Published in final edited form as:

J Phys Chem Lett. 2023 February 02; 14(4): 954–964. doi:10.1021/acs.jpclett.2c03706.

Path Topology in Molecular and Materials Sciences

Dong Chen,
School of Advanced Materials, Peking University, Shenzhen Graduate School, Shenzhen 518055, China

Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, United States

Jian Liu,
School of Mathematical Sciences, Hebei Normal University, Heibei 050024, China

Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China

Jie Wu,
Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China

Guo-Wei Wei,
Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, United States

Department of Electrical and Computer Engineering and Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States

Feng Pan,
School of Advanced Materials, Peking University, Shenzhen Graduate School, Shenzhen 518055, China

Shing-Tung Yau
Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China

Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, China

Abstract

Corresponding Authors wujie@bimsa.cn.

Author Contributions

Chen designed the project, modified the method, wrote the code, performed computational studies, wrote the first draft, and revised the manuscript. Liu contributed methods, wrote the methods section, and revised the manuscript. Wu supervised the project and acquired funding. Wei conceptualized and supervised the project, acquired funding, and revised the manuscript. Pan supervised the project, acquired funding and revised the manuscript. Yau supervised the project and acquired funding.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.2c03706.

Persistent path homology analysis for molecular structures; paths and boundary operator; path complexes and their homology; topological perturbation analysis; persistent path homology (PDF)

Transparent Peer Review report available (PDF)

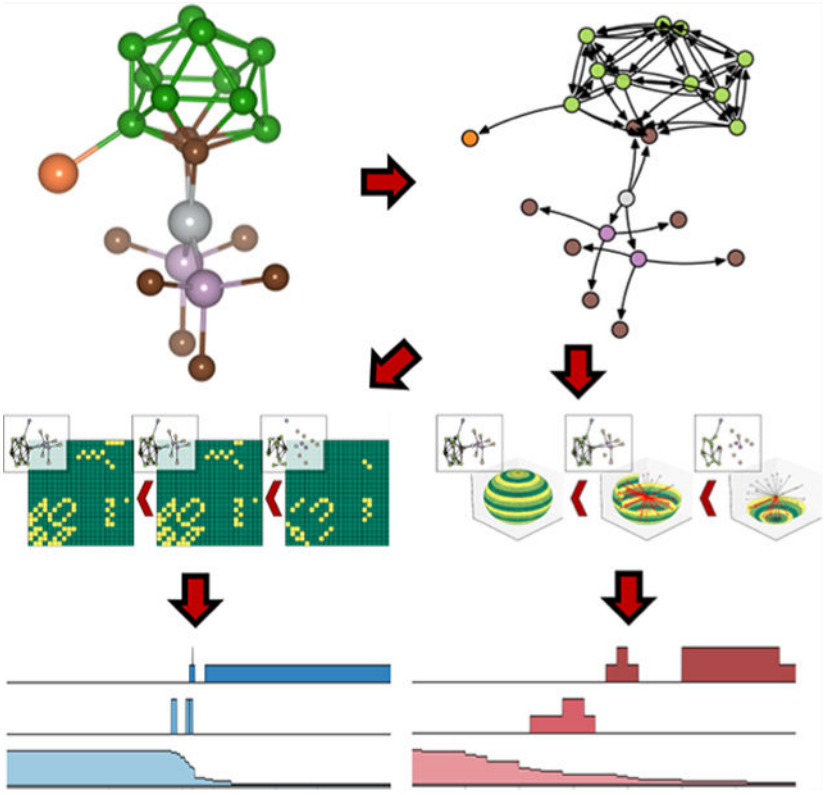
Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.2c03706

The authors declare no competing financial interest.

The persistent path topology analysis applies our established code, which is publicly available in the Github repository: https://github.com/WeilabMSU/PathHom

The structures of molecules and materials determine their functions. Understanding the structure and function relationship is the holy grail of molecular and materials sciences. However, the rational design of molecules and materials with desirable functions remains a grand challenge despite decades of efforts. A major obstacle is the lack of an intrinsic mathematical characteristic that attributes to a specific function. This work introduces persistent path topology (PPT) to effectively characterize directed networks extracted from functional units, such as constitutional isomers, cis−trans isomers, chiral molecules, Jahn−Teller isomerism, and high-entropy alloy catalysts. Path homology (PH) theory is utilized to decipher the role of mirror-symmetric sublattices that hinder the formation of periodic unit cells in amorphous solids. Topological perturbation analysis (TPA) is proposed to reveal the critical target in the blood coagulation system. The proposed topological tools can be directly applied to systems biology, omics sciences, topological materials, and machine learning study of molecular and materials sciences.

Graphical Abstract



The world is composed of molecules, including inorganic, organic, and biological ones. The understanding of molecular properties is part of human history. The structure-based rational design of functional molecules and materials is a grand challenge in scientific research, ranging from drug discovery and renewable energy to space shuttle coating. Although molecular energies can be predicted from molecular structures by quantum mechanism, there is still a big gap between molecular functions and their energies. Additionally, quantum calculations become intractable for large molecules. Recently, machine learning has emerged as a new paradigm for molecular and material design and discovery.[1] However,

how to embed the intrinsic structural characteristic of molecules and materials for machine learning remains a major obstacle.

Topology is well-known for its fundamental role in understanding topological insulators in terms of intrinsic symmetry-protected topological order.[2] Due to its ability to capture intrinsic topological characteristic, persistent topology or persistent homology,[3–5] has had tremendous success in simplifying the structural complexity and reducing the dimensionality of biological data.[6,7] Paired with artificial intelligence (AI), persistent homology accurately forecasted emerging dominant coronavirus variants about two months ahead.[8]

Inspired by the success of persistent homology, many advanced mathematical-based methods have been proposed for data analysis.[6] For example, persistent cohomology was introduced for protein stability analysis.[9] Algebraic graph theory is introduced to capture the geometrical information on molecules and protein−ligand complexes.[10,11] The evolutionary de Rham-Hodge theory and persistent spectral graph have been developed to capture both the topological invariants and homotopic shape evolution of data during filtration.[12–14] In addition, methods such as neighborhood complex,[15] Morse theory,[16] and hypergraph[17] have been applied to characterize complex molecular structures.

However, the aforementioned methods are insensitive to asymmetric or directional relations. In molecular and materials sciences, atoms have diverse properties, and their chemical bonds are typically heteronuclear. Metabolic pathways involve directed graphs (digraph) or networks.[18] Therefore, it is desirable to have mathematical tools to describe directed relationships in molecular and materials sciences, such as the directionality and polarity of bonds and reaction paths. Indeed, path homology (PH) proposed by Grigor'yan et al.[19–21] is designed for directed graphs and directed networks.[22] Instead of relying on homogeneous point clouds (or simplicial complexes in mathematical terminology), path homology is built on directed paths, which are essential to polarized chemical interactions and directed networks. Persistent path topology (PPT), also known as persistent path homology, was introduced to further empower path homology by multiscale filtration.[23] A one-dimensional persistent path homology was developed for applications.[24] Persistent path Laplacian (PPL) has also been proposed to reveal homotopic shape evolution of data.[25] However, these potentially powerful mathematical tools have hardly been applied to molecular and materials sciences.

This work fills this gap by introducing PH and PPT to molecular and materials sciences. To embed intrinsic geometric characteristics into topological invariants, we propose angle-based persistent path topology (APPT) to overcome the drawback of the existing distance-based persistent path topology (DPPT). APPT and DPPT complement each other in applications. We demonstrate the utility of these two forms of PPT methods for the identification and differentiation of three types of isomers in molecular systems, including D-fructose and D-glucose isomers, cis-trans isomers, and chiral molecules. We further apply PPT to characterize the Jahn−Teller effect and distinguish the isomers of high entropy alloy catalysts, showcasing PPT's advantage in unveiling atomic ordering induced physical and chemical properties. Using the PH method, we deciphered that the coexistence of mirror-symmetric tetrahedral or octahedral lattice pairs rules out the possibility for these

two lattices to become unit cells with long-range order in the amorphous solid. Finally, we propose topological perturbation analysis (TPA) to identify the critical node in the blood coagulation network, revealing a potential drug target of this process in systems biology.

Overview of Persistent Path Topology.

We use PPT to extract intrinsic topological characteristic of complex molecules and materials. Figure 1 shows the procedure of constructing PPT for the point cloud input data of a molecule (see Figure S1.). In this case, we start with atomic coordinates, where element types are labeled by colors as shown in Figure 1a. Unlike persistent homology, which regards all atoms in a molecule equally, PPT deals with paths, which can be extracted from the bond polarity determined by electronegativities of the bonded atoms. Four simplest paths, i.e., 0-path, 1-path, 2-path, and 3-path, are illustrated in Figure 1b. These are the basic building blocks for a topological space called p-path complex, which is a collection of all paths in the system and is characterized by the longest path length p in the system.

Although the path complex allows the construction of path homology, the process requires a standard mathematical operation called boundary operator, which exhausts all possible ways of removing one path in the system to end up with the $p-1$-path. The sequential applications of the boundary operator to the original p-path complex lead to a chain complex, which eventually vanishes after the use of the boundary operator $p-1$ times. The homology groups (H_p, H_{p-1}, ⋯, H_0) defined on the chain complex capture the topology of the original system. Specifically, the rank of the homology group H_n at each topological dimension ($n = p, p-1, ..., 0$) is a topological invariant, called Betti-n (β_n) number. Roughly speaking, Betti-0 (β_0) is the number of independent components in the system; Betti-1 (β_1) represents the number of looped objects composed of 1-paths in the system, and Betti-2 (β_2) is all voids under the meaning of paths composed of 2-paths in the system.

