hookrightarrow \mathcal{K}_{\varepsilon'}$ for any $\varepsilon \leq \varepsilon'$. For $a \leq b$, we have a morphism of chain complexes $j_{*,b}^{a,b} : C_*(\mathcal{K}_a) \hookrightarrow C_*(\mathcal{K}_b)$, which induces a morphism of

homology groups $H_*(\mathcal{K}_a) \rightarrow H_*(\mathcal{K}_b)$. The (a, b) -persistent homology is defined by

$$H_p^{a,b} := \text{Im}(H_p(\mathcal{K}_a) \rightarrow H_p(\mathcal{K}_b)), \quad p \geq 0.$$

The rank of the persistent homology $\beta_p^{a,b} = \text{rank } H_p^{a,b}$ is the (a, b) -persistent Betti number, which is always represented by barcode or persistence diagram in application. In particular, we have $\beta_p^{a,a} = \beta_p^a$.

The (a, b) -persistent Laplacian $\Delta_p^{a,b} : C_*(\mathcal{K}_a) \rightarrow C_*(\mathcal{K}_a)$ is defined by

$$\Delta_p^{a,b} = (d_p^a)^* \circ d_p^a + (d_{p+1}^{a,b})^* \circ d_{p+1}^{a,b}, \quad p \geq 0.$$

Here, d_p^a is the boundary operator of $C_*(\mathcal{K}_a)$, and $d_{p+1}^{a,b}$ is the composition of

$$\{x \in C_{p+1}(\mathcal{K}_b) \mid d_{p+1}^b x \in C_{p+1}(\mathcal{K}_a)\} \hookrightarrow C_{p+1}(\mathcal{K}_b) \xrightarrow{d_{p+1}^b} C_p(\mathcal{K}_a).$$

In particular, when $a = b$, the persistent Laplacian $\Delta_p^{a,a}$ is exactly the Laplacian Δ_p^a on chain complex $C_*(\mathcal{K}_a)$.

Persistent homology and persistent Laplacians can also be defined for hypergraphs, digraphs, and hyperdigraphs [CLWW23]. This allows for the characterization of multiscale topological information in these structures.

In applications, one often focuses on the 0th and 1st Betti numbers because they are more computationally tractable. Sometimes, researchers only compute the Betti numbers at each parameter and get the Betti curves β_0^a and β_1^a . If the barcode obtained from the data set has a large number of bars, then using the persistent Betti number is often a reasonable choice.

When it comes to persistent Laplacians, researchers seldom find it necessary to compute the persistent Laplacian $\Delta_p^{a,b}$ for all $a \leq b$ since it would make the calculation overly complex. In most cases, people only compute the Laplacians Δ_0^a and Δ_1^a . It is yet to demonstrate that computationally sophisticated persistence offers any real benefit for data analysis. Therefore, we might misuse the term “persistence”, which is largely due to computational considerations and does not affect the theoretical notion of persistence. In our examples, we have not provided the ChatGPT code obtained with our supervision for computing the barcodes and (a, b) -persistent Laplacians for $a < b$. We only calculate the Betti numbers and Laplacians at various filtration parameters.

4.2 ChatGPT for persistent harmonic space

It is proved that the persistent harmonic space for a filtration of simplicial complexes is naturally isomorphic to the persistent homology for the filtration [LLW23, Theorem 3.6]. It is worth noting that the basis for the persistent homology is a family of equivalent classes, which give rise to an obstruction for computing the generators in the persistent homology. However, the persistent harmonic space has a good property that the basis of the space is unique under linear isomorphisms, which provides us with a chance to find the basis of the persistent harmonic space and to track the generators as the persistence parameter varies.

Although there are many existing algorithms for computing persistent homology, there is not algorithm for computing the persistent harmonic space, to the best of our knowledge. It is interesting and important to demonstrate that ChatGPT can develop unseen codes in computational topology.

Consider the simplicial complex $\mathcal{K}_\varepsilon$ parametrized by a real number ε such that $\mathcal{K}_\varepsilon \hookrightarrow \mathcal{K}_{\varepsilon'}$ for any $\varepsilon \leq \varepsilon'$. Recall that the (a, b) -persistent Laplacian $\Delta_p^{a,b} : C_*(\mathcal{K}_a) \rightarrow C_*(\mathcal{K}_a)$ is $\Delta_p^{a,b} = (d_p^a)^* \circ d_p^a + (d_{p+1}^{a,b})^* \circ d_{p+1}^{a,b}$. The (a, b) -persistent harmonic space is defined by

$$\mathcal{H}_p^{a,b} = \ker \Delta_p^{a,b}, \quad p \geq 0.$$

Note that there is a decomposition of $C_*(\mathcal{K}_a) = \mathcal{H}_*^a \oplus \text{Im } d^a \oplus \text{Im } (d^a)^*$, where $\mathcal{H}_*^a = \ker \Delta_*^a$. Let $p^a : C_*(\mathcal{K}_a) \rightarrow \mathcal{H}_*^a$ be the projection of $C_*(\mathcal{K}_a)$ to the harmonic part. By [LLW23, Proposition 3.8], one has an

isomorphism

$$\mathcal{H}_p^{a,b} \cong p^b j_*^{a,b}(\mathcal{H}_p^a) \quad (3)$$

for any $a \leq b$. This implies the persistence on the persistent harmonic space. There is a natural isomorphism between the persistent harmonic space and the persistent homology

$$\mathcal{H}_p^{a,b} \cong H_p^{a,b}, \quad p \geq 0.$$

To compute the persistent homology, it suffices to give the code for the persistent harmonic space.