An individual homology group cannot discriminate all molecules as reflected in the joke that topologists cannot tell the difference between a coffee mug and a donut. This difficulty is resolved by filtration in PPT as shown in Figure 1c. Essentially, a family of paths is systemically and gradually established based on a distance thresholding. In the beginning, no path among nodes is allowed because the initial distance threshold, or the filtration parameter, is zero. Then, as the threshold is gradually increased, more and more paths are established when their distances between nodes are equal to or greater than the threshold as shown in the right charts of subgraphs in Figure 1c. The rectangular grid to the left of each subgraph records the connection of nodes in the system, where each pair of yellow entries indicates a path.

The topological invariants for H_2, H_1, and H_0 for the system over the filtration are depicted in Figure 1d. For H_0, the highest value of β_0 is 22, corresponding to 22 nodes. The β_0 decreases rapidly after the filtration parameter reaches 1.5. For H_1, there were not too many events, and β_1 had three short-living bars from 1.5 and 1.8, respectively. For H_2, there were one short-living void and one long-living void. We call persistent Betti numbers the DPPT fingerprints of the original colored point cloud data. Such fingerprints are readily used in the machine learning predictions.

The DPPT fingerprints may not be unique. To avoid DPPT isotopology from two molecules, we propose APPT. The essential idea, as shown in Figure 1e, is to translate all possible paths into a unit sphere such that the most highly aligned paths are located close to the south pole. Angle is increased to establish more paths during filtration. Path homology groups, H_2, H_1, and H_0, are constructed and their corresponding persistent Betti numbers are computed as shown in Figure 1f.

The topological invariants of DPPT and APPT methods extracted from a given system are distinguished from each other as shown in Figure 1d and Figure 1f. These two approaches complement each other and thus provide a more reliable framework for molecular characterization. Moreover, due to some empirical chemical-physical constraints, the final object used for PPT analysis can be either a fully connected digraph (Figure S1d) or a predefined digraph (Figure S1c) under specific conditions. DPPT and APPT are also presented in Figure S3 and Figure S4 when the fully connected digraph is used as the final path complex. The more mathematical details of the PH and PPT can also be found in the methods section and the Supporting Information Note 3.

Four-Vertex System.

To further illustrate the basic properties of PH, we consider a four-vertex system with different directed edge configurations. As shown in Figure 2a–d, for a 4-vertex digraph without heavy edges, there are only four cases. For this 4-vertex system, one observation is that there will be no high-dimensional holes if the traditional homology does not support such holes when all paths are replaced with edges. Indeed, $\beta_1 = 1$ in all systems Figure 2a–c. However, traditional homology admitting a high dimensional hole is not sufficient for PH to have a high dimensional hole. As shown in Figure 2d, there is no high dimensional hole in this case ($\beta_1 = 0$). There is no simple interpretation for the topological variants in PH, though the issue was discussed in the literature.[26]

Analysis for directed cubes and octahedrons.

Cubes and octahedrons are the basic building blocks of solid materials. To understand their properties, we carry out the systematical PH analysis of all possible scenarios in directed cubes and directed octahedrons as shown in Figure 2e and Figure 2f, respectively. Geometrically, in the nondirected case, the cube has the octahedral symmetry O_h of 48 symmetries, including 24 rotations. However, directed edges give rise to very different properties. For each of the 12 edges in the cube, there are two possible directions, which results in 2^{12} configurations. However, after 24 rotational symmetry operations, only 186 unique structures were obtained. Figure 2e shows the PH classification of the unique directed cubes in terms of Betti-0, 1, and 2 (i.e., β_0, β_1, and β_2). Illustrated in different colors, only six types of path homologies were obtained from 186 unique structures. The lines in the figure link the isomorphic structures. It is easy to notice that each directed cube has at most one isomorphic structure, and each pair of isomorphic structures always has the same set of topological invariants (i.e., β_0, β_1, and β_2). There are 74 pairs of isomers in total. We also found that all isomorphic structures are mirror-symmetric to each other, as shown in Figure 2g (left). PH offers dramatic simplification and interesting new characterization of

directed cubes. However, PPT can distinguish isomorphic structures as shown in Figure 2g (right).

Octahedron is dual to cube and also has O_h symmetry. For directed octahedrons, there are also 2^{12} configurations. After the 24 rotation symmetry operations, 186 unique directed octahedrons were found. Similarly, 74 pairs of these structures are isomorphic to each other, related by mirror symmetry. As shown in Figure 2f, for directed octahedrons, there are seven different combinations of topological invariants in dimensions 0, 1, and 2, which are different from the results for directed cubes. Notably, two of these structures have PH $\beta_0 = 1$, $\beta_1 = 0$, and $\beta_2 = 2$, and they are mirror-symmetric to each other. The persistent Betti numbers of these cases are displayed in Figure 2h. Therefore, PH analysis can distinguish directed cubes and directed octahedrons.

As a generalization of undirected structures, directed structures can bring richer structural information. Path homology, as defined over directed structures, is a powerful tool to analyze directed structures or directed networks. However, the PH alone cannot distinguish some stereoisomerism, such as the mirror-symmetric isomers shown in Figure 2g and Figure 2h. We introduced DPPT and APPT for the analysis of three types of isomers in stereochemistry.

It is natural to use a directed edge (path) to describe the unbalanced relationship in the real world. For a pair of atoms, differences between them often lead to asymmetric relationships, which in turn give rise to polarities at the molecular level. The differences between atoms can be deduced from atomic weights, atomic radii, electronegativities, and local charge densities, etc. In the present work, all directed edges between atoms are assigned from the atom with smaller electronegativity to the one with larger electronegativity. For atoms of the same element type, we use a pair of edges with opposite directions to connect two atoms, representing the relationship between them (alternatively, one can track the electronegativity of their neighbor atoms.).

D-Fructose and D-glucose.

The molecular formulas of D-fructose and D-glucose are identical, which is $C_6H_{12}O_6$, but their spatial configurations are different as shown in Figure 3a and Figure 3b. This difference in spatial structures causes a difference in optical activity. The naturally existing compounds are mostly D-glucose, while the artificially synthesized ones are achromatic, i.e., a mixture of D-fructose and D-glucose. Enzymes in an organism can recognize substances with different spin properties. D-fructose is not physiologically active, so the human body can only utilize D-glucose, rather than D-fructose. Their different geometric structures lead to different directed graphs, as shown in Figure 3a (right) and Figure 3b (right). To identify these two structures, we added orientation information to the structure and analyze the structures by using both DPPT and APPT.

Figure 3c shows the DPPT analysis of D-glucose (dark blue) and D-fructose (light blue). The horizontal axis denotes the distance-based filtration parameters, unit in Å, and the vertical axis denotes persistent Betti numbers. In this work, we only consider H_0, H_1, and H_2. The dark blue area indicates the DPPT for the D-glucose, the light blue area indicates the DPPT

for the D-fructose, and the area covered by the grid indicates their overlap. The H_0 records the β_0 values for each filtration parameter. The decrease of β_0 indicates that the newly formed paths connect some independent components, which has the same meaning as the Betti numbers of simplicial complexes in dimension 0. The H_1 and H_2 record the closed loops and the cavities along distance-based filtration in the sense of path complex. Although the H_0 of the two molecules overlaps exactly, we can still find differences between them in the H_1 and H_2, as shown in Figure 3c.

Figure 3d shows the APPT analysis for the aforementioned molecules. To apply APPT, we can find the difference between the two molecular structures in the first and second topological dimensions. This shows that the different spatial configurations lead to a large change in the directed edges, and APPT is so sensitive to the change in angle, leading to obvious angular changes. More than 20 directed loops were found in H_1, the difference between D-glucose and D-fructose can be captured as the angle-based filtration parameter changes, and the structure of D-glucose can form more directed loops. It is interesting to note that for H_2, the structure of D-glucose can form the two short living cavities in the sense of path complex but the structure of D-fructose has only one.

Cis−trans isomerism.

Cis−trans isomerism, also known as geometric isomerism, is a common form of stereoisomerism. Cis−trans isomerism refers to the diastereoisomerization of a compound molecule due to the different spatial arrangements of individual groups as a result of a freely rotating limiting factor. This restriction is generally caused by the presence of functional groups in the structure of organic compounds such as C=C double bond, C=N double bond, C=S double bond, N=N double bond, or an alicyclic ring that cannot rotate freely. The cis−trans isomers generally differ in physical, chemical, and biological properties.

To demonstrate the ability of PPT in distinguishing between cis−trans isomers, we use *trans*-1,2-dichlorocyclohexane (Trans) and *cis*-1,2-dichlorocyclohexane (Cis) shown in Figure 3e and Figure 3f. As shown in Figure 3g, the Cis and Trans structures share the same digraph, the direction of the edge depends on the electronegativity values.