Let B_p^a be the representation matrix of the boundary operator $d_p^a : C_p(\mathcal{K}_a) \rightarrow C_{p-1}(\mathcal{K}_a)$. The representation matrix of the adjoint operator $(d_p^a)^* : C_{p-1}(\mathcal{K}_a) \rightarrow C_p(\mathcal{K}_a)$ of d_p^a is given by the transpose matrix $(B_p^a)^T$ of B_p^a with respect to the standard orthogonal basis given by the simplices. Thus, the harmonic space $\mathcal{H}_p^a$ can be represented by the solution space W^a of the equation

$$W^a \begin{pmatrix} B_p^a & (B_{p+1}^a)^T \end{pmatrix} = \mathbf{0}.$$

All the generators in W^a have the index a as the birth time. When the time comes to $b \geq a$, we will determine if there are generators that die at b . Note that $j_*^{a,b}$ is an inclusion, by Eq. (3), we have that the (a, b) -persistent harmonic space $\mathcal{H}_p^{a,b}$ is the projection of $\mathcal{H}_p^a$ to $\mathcal{H}_p^b$. Note that $\mathcal{H}_p^a \subseteq \ker d_p^b$ and $\mathcal{H}_p^b = \ker d_p^b \cap \ker (d_{p+1}^b)^*$. Let U be the orthogonal complement space of $\mathcal{H}_p^{a,b}$ in $\mathcal{H}_p^a$. It follows that

$$(d_{p+1}^b)^* \mathcal{H}_p^a \cong U.$$

Thus, if $W^a (B_{p+1}^b)^T = \mathbf{0}$, there is no generator dead at b . Moreover, $\dim W^a (B_{p+1}^b)^T$ is the number of generators dead at b . The subspace of W^a corresponding to $W^a (B_{p+1}^b)^T$ are the space generated by the dead generators. The left generators contribute to the persistent harmonic space $\mathcal{H}^{a,b}$, represented by the matrix $W^{a,b}$.

Let E_p^a be the basis of $C_p(\mathcal{K}_a)$. Let $E_p^b = \begin{pmatrix} E_p^a \\ F_p^b \end{pmatrix}$ be the basis of $C_p(\mathcal{K}_b)$. Then we assume

$$d_p^b \begin{pmatrix} E_p^a \\ F_p^b \end{pmatrix} = \begin{pmatrix} B_p^a & \mathbf{0} \\ \alpha_p^b & \beta_p^b \end{pmatrix} \begin{pmatrix} E_{p-1}^a \\ F_{p-1}^b \end{pmatrix}.$$

Let $W_b = \begin{pmatrix} W^{a,b} & \mathbf{0} \\ W_1^b & W_2^b \end{pmatrix}$. It follows that the the harmonic space at b is given by $W^b \begin{pmatrix} B_p^b & (B_{p+1}^b)^T \end{pmatrix} = \mathbf{0}$.

Thus one has

$$\begin{pmatrix} W_1^b & W_2^b \end{pmatrix} \begin{pmatrix} B_p^a & \mathbf{0} & (B_{p+1}^a)^T & (\alpha_{p+1}^b)^T \\ \alpha_p^b & \beta_p^b & \mathbf{0} & (\beta_{p+1}^b)^T \end{pmatrix} = \mathbf{0}. \quad (4)$$

Note that if $(\mathbf{x}, \mathbf{0})$ is a solution vector, then $\mathbf{x} B_p^a = \mathbf{0}$ and $\mathbf{x} (\alpha_{p+1}^b)^T = \mathbf{0}$. It follows that $\mathbf{x}$ is a generator representation in $\mathcal{H}_p^a$. Thus, all the generators represented in the form $(\mathbf{x}, \mathbf{y})$ consist the solution space of Eq. (4), where $\mathbf{y} \neq \mathbf{0}$. Therefore, one can obtain the desired generators and mark them with the birth time b . By continuing this iteration, we can obtain the persistent harmonic space. In particular, the corresponding dimensions for the persistent harmonic space lead to the persistence diagram or barcode for the persistent harmonic space, which is the consistent with the persistence diagram of the corresponding persistent homology. In the example code of ChatGPT, we show the basic calculations for computing the 1-dimensional persistent harmonic space.

```
import vr_complex
import numpy as np
import boundary_matrix
```

```

def generate_complex(A, B, d):
    result = [A[i] for i, distance in enumerate(B) if distance < d]
return result
def rips_complex_generate(vr,a):
    rips_complex_1 = generate_complex(vr[1],vr[2],a)
    rips_complex_2 = generate_complex(vr[3],vr[4],a)
    rips_complex = vr[0] + rips_complex_1 + rips_complex_2
return rips_complex

def dim_index(hypergraph):
    num_hyperedges = len(hypergraph)
    m=len(hypergraph[num_hyperedges-1])+1
    m=max(m,4)
    b = [0] * m
    for i in range(num_hyperedges):
        b[len(hypergraph[i])]+=1;
    c = [0] * m
    for i in range(m-1):
        c[i+1]=c[i]+b[i+1]
    return c

def left_null_space(matrix):
    # Calculate the rank of the input matrix
    rank_matrix = np.linalg.matrix_rank(matrix)

    # Perform SVD to get the left singular vectors
    U, S, VT = np.linalg.svd(matrix)

    # Extract the left null space vectors from U
    left_null_space_vectors = U[:, rank_matrix:].T

    return left_null_space_vectors

def triangular_matrix(Ar):
    # Initialize the left multiplication matrix as the identity matrix
    La = np.identity(Ar.shape[0])
    # Iterate through each column and perform Gaussian elimination
    for j in range(Ar.shape[1]):
        for i in range(j + 1, Ar.shape[0]):
            # Calculate the steps of Gaussian elimination to obtain elements of L
            factor = Ar[i, j] / Ar[j, j]
            # Update elements of L
            La[i, j] = factor
            # Update rows of A to achieve elimination
            Ar[i, :] -= factor * Ar[j, :]
    return La, Ar

```

```

# Calculate the inverse transformation of the matrix row operations for L
def lower_triangular_inverse(A):
    n = A.shape[0]
    B = np.eye(n)

    if np.linalg.matrix_rank(A) < n:
        print("The matrix is singular. Try another one!")
        return B

    for k in range(n):
        B[0, k] = B[0, k] / A[0, 0]
        for i in range(1, n):
            B[i, k] = (B[i, k] - np.dot(A[i, :i], B[:i, k])) / A[i, i]
    return B

#for a harmonic_matrix, print the generators for the harmonic spaces
def print_harmonic_generators(rips_complex, harmonic_matrix):
    # Get the number of rows in the matrix
    num_rows = harmonic_matrix.shape[0]
    num_cols = harmonic_matrix.shape[1]
    #print the generators
    for i in range(num_rows):
        for j in range(num_cols):
            if j > 0 and harmonic_matrix[i][j]>0 and j<num_cols-1:
                print("+",round(harmonic_matrix[i][j],7),rips_complex[j],end="")
            elif harmonic_matrix[i][j] ==0 and j<num_cols-1:
                continue
            elif j<num_cols-1:
                print(round(harmonic_matrix[i][j],7),rips_complex[j],end="")
            elif i<num_rows-1:
                print(" and")
            elif harmonic_matrix[i][j]>0;
                print("+",round(harmonic_matrix[i][j],7),rips_complex[j])
            else:
                print(round(harmonic_matrix[i][j],7),rips_complex[j]))
    return

# since not each matrix can be triangulated, we use the gaussian elimination
def gaussian_elimination_partial_pivoting(matrix):
    m, n = matrix.shape
    for i in range(min(m, n)):
        # Use partial pivoting to select the main element
        max_row = np.argmax(np.abs(matrix[i:, i])) + i
        if matrix[max_row, i] == 0:
            continue # If a non-zero main element cannot be found, move to the
next column
        matrix[[i, max_row]] = matrix[[max_row, i]]
        for j in range(i + 1, m):
            factor = matrix[j, i] / matrix[i, i]
            matrix[j, i:] -= factor * matrix[i, i:]
    return matrix

```

```

discrete_points = np.array([[0,0,0],[0,1,0],
                             [1,0,0],[1,1,0],
                             [0,0,2],[0,1,2],
                             [1,0,2],[1,1,1.5]
                             ])

a = 1.1
b = 1.3
#Give the maximal VR complex with all 0,1,2-dimensional simplices
vr_list=vr_complex.vr_compute(discrete_points)

#generate the rips complexes at some parameters
RipsC_a = rips_complex.generate(vr_list,a)
RipsC_b = rips_complex.generate(vr_list,b)

print("The Rips complex at parameter a is:",RipsC_a)
print("The Rips complex at parameter b is:",RipsC_b)

#generate the boundary matrices for rips complexes
boundary_a=boundary_matrix.compute_boundary_matrix(RipsC_a)
boundary_b=boundary_matrix.compute_boundary_matrix(RipsC_b)

index_a = dim_index(RipsC_a)
index_b = dim_index(RipsC_b)

rips_a = RipsC_a[index_a[1]:index_a[2]]
rips_b = RipsC_b[index_b[1]:index_b[2]]

#compute W_a, which is the harmonic matrix at a,
#where dirac_1 is the union of boundary and coboundary matrices
dirac_1 = np.hstack((boundary_a[0], boundary_a[1].T))
W_a = left_null_space(dirac_1)

#If W_a is emptyset, the persistent harmonic space is none, or we compute the matrix
U in our paper
if W_a.size == 0:
    print("The harmonic space at a is none.")
    print("Then the harmonic space from a to b is also none.")
else:
    print("The harmonic space at a is generated by:")
    print_harmonic_generators(rips_a,W_a)

#U_matrix is the space of dead elements at b

#The number of one-dimensional simplices added in Rips complex at time b compared
to time a.
rank_2 = index_b[2]-index_b[1]-index_a[2]+index_a[1]