Since the differences between cis and trans isomers usually arise only from the differences in the positions of individual functional groups, such as the Cl in this example, it is difficult to distinguish them by DPPT as shown in Figure 3h. DPPT persistent Betti numbers only exhibit very small differences at H_2. The path complex cavity of Cis can live longer. On the contrary, the characteristic curves based on the APPT method shown in Figure 3g display a very clear difference between the cis and trans isomers. For H_0, Trans shows that β_0 begins at 18. It means the independent atoms in the structure. As shown in Figure 3f, the functional groups (Cl, orange) of Cis have almost opposite orientations, while the functional groups in Trans have similar orientations. This difference can be obtained in H_1, where the structure of Trans can produce more path rings compared to the structure of Cis. The maximum β_1 is 17 for Trans, and for Cis, the maximum β_1 is only 4.

Mirror symmetric isomers.

A chiral molecule is a molecule that has a nonsuperposable mirror image.[27] All chiral molecules are optically active, and almost all optically active compounds are chiral. Chirality is important in the creation and evolution of life. For example, the pharmacological effects of drugs acting in living organisms are often matched in the chirality. In some cases, one chiral compound is therapeutically effective, while the other may have deleterious properties. Therefore, it is important to effectively distinguish chiral molecules in computational modeling and machine learning representation. In this study, the *R*-alanine and *S*-alanine shown in Figure 3j and Figure 3k are used to demonstrate PPT's capability of dealing chiral molecules. Due to the uncertainty of experimental measurements, their structures only have approximate mirror symmetry. Their digraph representation is depicted in Figure 3l.

For molecular structures with mutual chiral symmetry, the distances between the atoms inside the molecules are almost the same. For all given distance-based filtration parameters, the corresponding path complexes are all the same, which results in the same path homology. The DPPT cannot distinguish the chiral molecules. As shown in Figure 3m, DPPT persistent Betti numbers in dimensions 0, 1, and 2 are identical.

For APPT analysis, the filtration process is based on a sequence of angles in a prescribed order, e.g., $(S^2, \leq)$. The details of the ordered set $(S^2, \leq)$ can be obtained in the methods section. For chiral molecules, as shown in Figure 3n, APPT demonstrates a very clear difference between *R*-alanine and *S*-alanine, especially in the H_1. *R*-Alanine has maximal $\beta_1 = 3$, while *S*-alanine is unable to form a nontrivial path homology throughout the filtration process. It means that, for R-alanine, there are as many as three loops in the path sense forming during the angle-based filtration, while for its chiral molecule, *S*-alanine, no such loops are forming. For designed mirrorsymmetric structures, as shown in Figure 2g and Figure 2h, the difference between the isomorphic cube and the mirrorsymmetric isomorphic octahedron can be captured by APPT, while DPPT cannot distinguish these isomorphic structures.

In this section, we analyzed the three spatial isomers in molecular science using DPPT and APPT, respectively. APPT persistence Betti numbers can differentiate all the above-mentioned isomers, while DPPT fails to discriminate chiral molecules and mirror-symmetric isomeric structures.

Jahn−Teller effect.

The Jahn−Teller effect, also called the Jahn−Teller distortion, describes the geometric distortion of molecules and ions caused by spatially degenerate electronic ground configurations. The elimination of the degeneracy through structural distortion reduces the overall energy of the compound.

Figure 4a shows how the Jahn−Teller effect occurs in the $Li_2Mn_2O_4/LiMn_2O_4$ system.[28] $Li_2Mn_2O_4$ is used as a cathode material for the Li-battery. During the charging and discharging processes, the chemical valence state of Mn gradually becomes +3 with the injection of lithium ions, which leads to the distortion of the octahedron in the MnO_6

structure with Mn in the center and oxygen ion elongation along the *z*-axis, as shown in the middle of Figure 4a. Figure 4b shows the DPPT analysis of the structural distortion caused by the Jahn−Teller effect, including one octahedron. The *x*-axis represents the filtration parameters ranging from 0 to 5 Å. The structural differences before and after the distortion can be captured by the information in H_0 and H_1 starting from 0.2 Å. For the APPT analysis, as shown in Figure 4c, the persistent Betti numbers in H_0 are 5, which is smaller than the number of the atoms at the beginning. It is because some directed edges are covered by the spherical surface in the first angle-based filtration step, as demonstrated in the angle-based filtration Figure 1f. Since the distortion changes the position of oxygen atoms, the directed edges between oxygen atoms, and between oxygen and manganese atoms are all changed. This leads to a different H_0 for these structures at the beginning of the filtration step; the structure of $Li_2Mn_2O_4$ forms fewer edges (1-path) at the beginning. After the distortion, the sublattice has more directed loops (H_1) and more directed cavities (H_2) along with the angle-based filtration.

High-entropy alloy catalysts.

High entropy alloys (HEA) are alloys that typically contain multiple major elements. Due to their large number of different types of active surface sites, they have recently emerged as promising catalysts capable of fine-tuning the catalytic performance of many important energy conversion reactions. However, the huge number of active sites makes both theoretical modeling and experimental studies very difficult. In addition, even with the same material composition, the configurational variations can lead to changes in adsorption energy of up to 1 eV.[29,30]Figure 4 panels d and e show the slab model of RuRhIrPdPt HEA and OH* complexes. They are geometrically the same but chemically have very different arrangements of atoms. This leads to a difference of 5.2 eV in the formation energy of HEA for OH* adsorbates, (formation energy of −456.9 eV on the left and −451.7 eV on the right).[31]

Since the HEA contains different types of atoms, it allows us to define structures with directionality depending on the electronegativity of the atoms. As a demonstration, we just use the complex formed by the first two layers of HEA and OH* for the analysis. The results of the analysis by the DPPT method are shown in Figure 4f. For these two structures, their H_0 persistent Betti numbers are almost the same throughout filtration, which indicates similar geometric information on these two structures. However, their H_1 and H_2 persistent Betti numbers are sharply different, which means that the DPPT can capture the difference in the arrangement of the atoms within HEA. The dotted line points to the corresponding path complexes when the filtration parameter is equal to 2.7 and 3.4 Å. As for the APPT analysis (Figure 4g), the difference can be discovered from all H_0, H_1, and H_2. The dotted line points to the corresponding path complexes in the 81th and 207th angle-based filtration steps. The above results show that PPT can effectively capture the characteristics of HEAs, especially important features with chemical significance, such as atom arrangements, to achieve the identification and classification of different HEAs in machine learning predictions.

Primitives of amorphous solid.

Amorphous solids, such as glass and plastics, are ubiquitous and have a wide range of applications in our daily lives. However, due to the lack of long-range periodic ordering, it is difficult to directly determine the three-dimensional (3D) atomic structure of amorphous solids. The medium-range ordering has also been found in amorphous solids.[32] In this work, we analyzed the two most common geometries within amorphous solid, such as the tetrahedral lattice in Figure 4f and the octahedral lattice in Figure 4k. Specifically, we consider a high entropy alloy mixed in a tetrahedral vertex and an octahedral vertex.

For a tetrahedral lattice, up to four metallic elements can be present at the same time, as shown in Figure 4f. We use four colors to indicate the four types of atoms that may be present in the lattice. To naturally induce a directed lattice, the arrows between the different colors can be determined by the electronegativity (or other properties) of the elements. Figure 4i shows the nine unique directed tetrahedral lattices after the rotation operation. Among these structures, we found two structures that are mirror-symmetric to each other, which have been plotted in a frame in Figure 4i. The results of the APPT analysis are shown in Figure 4j. We can distinguish them directly by H_0 information.

Similarly, we use multiple atoms from at most six element types to induce the octahedral-based directed structure, as shown in Figure 4k. After removing duplicated atomic mixing schemes, we can obtain 2226 unique directed octahedrons with different atomic mixing schemes. The PH analysis of all possible unique directed octahedrons is shown in Figure 4l. Topologically, only 4 sets of PH systems characterized by four different combinations of β_0, β_1, and β_2 were found and shown in four colors. As many as 145 geometric digraphs exhibit the uniqueness under rotation operations, which is represented by circles on the horizontal axis. The size of each circle indicates the number of repeated atomic mixing schemes. The connecting lines in the figure indicate structures that are isomorphic to each other. It can be found that isomers always have the same path homology.

Also, as shown in Figure 2e,f, all isomers are mirrorsymmetric to each other. There are usually 5 or more elements in high entropy alloy nanoparticles, then it is highly likely that there are various configurations of tetrahedral lattice and octahedral lattice. Although in the above examples we did not consider the structural distortion caused by different elements. We can find that tetrahedral lattice and octahedral lattice are hard to be the unit cell lattice to form the long-range order in the HEA, because the mirror-symmetric configurations with tetrahedral lattice or octahedral lattice cannot be arranged periodically by translation and rotation.

Here, we applied PPT to three systems in materials science and found that the physicochemical information in material structures can be well encoded by digraphs (path complexes). PH and PPT can also be well extracted from the digraph for material structural analysis. In addition, we also analyzed a nanocage structure in Supporting Information using DPPT and APPT, as shown in Figure S5. It is found that PPT can generate rich structural information for materials sciences.

Topological perturbation analysis of directed networks.

Systems pharmacology is designed for the quantitative analysis of the dynamic interactions between drugs and biological systems. It bridges pharmacometrics and systems biology. In the past decade, many computational models have been developed for systems pharmacology.[33–35] However, understanding complex pharmacokinetic and pharmacodynamic issues in drug development remains a challenge. To provide a new perspective, we apply path homology analysis to reveal the topological significance of complex systems biology processes.

To demonstrate the potential capability of the PH in systems pharmacology and systems biology, we consider the human coagulation process (see Figure 5a). In this case, our goal is to discover the most important drug target in the reactive chemical system Figure 5a. To this end, we extract a directed network from the system as shown in Figure 5b. PH analysis shows $\beta_0 = 1$ and $\beta_1 = 7$ for Figure 5b. Here, $\beta_0 = 1$ is due to the connected physiological system, and $\beta_1 = 7$ indicates 7 closed loops in the sense of path complex. We introduce a topological perturbation analysis (TPA) to explore the importance of each node (protein) in this directed network. Specifically, we systematically remove one node in the network at a time (see, for example, Figure 5c) and then compute PH on the remaining directed network to obtain topological invariants. We track the changes in Betti numbers $\Delta\beta_n$ induced by the perturbation and denote the node causing the maximum change $\Delta\beta_n^{\max}$ as the critical node. The critical node could be a bottleneck in metabolic pathways or an ideal target for drug discovery in a protein network.

The changes in Betti numbers $\Delta\beta_n$ for the human coagulation process are plotted in Figure 5d. In this study, we reveal that removing node thrombin (IIa) causes the most drastic disruption in the topological invariance of the network. Namely, β_0 changed from the original 1 to 2 ($\Delta\beta_0 = 1$), which means the original connected system is split into two parts. Additionally, β_1 changed from the original 7 to 1 ($\Delta\beta_1 = -6$), as shown in Figure 5d, which means that after removing the paths associated with thrombin, the number of directed rings in the path sense is drastically reduced. In comparison, the next important node has $\Delta\beta_1 = -3$. Our TPA indicates that thrombin is the most important drug target for treating thrombus. Physiologically, thrombin is the most important component of coagulation and is rapidly released in terms of feedback activation to ensure timely clotting of blood, protecting the wound, and preventing further blood loss. The proposed TPA is an effective new approach for the study of directed networks in systems biology.

TPA is the first ever PPT technique proposed for biological network analysis. It can single out critical nodes in a complex network and has a great potential for applications to biological networks such as gene regulatory networks, protein−protein interaction networks, signaling networks, metabolic networks, neuronal networks, DNA−DNA−chromatin networks transcriptomic networks, etc. The present work adds a new dimension to computational biology. PPT opens a new door to the future development in biological networks which is promising for drug target identification, discovery of gene motifs, directed evolution, protein engineering, and omics in general.

The successful application of PPT to three classes of materials is inspirational. Although it remains to verify the specific correlation between PPT-based persistent Betti numbers and topological materials for specific examples, it is safe to expect that the proposed PPT analysis will play an important role in the design of topological materials because the same topological principle is applied. Therefore, the proposed PPT will aid the design and development of electronic/photonic topological insulators,[2] materials for Libattery and catalysts, one-dimensional, two-dimensional, and/or three-dimensional topological materials, exotic electrical and optical semimetals, superconductors, and other forms of matter. This potential application of PPT may impact a wide range of fields such as ultralow-energy transistors, cancer-removal lasers, and broadband communication beyond 5G.

Although we have not demonstrated the use of PPT for the machine learning prediction of molecular and material properties in this work, it is straightforward to do so based on our early success with persistent homology.[36,9,6,8] The proposed PPT will give rise to robust and efficient representations of molecules and materials for machine learning prediction. These representations can be easily integrated into a wide variety of machine learning, including deep learning, algorithms. They can also be concatenated with other descriptors. PPT-based AI will enable a new generation of advanced algorithms in molecular and materials predictions.

While persistent homology treats all atoms indiscriminately, PPT characterization of molecules allows the embedding of element types into the topological analysis. This additional function enables PPT to deal with chemistry/biochemistry without resorting to the element-specific techniques[36] and/or persistent cohomology.[9]

The proposed APPT complements the traditional distance-based topological methods. It can be generalized with a variety of vector alignment/ordering methods so that the advantages of the PPT method can be fully utilized when facing different complex problems in molecular and materials sciences. In addition, the present PPT can be easily generalized to a persistent hyperdigraph. This approach will allow the incorporation of subnetworks in path homology analysis.

Unlike persistent homology, whose high-dimensional cycles can be easily comprehended, PPT renders high-dimensional cycles that cannot easily be interpreted with traditional geometry. Further efforts are needed to make PPT interpretable. Compared with traditional persistent homology, PPT may rapidly increase its high-dimensional paths and cycles, leading to computational challenges.[25] Therefore, computational complexity can be a problem for the PPT analysis of large networks. The development of efficient algorithms for computing PPT, particularly high-dimensional path complexes, is an important task.

METHODS

The path homology theory, introduced by Grigor'yan et al.,[20,21] is a new mathematical tool. Persistent path topology theory was proposed in 2018.[23] In this work, we further introduced the PPT based on weight functions, angle-based filtration of PPT, and topological perturbation analysis.

Path complexes and their homology.

Let V be a nonzero finite set. For any integer $p \geq 0$, an *elementary p-path* on V is a sequence $i_0 i_1 \cdots i_p$ of elements in V. Let $\mathbb{K}$ be a field and let $\Lambda_p = \Lambda_p(V)$ be a $\mathbb{K}$-linear space generated by all the elementary p-paths. More precisely, we denote $e_{i_0 i_1 \cdots i_p}$ the generator corresponding to the elementary p-path $i_0 i_1 \cdots i_p$ and the family $\left\{e_{i_0 i_1 \cdots i_p}, i_0, i_1, \ldots, i_p \in V\right\}$ is a basis of Λ_p over $\mathbb{K}$. We always make the convention that $\Lambda_{-1} = 0$ is the null space. An element in Λ_p is called a *p-path*. Thus, each p-path v can be uniquely written as

$$v = \sum_{i_0, i_1, \cdots, i_p \in V} a^{i_0 i_1 \cdots i_p} e_{i_0 i_1 \cdots i_p}, a^{i_0 i_1 \cdots i_p} \in \mathbb{K} \tag{1}$$

For any integer $p \geq 0$, we have a $\mathbb{K}$-linear map $\partial : \Lambda_p \to \Lambda_{p-1}$ defined by $\partial e_{i_0} = 0$ f o r $e_{i_0} \in \Lambda_0$ and $\partial e_{i_0 i_1 \cdots i_p} = \sum_{k=0}^{p} (-1)^k e_{i_0 \cdots \widehat{i_k} \cdots i_p}$ for $p \geq 1$, where $\widehat{i_k}$ means omission of the index i_k. It can be directly verified that ∂ is a boundary operator on $(\Lambda_p)_p$, that is,$\partial^2 = 0$.

A *path complex* on a nonempty finite set V is a nonempty collection $\mathscr{P}$ of elementary paths on V such that $i_0 i_1 \cdots i_p \in \mathscr{P}$ implies $i_0 i_1 \cdots i_{p-1} \in \mathscr{P}$ and $i_1 i_2 \cdots i_p \in \mathscr{P}$. Let $(\mathscr{P}, V), (Q, W)$ be two path complexes. A *morphism of path complexes* is a map $\theta : \mathscr{P} \to Q$ induced by the map of finite sets $V \to W$. A simplicial complex is a path complex. Let G be a digraph. The set of paths on G is a path complex, denoted by $\mathscr{P}(G)$. The paths in $\mathscr{P}$ are called *allowed paths*. We denote

$$\begin{aligned} \mathscr{A}_p &= \mathscr{A}_p(\mathscr{P}) \\ &= \left\{ \sum_{i_0, i_1, \ldots, i_p \in V} a^{i_0 i_1 \cdots i_p} e_{i_0 i_1 \cdots i_p} \mid i_0 i_1 \cdots i_p \in \mathscr{P}, a^{i_0 i_1 \cdots i_p} \in \mathbb{K} \right\} \end{aligned} \tag{2}$$

We make the convention that $\mathscr{A}_{-1} = 0$ is the null space. The space of ∂-invariant p-paths is given by

$$\Omega_{-1} = 0, \quad \Omega_p = \Omega_p(\mathscr{P}) = \{x \in \mathscr{A}_p \mid \partial x \in \mathscr{A}_{p-1}\}, p \geq 0 \tag{3}$$

Then $(\Omega_p)_p$ is a subchain complex of $(\Lambda_p(V))_p$. This leads to the definition of *path homology*

$$H_p(\mathscr{P}; \mathbb{K}) := H_p(\Omega_p(\mathscr{P})) = \frac{\ker \partial|_{\Omega_p}}{im \partial|_{\Omega_{p+1}}}, \quad p \geq 0 \tag{4}$$

The path homology of a digraph G is the path homology of $\mathscr{P}(G)$. The *p-th Betti number* of digraph G is the rank of the homology $H_p(G; \mathbb{K})$, denoted as $\beta_p(G)$. In particular, if G is a

(weakly) connected digraph, that is, there is an undirected path between any pair of vertices, then $\beta_0(G) = 1$.

Persistent path topology.

Persistent path topology was introduced by Chowdhury and Mémoli,[23] following persistent homology.[5] In this work, we propose angle-based persistent path topology to complement distance-based persistent path topology.