```

```

    #The number of points
    rank_1 = W_a.shape[0]
    zero_add = np.zeros((rank_1, rank_2))

    W_a_add = np.hstack((W_a, zero_add))

    U_matrix = W_a_add@(boundary_b[1].T)
    if U_matrix.size == 0:
        print("The harmonic space from a to b is generated by:")
        W_ab = W_a_add
        print_harmonic_generators(rips_b, W_ab)
    else:
        rank_U = np.linalg.matrix_rank(U_matrix)
        if rank_U == rank_1:
            print("The harmonic space from a to b is none.")
        else:
            # compute the persistent harmonic space matrix W_ab from a to b
            Left_triangle, triangle_matrix = triangular_matrix(U_matrix)
            copy_triangle = Left_triangle[:].copy()
            L_inv = lower_triangular_inverse(copy_triangle)
            W_temp = L_inv @ W_a
            W_ab = W_temp[:rank_U, :]

    #W_b is the harmonic matrix at b
    dirac_2 = np.hstack((boundary_b[0], boundary_b[1].T))
    W_b = left_null_space(dirac_2)

    if W_b.size == 0:
        print("The harmonic space at b is none.")
    else:
        print("The harmonic space at b is generated by:")
        print_harmonic_generators(rips_b, W_b)
        _, birth_b = triangular_matrix(W_b)
        reflect_matrix = np.fliplr(birth_b)
        upper_triangular_matrix = gaussian_elimination_partial_pivoting(reflect_matrix)
        rr_matrix = np.fliplr(upper_triangular_matrix)
        if W_a.size == 0:
            print("The harmonic generators birth at b is generated by:")
            print_harmonic_generators(rips_b, W_b)
        else:
            number_of_generators = W_b.shape[0] - W_ab.shape[0]
            print("The harmonic generators born at b is generated by:")
            print_harmonic_generators(rips_b, W_b[:number_of_generators, :])

```

Figure 14: There are ten functions appeared in the ChatGPT code. The functions `compute_boundary_matrix` and `vr_compute` are imported from the previous files `boundary_matrix` and `vr_complex`, respectively. Apart from the function `print_harmonic_generators`, the functions `rips_complex_generate` and `dim_index` were designed by ChatGPT with our guidance. All the other functions, including `generate_complex`, `left_null_space`, `triangular_matrix`, `lower_triangular_inverse`, and `gaussian_elimination_partial_pivoting`, were provided in responses by ChatGPT. Specifically, the main idea for the function `gaussian_elimination_partial_pivoting` was also provided by ChatGPT.

Example 4.1. Consider the point set X embedded into the Euclidean space $\mathbb{R}^3$ as follows.

$$X = \{(0, 0, 0), (0, 1, 0), (1, 0, 0), (1, 1, 0), (0, 0, 2), (0, 1, 2), (1, 0, 2), (1, 1, 1.5)\}.$$

We guided ChatGPT to complete the code for computing the 1-dimensional persistent harmonic space. We take the persistent parameter $a = 1.2$ and $b = 1.3$. Then the result window displays as follows.

The Rips complex at parameter a is: $[[0], [1], [2], [3], [4], [5], [6], [7], [0, 1], [0, 2], [1, 3], [2, 3], [4, 5], [4, 6]]$
The Rips complex at parameter b is: $[[0], [1], [2], [3], [4], [5], [6], [7], [0, 1], [0, 2], [1, 3], [2, 3], [4, 5], [4, 6], [5, 7], [6, 7]]$
The harmonic space at a is generated by: $0.5 [0, 1] - 0.5 [0, 2] + 0.5 [1, 3] - 0.5 [2, 3]$
The harmonic space from a to b is generated by: $0.5 [0, 1] - 0.5 [0, 2] + 0.5 [1, 3] - 0.5 [2, 3]$
The harmonic space at b is generated by: $-0.5 [0, 1] + 0.5 [0, 2] - 0.5 [1, 3] + 0.5 [2, 3]$ and $-0.5 [4, 5] + 0.5 [4, 6] - 0.5 [5, 7] + 0.5 [6, 7]$
The harmonic generators born at b is generated by: $-0.5 [4, 5] + 0.5 [4, 6] - 0.5 [5, 7] + 0.5 [6, 7]$

Figure 15: Shown results on persistent harmonic spaces (I).

Besides, for another example, X is the point set with the points $(0.5, \sqrt{3}/2, 0)$, $(0.5, \sqrt{3}/2, 0)$, $(-1, 0, 0)$, $(1, 0, 0)$, $(0.5, -\sqrt{3}/2, 0)$, $(-0.5, -\sqrt{3}/2, 0.3)$. We take $a = 1$ and $b = 1.2$. The result is shown in Fig. 16.