Consider the order set $(S, \leq)$, which can be regarded as a category with elements in S as objects and all the binary orders as morphisms. A *filtration of path complexes* is a covariant functor $\mathscr{F}:(S, \leq) \to \mathbf{Path}$ from the category $(S, \leq)$ to the category of path complexes. More precisely, $\mathscr{F}_a$ is a path complex for each $a \in S$, and for $a \leq b \leq c$, we have $f_{b,c} \circ f_{a,b} = f_{a,c}$ where $f_{a,b}: \mathscr{F}_a \to \mathscr{F}_b$ is the morphism induced by $a \to b$. The morphism $f_{a,b}$ induces a morphism of path homology $\tilde{f}_{a,b}: H_p(\mathscr{F}_a; \mathbb{K}) \to H_p(\mathscr{F}_b; \mathbb{K})$. The (a, b)-*persistent path topology* of $\mathscr{F}$ is defined by

$$H_p^{a,b}(\mathscr{F}; \mathbb{K}) = \mathrm{im}(H_p(\mathscr{F}_a; \mathbb{K}) \to H_p(\mathscr{F}_b; \mathbb{K})), \quad p \geq 0 \tag{5}$$

Moreover, the (a, b)-persistent Betti number is defined to be the rank of $H_p^{a,b}(\mathscr{F}; \mathbb{K})$. The persistent Betti numbers can be generated from persistent diagrams and persistent barcodes.[37]

The path complexes considered in the application are often defined on digraphs. Let Digraph be the category of digraphs. A *filtration of digraphs* is a covariant functor $\mathscr{D}:(S, \leq) \to$ **Digragh** from the category $(S, \leq)$ to the category of digraphs. A filtration of digraphs can induce a filtration of path complexes, which leads to the persistent homology of digraphs.

Distance-based filtration.

Let $G = (V, E)$ be a digraph such that V is a set of data points in a metric space $(X, \|\cdot\|)$. There is a weight function $d: E \to \mathbb{R}$ on the edge set E given by

$$d(x_1, x_2) = \| x_1 - x_2 \|, \quad (x_1, x_2) \in E \subseteq X \times X \tag{6}$$

Note that the function d here can be any weight function, such as the Lorentz function, the exponential function, etc. In this work, for demonstration purposes, the metric space $(X, \|\cdot\|)$ is assumed to be the Euclidean space with L_2-norm. Let $E_t = \{(x, y) \in E \mid d(x, y) \leq t\}$ and $\mathscr{G}_t = (V, E_t)$. It can be verified that $\mathscr{G}:(\mathbb{R}, \leq) \to$ **Digragh**, $t \mapsto \mathscr{G}_t$ is a filtration of digraphs, which gives a persistent diagram $\mathscr{D}(\mathscr{G})$ of G.

Consider the isometric transformation T on the space X, that is, a composition of rotation and translation transformations on X. Then $T\mathscr{G}$ is also a filtration of digraphs. Moreover, we have $\mathscr{D}(\mathscr{G}) = \mathscr{D}(T\mathscr{G})$. It means that the persistent diagram is an invariant under isometric

transformation on Euclidean space. Figure 1c and Figure 1d show the distance-based filtration on $C_8H_{27}B_{10}BrNiP_2$[38] (Figure S1a). The hydrogen atoms are not included for simplicity.

Angle-based filtration.

Consider a digraph $G = (V, E)$ with V in Euclidean space $\mathbb{R}^3$. We give a filtration of digraphs described as follows. First, we need to fix a coordinate system by a predefined rule. The predefined rules here can be nonunique, but the resulting coordinate system should be unique for a given rule. Second, each point on S^2 is given by a pair $(\alpha, \gamma), \alpha \in [0, 2\pi], \gamma \in [0, \pi]$. Let $S^2 = \{(\alpha, \gamma) \mid \alpha \in [0, 2\pi], \gamma \in [0, \pi]\}$ be an order set with its order given by

$$(\alpha, \gamma) \leq (\alpha', \gamma') \text{ if } (\alpha < \alpha') \text{ or } (\alpha = \alpha', \gamma \leq \gamma') \tag{7}$$

Namely, the priority of an element in the set is based on its parameter α. Parameter γ is considered only if α values for two elements are the same. For the discrete case, let m, k be positive integers. We can choose an order set S^2 by

$$\left(\frac{2\pi t}{k}, \frac{\pi s}{m}\right), \quad t = 0, \ldots, k-1, \quad s = 0, \ldots, m-1 \tag{8}$$

In this way, we give a filtration based on the angle shown as spiral progress in a polar coordinate system. In this work, we use $m = 6$, $k = 12$, and the detailed descriptions of angle-based filtration can be found in the Supporting Information. Figure 1e and Figure 1f illustrate the angle-based filtration on $C_8H_{27}B_{10}BrNiP_2$.[38]

The progress of angle changing gives an ordered set $\left(S^2, \leq\right)$ on a unit 2-sphere. By a straightforward verification, we have a filtration of digraphs based on angles $\mathscr{A}: \left(S^2, \leq\right) \to$ **Digraph**, $\theta \mapsto \mathscr{A}_\theta$ Here, $\mathscr{A}_\theta = (V_\theta, E_\theta)$ is given by $V_\theta = \left\{x \in V \middle| \frac{x}{\| x \|} \leq \theta \in S^2\right\}$ and $E_\theta = E \cap (V_\theta \times V_\theta) = \{(x, y) \in E \mid x, y \in V_\theta\}$. For distance-based filtration, it can be shown that

$$\mathscr{D}(\mathscr{G}) = \mathscr{D}(g\mathscr{G}), \quad g \in O(3) \tag{9}$$

where $O(3)$ is the orthogonal group $O(3)$. However, for $g \in O(3)$, it does not have to be $\mathscr{D}(\mathscr{A}) = \mathscr{D}(g\mathscr{A})$, since a symmetric transformation may change the persistent diagram. Fortunately, we always have

$$\mathscr{D}(\mathscr{A}) = \mathscr{D}(g\mathscr{A}), \quad g \in SO(3) \tag{10}$$

where $SO(3)$ is the rotation group. This means that the angle-based filtration can help us distinguish structures with mirror symmetry. It is worth noting that we can obtain a multidimensional filtration by combining the distance-based filtration and the angle-based filtration. We also provide a detailed description of the construction multidimensional filtration in the Supporting Information.

Topological perturbation analysis.

The topological perturbation analysis (TPA) is designed for general network analysis. It can be implemented with persistent homology and/or topological Laplacians.[25] In the present work, we implement TPA with path homology for the study of digraph properties by computing the path homology of the sub digraphs of the original digraph. Mathematically, TPA is related to relative homology.[39] Let $G = (V, E)$ be a digraph. We consider the sub digraph of G by deleting a vertex $v \in V$ and the edges connecting the vertex, which is called a *perturbation of G at v*, denoted by G_v. Let $i_v: G_v \to G$ be an inclusion of digraphs. It induces a morphism of path homology groups

$$i_v^*: H_p(G_v) \to H_p(G), \quad p \geq 0$$

In general, the morphism i_v^* can be neither injective nor surjective.

In application, we are interested in the changes of topological invariants,

$$\Delta\beta_p(v) = rank(H_p(G_v)) - rank(H_p(G)), \forall v \in V, p \geq 0$$

The importance of each node can be judged from $\Delta\beta_p(v)$. In particular, the maximal change, $\Delta\beta_p^{\max} = \max_v|\Delta\beta_p(v)|$, is important for the graph. The proposed TPA is valuable for practical applications and can be used to measure the impact of different nodes on the network.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The work of Chen and Pan was financially supported by the Shenzhen Science and Technology Research Grant (No. JCYJ20200109140416788), the Soft Science Research Project of Guangdong Province (No. 2017B030301013) and the Major Science and Technology Infrastructure Project of Material Genome Big-science Facilities Platform supported by Municipal Development and Reform Commission of Shenzhen. This work is also supported in part by Natural Science Foundation of China (NSFC grant no. 11971144), High-level Scientific Research Foundation of Hebei Province and the start-up research fund from BIMSA. The work of Jian Liu was supported in partial by Tianjin Natural Science Foundation (Grant No. 19JCYBJC30200). The work of Chen and Wei was supported in part by NIH grants R01GM126189 and R01AI164266, NSF grants DMS-2052983, DMS-1761320, and IIS-1900473, NASA grant 80NSSC21M0023, Michigan Economic Development Corporation, MSU Foundation, Bristol-Myers Squibb 65109, and Pfizer. Chen thanks BIMSA for the hospitality during his visit in summer 2022.

REFERENCES

(1). Hey A, Tansley S, Tolle K; Gray JThe fourth paradigm: data-intensive scientific discovery; Microsoft research: Redmond, WA, 2009; Vol. 1.

(2). Moore JE The birth of topological insulators. Nature 2010, 464 (7286), 194–198. [PubMed: 20220837]

(3). Edelsbrunner H; Letscher D; Zomorodian ATopological persistence and simplification. In Proceedings 41st annual symposium on foundations of computer science, IEEE, 2000; pp 454–463.

(4). Zomorodian A; Carlsson G. Computing persistent homology. Discrete & Computational Geometry 2005, 33 (2), 249–274.

(5). Carlsson G. Topology and data. Bulletin of the American Mathematical Society 2009, 46 (2), 255–308.