The Rips complex at parameter a is: $[[0], [1], [2], [3], [4], [5], [0, 3], [1, 2], [3, 4]]$
The Rips complex at parameter b is: $[[0], [1], [2], [3], [4], [5], [0, 3], [1, 2], [3, 4], [0, 1], [2, 5], [4, 5]]$
The harmonic space at a is none.
Then the harmonic space from a to b is also none.
The harmonic space at b is generated by:
 $-0.4082483 [0, 3] + 0.4082483 [1, 2] - 0.4082483 [3, 4] + 0.4082483 [0, 1] + 0.4082483 [2, 5]$
The harmonic generators birth at b is generated by:
 $-0.4082483 [0, 3] + 0.4082483 [1, 2] - 0.4082483 [3, 4] + 0.4082483 [0, 1] + 0.4082483 [2, 5]$

Figure 16: Shown results on the persistent harmonic space (II).

In the example, one see that we can indeed obtain the persistent harmonic space, the corresponding harmonic generators, and the generators of birth and death. If we only consider the dimensions of harmonic spaces, the code can be developed into a code for computing the barcode for persistent dimensions of harmonic spaces.

Through our interactions with ChatGPT, we have discussed various aspects of matrix transformations, operations, and calculations. We have gained knowledge about matrix coding, and ChatGPT has improved its understanding of the questions we have posed during our Q&A sessions.

As demonstrated above, we can guide ChatGPT step by step to implement the required code. What we want to emphasize is that ChatGPT can not only help us find existing codes but also create new codes that are unknown to researchers. This highlights the creative potential of ChatGPT for computations.

5 Applications

In this section, the ChatGPT codes obtained with our supervision are used to calculate data sets as examples. We will present two examples: one involving a point set of six points arranged as nodes of a hexagon, and the other involving three molecules.

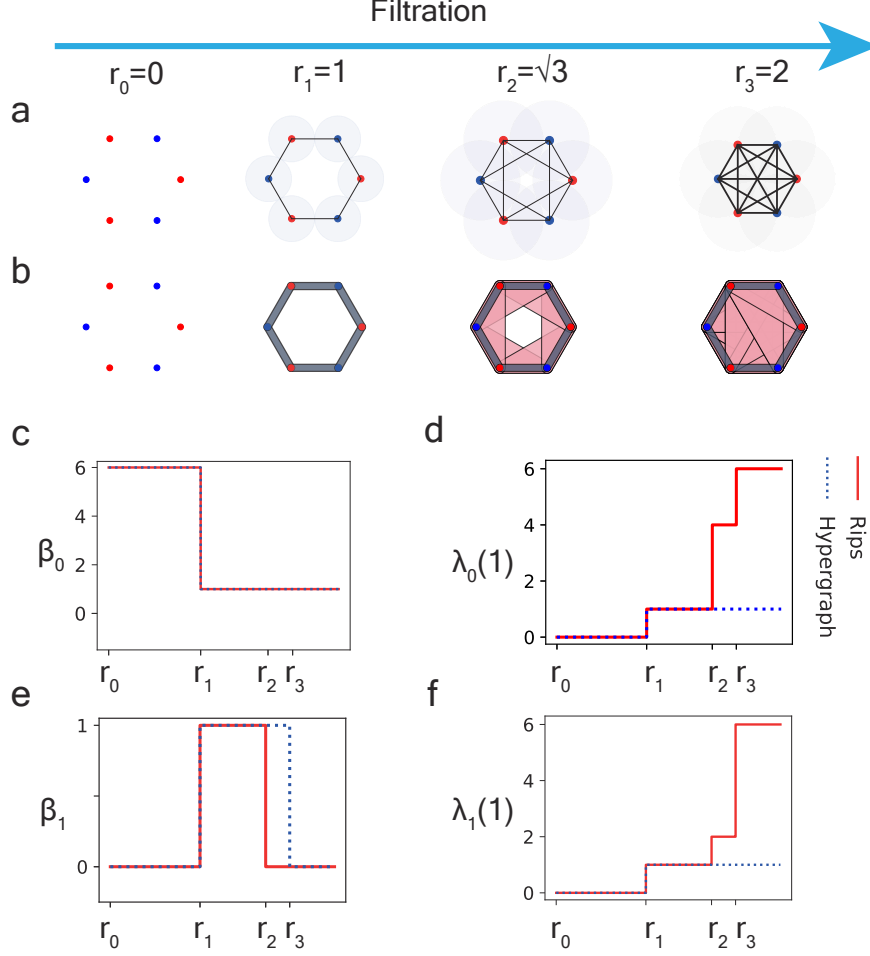


Figure 17: Comparison of persistent Laplacians generated by Rips complexes and hypergraphs. **a** The filtration of Rips complex. **b** The filtration of hypergraphs. The source code for generating virtualization of hypergraphs is from [CLWW23]. **c** Comparison of 0-dimensional Betti numbers for filtrations of Rips complexes and hypergraphs. **d** Comparison of spectral gaps of 0-dimensional Laplacians for filtrations of Rips complexes and hypergraphs. **e** Comparison of 1-dimensional Betti numbers for filtrations of Rips complexes and hypergraphs. **f** Comparison of spectral gaps of 1-dimensional Laplacians for filtrations of Rips complexes and hypergraphs.