(6). Nguyen DD; Cang Z; Wei G-W A review of mathematical representations of biomolecular data. Phys. Chem. Chem. Phys 2020, 22 (8), 4343–4367. [PubMed: 32067019]

(7). Townsend J; Micucci CP; Hymel JH; Maroulas V; Vogiatzis KD Representation of molecular structures with persistent homology for machine learning applications in chemistry. Nat. Commun 2020, 11 (1), 1–9. [PubMed: 31911652]

(8). Chen J; Wei G-W Omicron ba. 2 (b. 1.1. 529.2): High potential for becoming the next dominant variant. journal of physical chemistry letters 2022, 13 (17), 3840–3849. [PubMed: 35467344]

(9). Cang Z; Wei G-W Persistent cohomology for data with multicomponent heterogeneous information. SIAM journal on mathematics of data science 2020, 2 (2), 396–418. [PubMed: 34222831]

(10). Nguyen DD; Wei G-W AGL-score: algebraic graph learning score for protein−ligand binding scoring, ranking, docking, and screening. J. Chem. Inf. Model 2019, 59 (7), 3291–3304. [PubMed: 31257871]

(11). Chen Dong; Gao Kaifu; Nguyen Duc Duy; Chen Xin; Jiang Yi; Wei Guo-Wei; Pan, Feng Algebraic graph-assisted bidirectional transformers for molecular property prediction. Nat. Commun 2021, 12 (1), 1–9. [PubMed: 33397941]

(12). Chen J; Zhao R; Tong Y; Wei G-W Evolutionary de rhamhodge method. Discrete and continuous dynamical systems. Series B 2021, 26 (7), 3785. [PubMed: 34675756]

(13). Wang R; Nguyen DD; Wei G-W Persistent spectral graph. International journal for numerical methods in biomedical engineering 2020, 36 (9), No. e3376.

(14). Meng Z; Xia K. Persistent spectral−based machine learning (perspect ml) for protein-ligand binding affinity prediction. Science Advances 2021, 7 (19), No. eabc5329.

(15). Liu X; Xia KNeighborhood complex based machine learning (NCML) models for drug design. In Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data; Springer, 2021; pp 87–97.

(16). Wu C; Ren S; Wu J; Xia K. Discrete Morse theory for weighted simplicial complexes. Topology and its Applications 2020, 270, 107038.

(17). Liu X; Wang X; Wu J; Xia K. Hypergraph-based persistent cohomology (HPC) for molecular representations in drug design. Briefings in Bioinformatics 2021, 22 (5), bbaa411.

(18). Schuster S; Fell DA; Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature biotechnology 2000, 18 (3), 326–332.

(19). Grigoryan A; Lin Y; Muranov Y; Yau S-T Homologies of path complexes and digraphs. arXiv 2012, No. 1207.2834, DOI: 10.48550/arXiv.1207.2834.

(20). Grigor'yan A; Jimenez R; Muranov Y; Yau S-T On the path homology theory of digraphs and Eilenberg−Steenrod axioms. Homology, Homotopy and Applications 2018, 20 (2), 179–205.

(21). Grigor'yan AA; Lin Y; Muranov YV; Yau S-T Path complexes and their homologies. Journal of Mathematical Sciences 2020, 248 (5), 564–599.

(22). Chowdhury S; Huntsman S; Yutin M. Path homologies of motifs and temporal network representations. Applied Network Science 2022, 7 (1), 1–23. [PubMed: 35013714]

(23). Chowdhury S; Mémoli FPersistent path homology of directed networks. In Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms; SIAM, 2018; pp 1152–1169.

(24). Dey TK; Li T; Wang Y. An efficient algorithm for 1-dimensional (persistent) path homology. Discrete & Computational Geometry 2022, 68 (4), 1102–1132.

(25). Wang R; Wei G-W Persistent path laplacian. Foundation of data Science 2023, 5 (1), 26–55.

(26). Chowdhury S; Gebhart T, Huntsman S; Yutin MPath homologies of deep feedforward networks. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA); IEEE, 2019; pp 1077–1082.

(27). Kelvin WTThe molecular tactics of a crystal; Clarendon Press, 1894.

(28). Huang W; Zhang M; Liu T; Zhao W; He L; Yin L; Tan Z; Lin C; Liu J; Zhao Q; Chen C; Qi R; Zuo C; Chen H; Lin H; Liu X; Amine K; Pan F; et al. Tunning the linkage of structure units to enable stable spinel-based cathode in the wide potential window. Nano Energy 2021, 89, 106457.

(29). Calle-Vallejo F; Loffreda D; Koper MTM; Sautet P. Introducing structural sensitivity into adsorption−energy scaling relations by means of coordination numbers. Nature Chem. 2015, 7 (5), 403–410. [PubMed: 25901818]

(30). Batchelor TAA; Pedersen JK; Winther SH; Castelli IE; Jacobsen KW; Rossmeisl J. High-entropy alloys as a discovery platform for electrocatalysis. Joule 2019, 3 (3), 834–845.

(31). Lu Z; Chen ZW; Singh CV Neural network-assisted development of high-entropy alloy catalysts: Decoupling ligand and coordination effects. Matter 2020, 3 (4), 1318–1333.

(32). Yang Y; Zhou J; Zhu F; Yuan Y; Chang DJ; Kim DS; Pham M; Rana A; Tian X; Yao Y; Osher SJ; Schmid AK; Hu L; Ercius P; Miao J; et al. Determining the three-dimensional atomic structure of an amorphous solid. Nature 2021, 592 (7852), 60–64. [PubMed: 33790443]

(33). Cheng L; Khoo MCK Modeling the autonomic and metabolic effects of obstructive sleep apnea: a simulation study. Frontiers in Physiology 2012, 2, 111. [PubMed: 22291654]

(34). Cheng L; Wei G-W; Leil T. Review of quantitative systems pharmacological modeling in thrombosis. Communications in information and systems 2019, 19 (3), 219. [PubMed: 34045928]

(35). Leil TA; Ermakov S. The emerging discipline of quantitative systems pharmacology. Frontiers in Pharmacology 2015, 6, 129. [PubMed: 26175687]

(36). Cang Z; Wei G-W Analysis and prediction of protein folding energy changes upon mutation by element specific persistent homology. Bioinformatics 2017, 33 (22), 3549–3557. [PubMed: 29036440]

(37). Ghrist R. Barcodes: the persistent topology of data. Bulletin of the American Mathematical Society 2008, 45 (1), 61–75.

(38). Qiu Z; Deng L; Chan H-S; Xie Z. Synthesis and structural characterization of group 10 metal-carboryne complexes. Organometallics 2010, 29 (20), 4541–4547.

(39). Hatcher AAlgebraic topology; Cambridge Univ. Press: Cambridge, 2002.

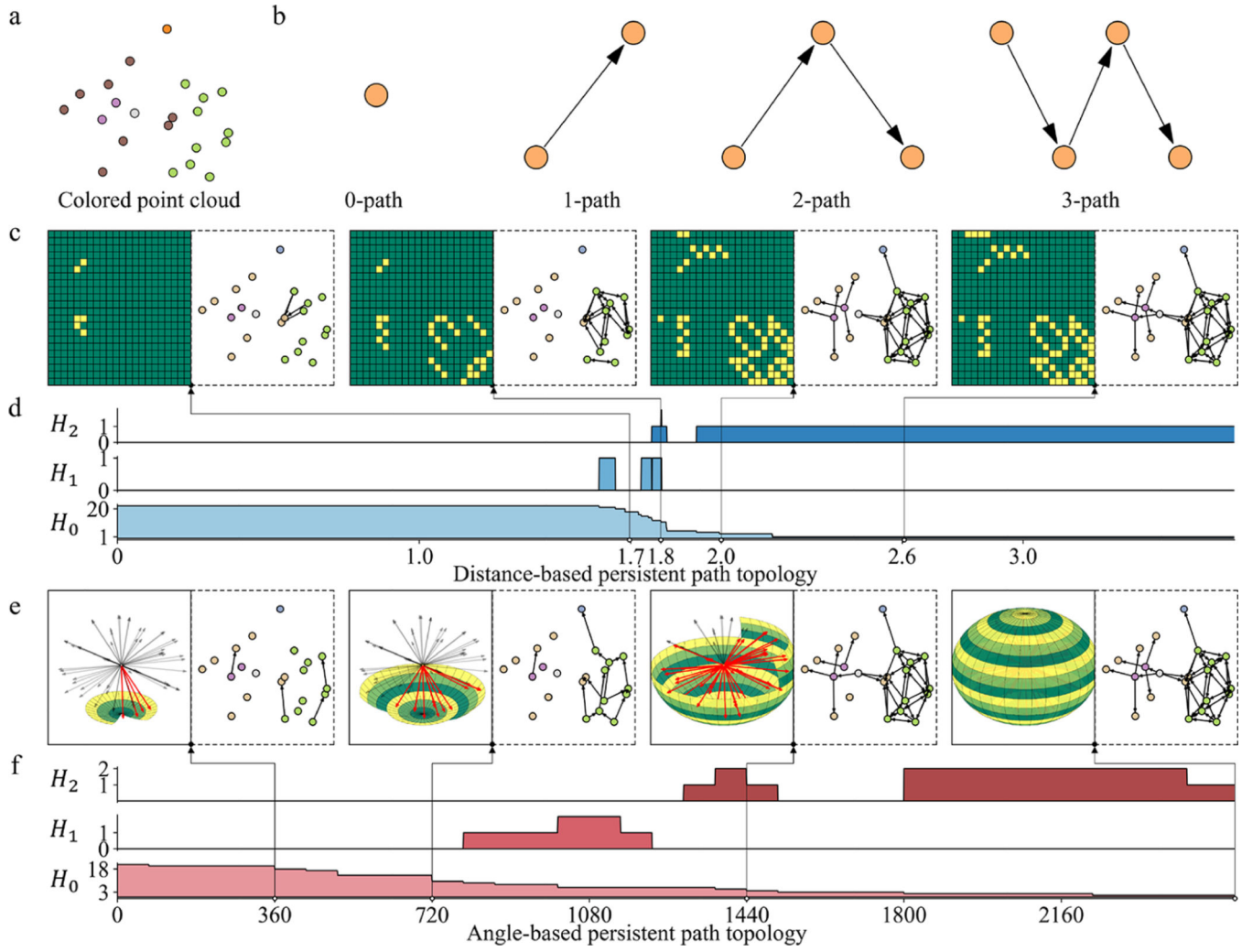


Figure 1.