In Fig. 17, we consider a point set composed of six vertices forming a regular hexagon with side lengths equal to 2. The six points are alternately colored in red and blue. For simplicial complex model, we use the Rips complex. For hypergraph model, the hyperedges are constructed from the corresponding Rips complex by requiring that each hyperedge has two colors. Specifically, the vertices are considered as 0-dimensional hyperedges. When the filtration parameter reaches $r_1 = 1$, the Rips complex forms a regular hexagon. The corresponding hypergraphs is given by $\mathcal{H} = \{\{0\}, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{0, 1\}, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{0, 5\}\}$. Here, $0, \dots, 5$ denotes the six points for convenience. When the filtration parameter $r_2 = \sqrt{3}$, the points 0 and 2 are connected by an edge. But $\{0, 2\}$ is not a hyperedge since 0 and 2 have the same color. Similarly, $\{1, 3\}, \{2, 4\}, \{3, 5\}, \{0, 4\}, \{1, 5\}$ are not hyperedges. However, sets as $\{0, 1, 2\}, \{1, 2, 3\}, \{2, 3, 4\}, \{3, 4, 5\}, \{0, 4, 5\}, \{0, 1, 5\}$ are 2-dimensional hyperedges. When $r_3 = 2$, all the six points are connected. The corresponding hypergraph have more 2-dimensional hyperedges, such as $\{0, 1, 3\}, \{0, 1, 4\}$ and so on. In 17c, the 0-th Betti numbers for the Rips complex model and the hypergraph model coincides. Fig. 17e shows difference of the 1-th Betti numbers between the two models. Fig. 17d and Fig. 17f show the spectral gaps of 0 and 1-dimensional Laplacians for the Rips complex model and the hypergraph model. When the

spectral gap equals 0, it means that all the eigenvalues are zero. While the Rips complex model and the hypergraph give different information for the given set of point cloud, the Rips complex model apparently offers rich information than a simple hypergraph does. However, hypergraph can be tailored to reveal specific structural features.

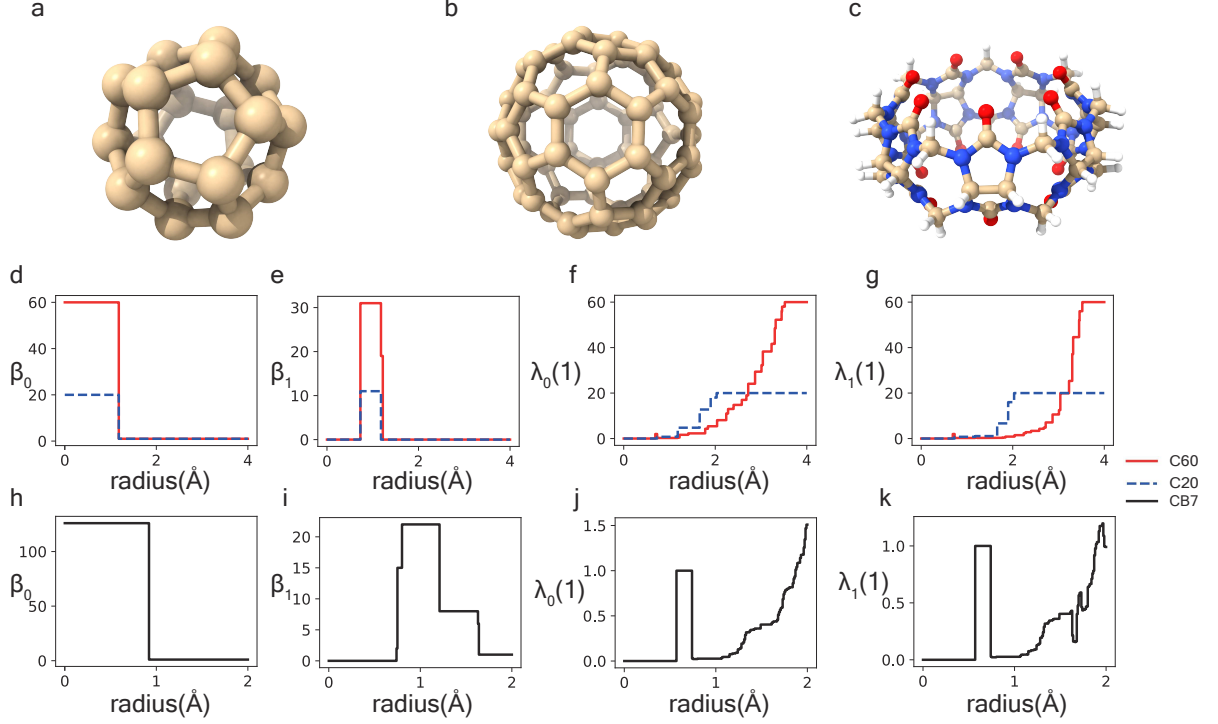


Figure 18: Illustration of persistent homology and persistent Laplacians for three molecules. **a** The structure of fullerene C_{20} . **b** The structure of fullerene C_{60} . **c** The structure of cucurbit[7]uril CB7. **d** Comparison of 0-dimensional Betti numbers for C_{20} and C_{60} . **e** Comparison of 1-dimensional Betti numbers for C_{20} and C_{60} . **f** Comparison of spectral gaps of 0-dimensional Laplacians for C_{20} and C_{60} . **g** Comparison of spectral gaps of 1-dimensional Laplacians for C_{20} and C_{60} . **h** The 0-dimensional Betti numbers for CB7. **i** The 1-dimensional Betti numbers for CB7. **j** The spectral gaps of 0-dimensional Laplacians for CB7. **k** The spectral gaps of 1-dimensional Laplacians for CB7.