Illustration of persistent path topology. (a) The weight function-based matrix is constructed from a molecular structure. (b) Illustration of the basic component that makes up the path complex, *p*-path, where $p = 0$, 1, 2, and 3. (c) Illustration of the distance-based filtration. As the filtration parameter increases, the path complex based on the weight matrix expands accordingly. In the left figure, the *x*-axis and the *y*-axis represent the atomic index in the structure, respectively. The yellow entries represent the formation of directed edges between the corresponding pairs of atoms. The right figure represents the corresponding path complexes. (d) The persistent Betti numbers of the distance-based persistent path topology. H_0, H_1, and H_2 show the 0-, 1-, and 2-dimensional path homology, respectively. The vertical axis represents the values of persistent Betti numbers, and the horizontal axis represents the filtration parameter in Å. (e) Illustration of the angle-based filtration. All possible directed edges are mapped to unit sphere. The path complex in the right figure expands with the increase of the directed edges covered by the growth of the spherical surface, where the increase of the spherical surface area is given by the angle-based rule, which is defined by eq 8 with $k = 36$, $m = 72$. (f) The persistent Betti numbers of the

angle-based persistent path topology. The vertical axis represents the values of persistent Betti numbers and the horizontal axis represents the angle-based filtration parameter.

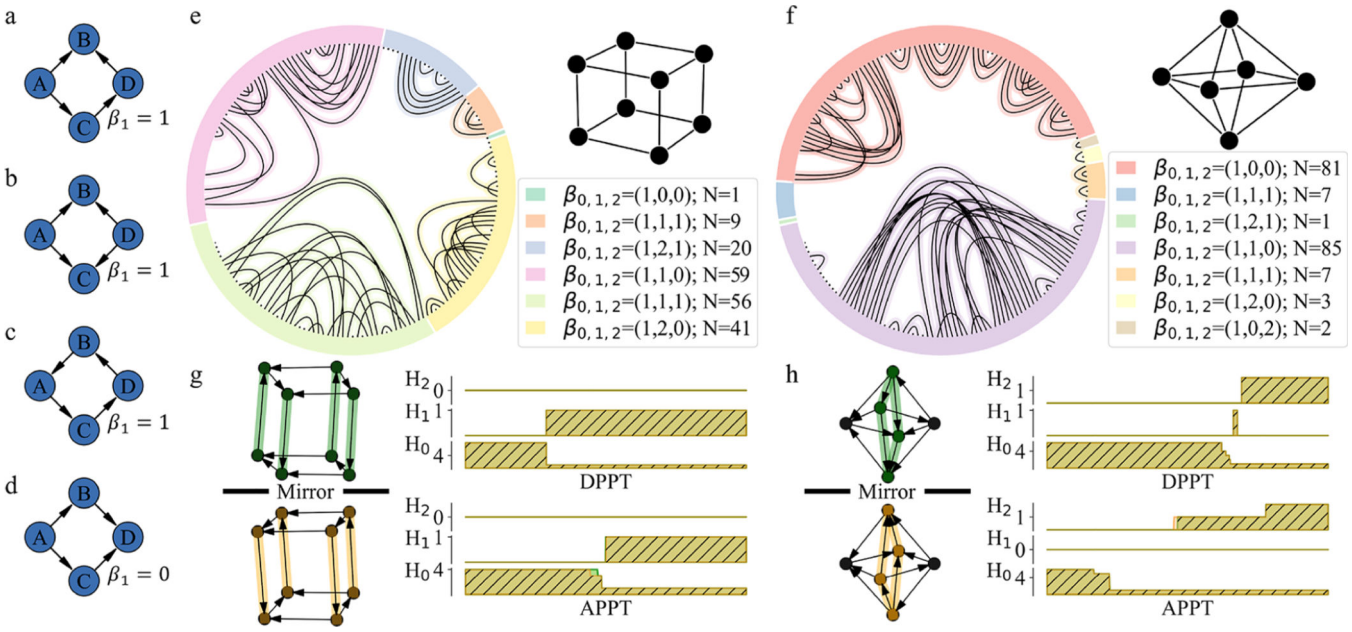


Figure 2.

Path homology and persistent path topology analysis of simple structures. (a−c) Three 4-vertex digraphs that have a nontrivial path homology ($\beta_1 = 1$) in the first dimension. (d) A 4-vertex digraph with $\beta_1 = 1$. (e) Path homology analysis of directed cubes. Colors represent different combinations of topological invariants from dimensions 0, 1 and 2, denoted as $\beta_{0,1,2}$. Where there are six types of combinations for directed cubes. The lines inside the ring connect structures that are isomorphic to each other. (f) Path homology analysis of directed octahedrons. Similar to (e), colors denote seven types of directed octahedrons coordinating to their $\beta_{0,1,2} = (\beta_0, \beta_1, \beta_2)$ characterization. (g) Persistent path topology analysis of two isomorphic directed cubes. Two isomorphic structures are mirror-symmetric to each other. Their DPPT and APPT persistent Betti numbers are shown on the right, where APPT can distinguish them. (h) Persistent path topology analysis of two isomorphic directed octahedrons, which are mirror-symmetric to each other. They have identical DPPT but different APPT.

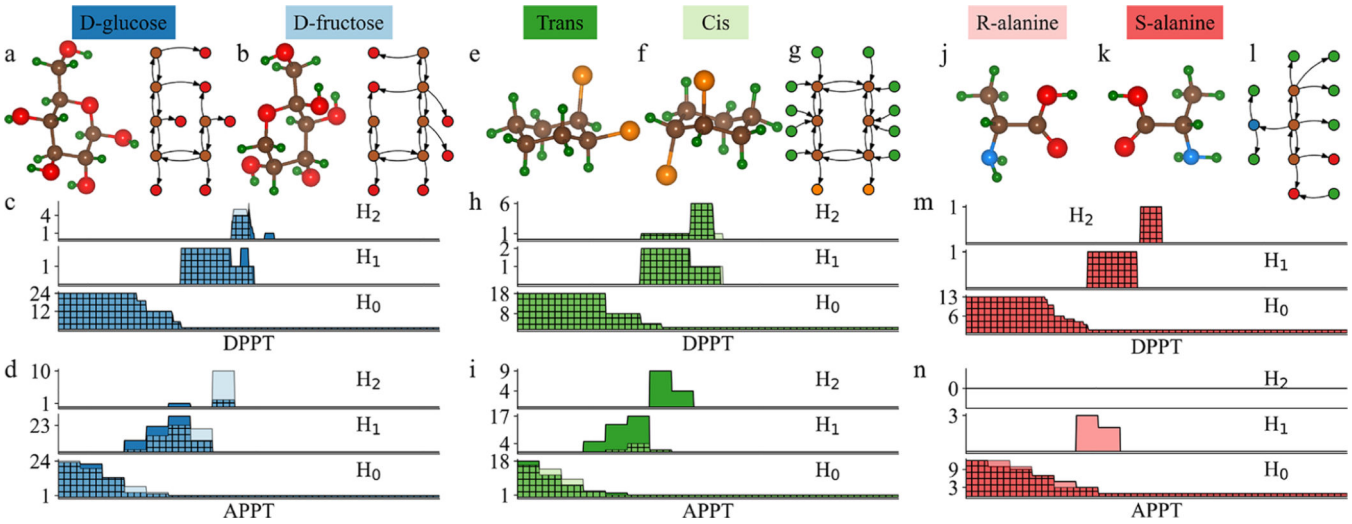


Figure 3.

Illustration of the DPPT and APPT analysis of spatial isomers. (a) The molecular structures of D-glucose (left) and associated digraphs (right). (b) The molecular structures of D-fructose (left) and associated digraphs (right). (c) The DPPT analysis of D-glucose and D-fructose. Shared parts are plotted in mesh. (d) The APPT analysis of D-glucose and D-fructose. Shared parts are displayed in mesh. (e) The structure of *trans*-1,2-dichlorocyclohexane. (f) The structure of *cis*-1,2-dichlorocyclohexane. (g) Shared digraph representation for trans and cis structures. (h) The DPPT analysis of *trans*-1,2-dichlorocyclohexane and *cis*-1,2-dichlorocyclohexane. Shared parts are plotted in mesh. (i) The APPT analysis of *trans*-1,2-dichlorocyclohexane and *cis*-1,2-dichlorocyclohexane. Shared parts are depicted in mesh. (j) The structure of *R*-alanine. (k) The structure of *S*-alanine. (l) Shared digraph representation of *R*-alanine and *S*-alanine. (m) The DPPT analysis of *R*-alanine and *S*-alanine. Shared parts are plotted in mesh. (n) The APPT analysis of *R*-alanine and *S*-alanine. Shared parts are presented in mesh.