Additionally, we consider examples of three molecules, including C_{20} , C_{60} , and CB7, and calculate the corresponding Betti numbers and Laplacian spectra over filtration. Note that C_{20} and C_{60} were studied in the literature using persistent Laplacians [WNW20], and CB7 was examined in the literature using persistent path Laplacians [WW23]. The topological features are computed from the Rips complexes of these small molecules. The atoms are regarded as points in Euclidean space and filtration lengths are determined by the distance between points.

In Fig. 18, subfigures **d**, **e**, **f**, and **g** show the comparison of Betti numbers and the smallest positive eigenvalues for the corresponding Laplacians of dimension 0,1 for C_{20} and C_{60} . Clearly, information obtained from persistent homology is limited. In contrast, persistent Laplacians not only return all the topological invariants as given by persistent homology in their harmonic spectra, but also reveal additional structural changes during the filtration in their nonharmonic spectra. Subfigures **h**, **i**, **j**, and **k** give the Betti numbers and the smallest positive eigenvalues for the corresponding Laplacians of dimension 0 and 1 of CB7. The curves of Betti numbers and the smallest positive eigenvalues based on persistent Laplacians are quite similar with those obtained with persistent path Laplacians [WW23].

6 Concluding remarks

In this study, we delve into the application of ChatGPT in computational topology from the perspective of a topologist who possesses limited knowledge of computational algorithms and lacks coding experience. To enable ChatGPT to generate accurate codes for computational topology tasks, we took several strategic steps. These steps involved training ChatGPT to understand fundamental mathematical concepts, guiding its responses in the right direction, leveraging ChatGPT to enhance our comprehension of the resulting algorithms, and designing test examples to validate the code provided by ChatGPT. This approach essentially involves a form of expert-supervised ChatGPT debugging.

One of the central mathematical models we focused on is the Vietoris-Rips complex, a widely used tool in applied topology that excels at characterizing the high-dimensional structure of data. Initially, we concentrated on algorithms for computing Betti numbers, Laplacians, and Dirac operators for simplicial complexes. Notably, the recently developed persistent Laplacians have greatly improved the effectiveness of applying topology in the field of molecular biology.

While simplicial complex models suit most applications, there are scenarios where they may not fully capture the data due to their constraints. This is where hypergraph and directed graph models come into play. Both hypergraphs and directed graphs exhibit high-dimensional structures, with hypergraphs characterized by embedded homology and directed graphs characterized by path homology. In our study, we harnessed the capabilities of ChatGPT to implement codes for computing Betti numbers and Laplacians specifically tailored for hypergraphs and directed graphs. We also demonstrate ChatGPT’s ability to compute the persistent harmonics space, which has been computed to the best of our knowledge.

By combining the expertise of a topologist with the capabilities of ChatGPT, we aim to advance the field of computational topology and make these powerful mathematical tools more accessible and applicable in various domains.

In our applications, we examined a point set comprising six vertices arranged in a regular hexagon. We carried out computations for the Vietoris-Rips complexes and determined the 0th and 1st-dimensional Betti numbers, as well as the smallest positive eigenvalue of Laplacians for hypergraphs. Subsequently, we generated plots illustrating the Betti curves and curves depicting the smallest positive eigenvalues. The ChatGPT-generated code yielded accurate results in these calculations. Additionally, we extended our analysis to include three molecules C₂₀, C₆₀, and C₇₀. For these molecules, we computed their 0-dimensional and 1-dimensional Betti numbers, along with the smallest positive eigenvalue of their Laplacians. We then visualized the corresponding Betti curves and smallest positive eigenvalue curves. The efficiency of the code was reasonable, with the calculations taking approximately several tens of minutes to run on a standard laptop.

In the future, there remain several unexplored avenues for further work and research. Firstly, one critical aspect we have not addressed in this work is algorithm optimization. On one hand, ChatGPT could directly aid us in optimizing algorithms if desired. On the other hand, we could leverage ChatGPT to explore mathematical optimizations of algorithms, ultimately achieving more efficient computations.

Secondly, while the algorithms presented in this paper can calculate persistent Betti numbers and persistent Laplacians, they do not encompass the computation of barcodes. We believe that ChatGPT can play a pivotal role in helping us complete this task.

Lastly, it is worth noting that commonly used topological invariants include (co)homology, Laplacian, and Dirac. However, there are other algorithms for topological invariants, such as cup-product, Steenrod algebra, K-theory, and others, which warrant further investigation and study.

This work signifies merely the initial stride in the utilization of chatbots for computational topology. We are confident that this step marks the commencement of a new era in mathematical research, forging connections between pure mathematical theories and computational tools in applied sciences. We envisage a promising future for this approach, as generative AI becomes increasingly potent and its capacity to comprehend mathematical concepts continuously enhances rapidly.