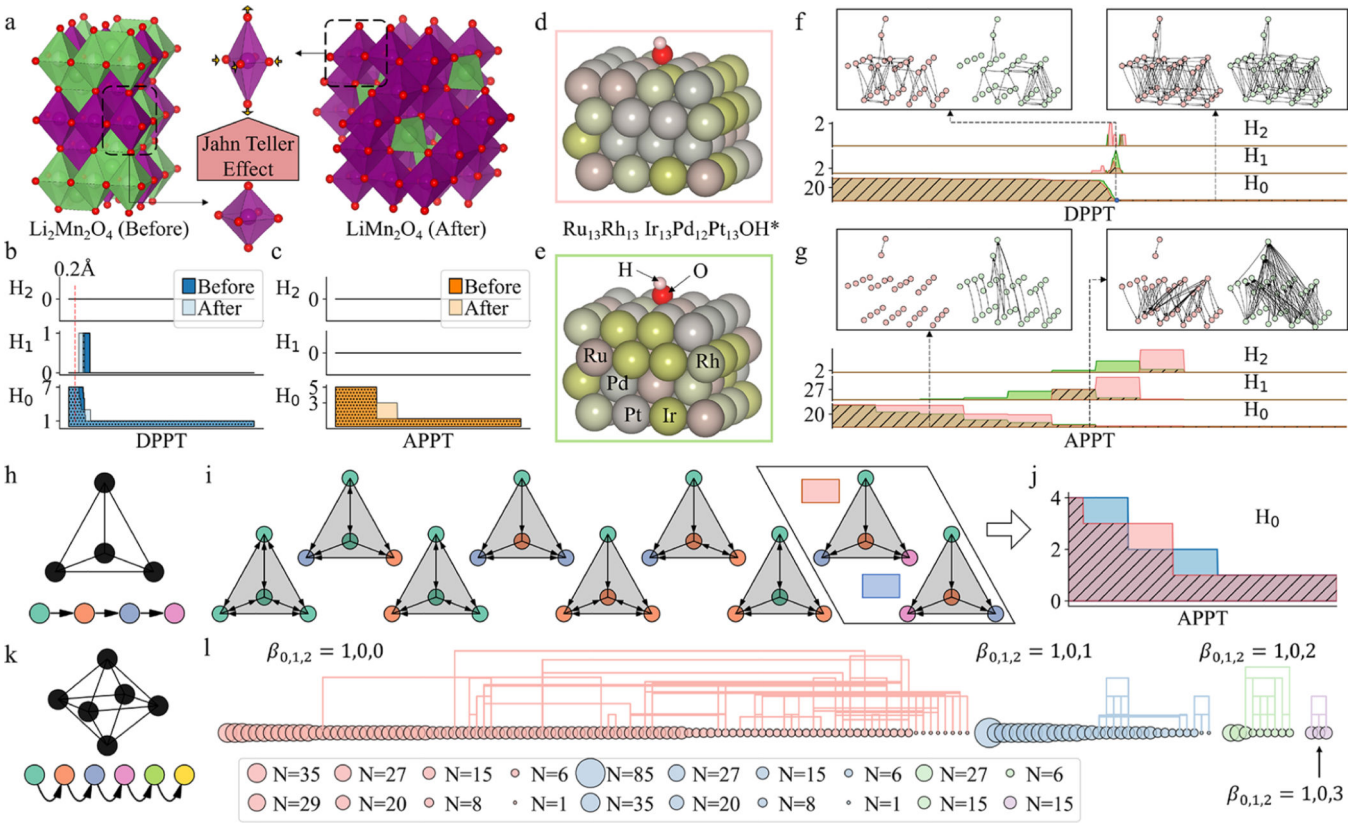


Figure 4.

PH and PPT analysis of Jahn−Teller effect, HEA catalysts, and amorphous solid. (a) Illustration of the Jahn−Teller effect associated with $Li_2Mn_2O_4$ and $Li_2Mn_2O_4$. (b) DPPT analysis of the Jahn−Teller effect in MnO_6 associated with $Li_2Mn_2O_4$ and $Li_2Mn_2O_4$. Shared persistent Betti numbers two MnO_6 structures are covered with mesh. (c) APPT analysis of the Jahn−Teller effect in MnO_6. Shared persistent Betti numbers two MnO_6 structures are covered with mesh. (d,e) The structures of HEA ($Ru_{13}Rh_{13}Ir_{13}Pd_{12}Pt_{13}$) and OH* complex with different atomic arrangements. (f) The DPPT analysis of HEA and OH* complex. The left and right frames in rectangles are respectively the digraph representations of (d) (pink) and (e) green at the distance filtration parameter 2.7 Å (left) and 3.4 Å (right). (g) APPT analysis of HEA and OH* complex. The left and right frames in rectangles are respectively the digraph representations of (d) (pink) and (e) green at the angle filtration stage 9 (left) and 23 (right). The $k = 36$, $m = 18$ is used to define the order set S^2 in this example. (h) An undirected tetrahedron as a basic building block for amorphous solid. (i) The schematic display of high entropy mixing in a tetrahedral lattice. At most 4 types of element will occupy the lattice sites randomly to form a high-entropy substructure in high entropy alloy nanoparticles. All 9 unique directed tetrahedral lattices. The two lattices that are mirror-symmetric to each other are framed by parallelogram boxes. (j) The 0-dimensional APPT analysis of two mirror lattices selected in (i). (k) An undirected octahedron as a basic building block for amorphous solid. At most 6 types of element will occupy the octahedral lattice sites randomly to form a high-entropy substructure in high entropy alloy nanoparticles. (l) The statistic analysis of all possible

directed octahedral lattices with at most 6 types elements. Different colors illustrate the four different combinations of Betti numbers $\beta_{0,1,2} = \beta_0, \beta_1, \beta_2$. The size of the circles indicates the number of unique atomic mixing scheme (2226 in total). The colored lines are connected to isomers.

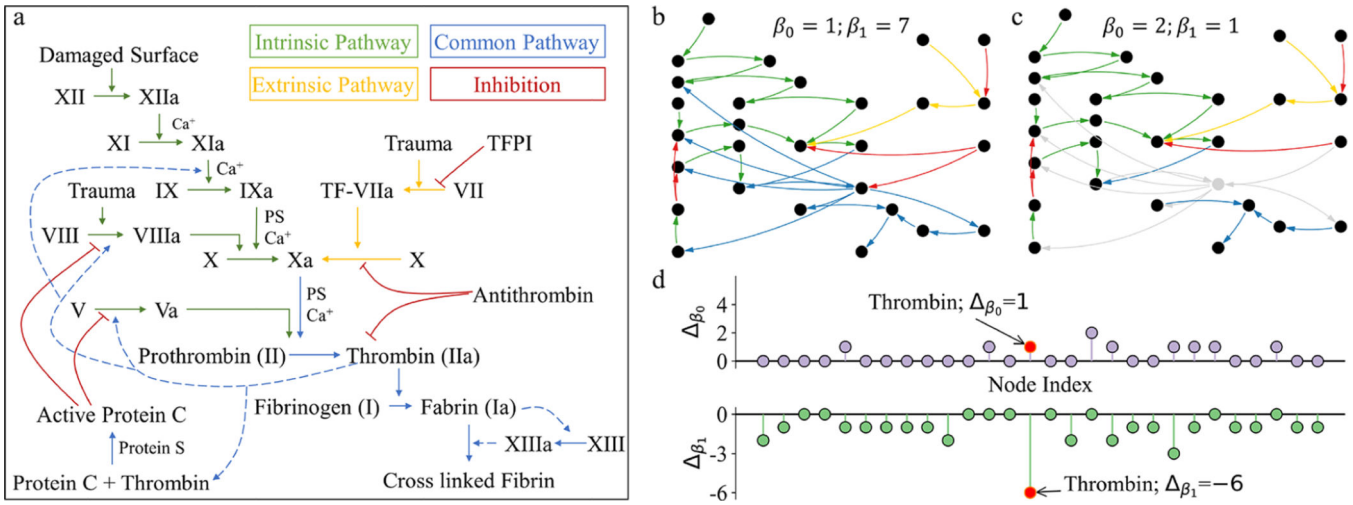


Figure 5.

Path homology-based topological perturbation analysis of blood coagulation pathways. (a) The classical blood coagulation pathways. The intrinsic pathway, extrinsic pathway, common pathway, and inhibition processes are colored in green, yellow, blue, and red, respectively. (b) The directed network representation of the blood coagulation pathway, whose path topological invariants in 0 and 1 dimensions are 1 and 7, respectively. (c) The condensation process after removing the key node (thrombin), and related edges are represented by a directed network, where the gray node and edges are removed objects. The resulting topological invariants are $\beta_0 = 2$ and $\beta_1 = 1$. (d) Topological perturbation analysis (TPA). The change in β_0 and β_1 caused by systematically removing one node and associated edges in the directed network. The topological changes in removing node thrombin are $\Delta\beta_1 = -6$ and $\Delta\beta_0 = 1$.