Data and Code Availability

The data and source code obtained in this work are publicly available in the Github repository: <https://github.com/JoybearLiu/ChatGPT-for-computational-topology>.

Acknowledgments

This work was supported in part by NIH grants R01GM126189, R01AI164266, and R35GM148196, National Science Foundation grants DMS2052983, DMS-1761320, and IIS-1900473, NASA grant 80NSSC21M0023, Michigan State University Research Foundation, and Bristol-Myers Squibb 65109.

References

- [AMS22] Bernardo Amenyro, Vasileios Maroulas, and George Siopsis. Quantum persistent homology. *arXiv preprint arXiv:2202.12965*, 2022.
- [BLRW19] Stephane Bressan, Jingyan Li, Shiquan Ren, and Jie Wu. The embedded homology of hypergraphs and applications. *Asian Journal of Mathematics*, 23(3):479–500, 2019.
- [Car09] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.
- [CHY22] Samir Chowdhury, Steve Huntsman, and Matvey Yutin. Path homologies of motifs and temporal network representations. *Applied Network Science*, 7(1):4, 2022.
- [CLW⁺23] Dong Chen, Jian Liu, Jie Wu, Guo-Wei Wei, Feng Pan, and Shing-Tung Yau. Path topology in molecular and materials sciences. *The Journal of Physical Chemistry Letters*, 14(4):954–964, 2023.
- [CLWW23] Dong Chen, Jian Liu, Jie Wu, and Guo-Wei Wei. Persistent hyperdigraph homology and persistent hyperdigraph Laplacians. *arXiv preprint arXiv:2304.00345*, 2023.
- [CM18] Samir Chowdhury and Facundo 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.
- [CQWW22] Jiahui Chen, Yuchi Qiu, Rui Wang, and Guo-Wei Wei. Persistent laplacian projected omicron ba. 4 and ba. 5 to become new dominating variants. *Computers in Biology and Medicine*, 151:106262, 2022.
- [CZCG04] Gunnar Carlsson, Afra Zomorodian, Anne Collins, and Leonidas Guibas. Persistence barcodes for shapes. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 124–135, 2004.
- [CZTW19] Jiahui Chen, Rundong Zhao, Yiyong Tong, and Guo-Wei Wei. Evolutionary de Rham-Hodge method. *arXiv preprint arXiv:1912.12388*, 2019.
- [EH⁺08] Herbert Edelsbrunner, John Harer, et al. Persistent homology-a survey. *Contemporary mathematics*, 453(26):257–282, 2008.
- [EH22] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.

- [ELZ02] Edelsbrunner, Letscher, and Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28:511–533, 2002.
- [GLMY12] Alexander Grigor’yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Homologies of path complexes and digraphs. *arXiv preprint arXiv:1207.2834*, 2012.
- [GLMY14] Alexander Grigor’yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Homotopy theory for digraphs. *arXiv preprint arXiv:1407.0234*, 2014.
- [GLMY15] Alexander Grigo’yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Cohomology of digraphs and (undirected) graphs. *Asian Journal of Mathematics*, 19(5):887–932, 2015.
- [KMM04] Tomasz Kaczynski, Konstantin Michael Mischaikow, and Marian Mrozek. *Computational homology*, volume 3. Springer, 2004.
- [LCLW22] Jian Liu, Dong Chen, Jingyan Li, and Jie Wu. Neighborhood hypergraph model for topological data analysis. *Computational and Mathematical Biophysics*, 10(1):262–280, 2022.
- [LLW23] Jian Liu, Jingyan Li, and Jie Wu. The algebraic stability for persistent Laplacians. *arXiv preprint arXiv:2302.03902*, 2023.
- [MWW22] Facundo Mémoli, Zhengchao Wan, and Yusu Wang. Persistent Laplacians: Properties, algorithms and implications. *SIAM Journal on Mathematics of Data Science*, 4(2):858–884, 2022.
- [MX21] Zhenyu Meng and Kelin Xia. Persistent spectral-based machine learning (perspect ML) for protein-ligand binding affinity prediction. *Science advances*, 7(19):eabc5329, 2021.
- [ST23] Gaurav Sharma and Abhishek Thakur. ChatGPT in drug discovery. 2023.
- [WBX23] JunJie Wee, Ginestra Bianconi, and Kelin Xia. Persistent Dirac for molecular representation. *arXiv preprint arXiv:2302.02386*, 2023.
- [WFW23] Rui Wang, Hongsong Feng, and Guo-Wei Wei. Chatbots in drug discovery: A case study on anti-cocaine addiction drug development with chatgpt. *arXiv preprint arXiv:2308.06920*, 2023.
- [WNW20] Rui Wang, Duc Duy Nguyen, and Guo-Wei Wei. Persistent spectral graph. *International journal for numerical methods in biomedical engineering*, 36(9):e3376, 2020.
- [WW23] Rui Wang and Guo-Wei Wei. Persistent path Laplacian. *Foundations of Data Science*, 5(1):26–55, 2023.
- [ZC04] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 347–356, 2004.