Chinese Journal of Structural Chemistry 42 (2023) 100120

Contents lists available at ScienceDirect

Chinese Journal of Structural Chemistry

journal homepage: www.journals.elsevier.com/chinese-journal-of-structural-chemistry

Review

Application of topology-based structure features for machine learning in materials science

Shisheng Zheng [a], Haowen Ding [a], Shunning Li [a], Dong Chen [a,b,**], Feng Pan [a,*]

[a] School of Advanced Materials, Peking University Shenzhen Graduate School, Shenzhen, Guangdong, 518055, PR China
[b] Department of Mathematics, Michigan State University, East Lansing, MI, USA

ARTICLE INFO

Keywords:
Machine learning
Structure feature
Structure graph
Algebraic topology

ABSTRACT

Structure features play an important role in machine learning models for the materials investigation. Here, two topology-based features for the representation of material structure, specifically structure graph and algebraic topology, are introduced. We present the fundamental mathematical concepts underlying these techniques and how they encode material properties. Furthermore, we discuss the practical applications and enhancements of these features made in specific material predicting tasks. This review may provide suggestions on the selection of suitable structural features and inspire creativity in developing robust descriptors for diverse applications.

1. Introduction

In the field of materials science, a vast amount of experimental and computational data on the properties and structures of different compounds has been accumulated. With access to databases containing millions of organic compounds [1,2] and over 200 thousands of inorganic compounds [3], it is essential to explore how to effectively manage and use these datasets to boost scientific breakthroughs. Machine learning, as a typical approach to data-driven patterns [4,5], is an important typology for accelerating statistical analysis of big data. It is a powerful tool that can extract and generalize potential underlying patterns from a large volume of existing data, and make predictions or classifications of unseen data. It has been extensively applied in various subdivisions of materials science (e.g., batteries [6,7], catalysts [8–12] and ferromagnetic materials [13,14]), enabling high-throughput screening of novel materials and predicting material properties.

The process of machine learning generally includes data collection and preprocessing, feature engineering, model selection and training, model evaluation and verification and model deployment and application [15]. Feature engineering involves the representation of material structures as descriptors for machine recognition. The appropriate representation of material structures through their relevant features is the key to enabling reliable predictions of material properties using machine learning [4]. These features serve as the building blocks that form the foundation of the model's ability to learn and identify patterns in the data, and can also highlight anomalies and unusual structures that may be of interest for further investigation [16]. Among the various structural features utilized in materials science, topology-based descriptors have unique advantages. Such representations are often characterized by their intuitive and concise nature, as well as their ability to effectively utilize mathematical operations to comprehend the complex structural features of materials. In this mini review, we introduce two structural features derived from topology-based mathematical sub-disciplines, namely structure graph and algebraic topology and demonstrate their utility in the realm of materials science.

2. Structure graph

Graph theory is a mathematical discipline in which graphs are the object of study. A graph, which exactly is the simplicial 1-complexes in topology, is an aggregate consisting of a set of vertices and a set of edges that represent connected relationships between the vertices. It is typically employed to represent relationships between objects, with vertices representing objects and edges connecting pairs of vertices that exhibit a particular relationship [17–19]. Material structures, consisting of atoms and chemical bonds, can be intuitively represented as structure graph [20–23]. In a structural graph, nodes represent atoms and edges are chemical bonds between atoms [24,25]. The constructed structure graphs contain only the topological information while neglecting the symmetry, bond lengths, and bond angles of the structures. The

* Corresponding author.
** Corresponding author.
E-mail addresses: chend@pku.edu.cn (D. Chen), panfeng@pkusz.edu.cn (F. Pan).

https://doi.org/10.1016/j.cjsc.2023.100120
Received 17 April 2023; Received in revised form 2 June 2023; Accepted 5 June 2023
Available online 7 June 2023

topological information itself can be applied to the classification and identification of materials [26]. The structure graph is directly related to an atomic adjacency matrix, where $a_{ij} = 1$ if atoms i and j are connected and $a_{ij} = 0$ if not, and thus is easily for machine recognition.

One can directly derive topological information from such structure graphs using graph theoretical quantifiers which can produce useful functions that correlate with properties of materials through the graph invariants or more frequently structural descriptors. Topological indices are one such structural descriptors which provide numeric functions of a molecular structure to facilitate quantitative structure-property/activity relationship (QSPR/QSAR) [27]. A large spectra of topological indices are available, among which the commonly utilized in QSPR/QSAR studies for grasping the relationships between the structure and the potential physicochemical characteristics are distance based, bond additive and degree based indices. Taking the field of zeolite research as an example, Clement et al. [28] have developed efficient technique to obtain exact analytical expressions for the various relativistic topological descriptors of the zeolite RHO structures (Fig. 1a and b) by graph-theoretical cut methods that reduce the complex structures with tunnels and cages into simpler graphs. All the acquired topological descriptors (Fig. 1c, d and e) are highly significant for the characterization of the QSAR properties of zeolite frameworks. Moreover, the developed topological indices that have the capability to include relativistic effects would be especially useful in the characterization of morphological changes to the materials that occur by the incorporation of heavier elements into the zeolite. They also developed topological indices to characterize other zeolite structures and to analyze the entropy measures of zeolite and benzenoid hydrocarbons [29–32].

When practically using structure graphs as features for machine learning, the graph is usually decomposed into subgraphs centered on different atoms to obtain the coordination environment of each atom [33, 34]. A simplified subgraph centered around an oxygen atom in spinel Co_3O_4 is shown in Fig. 2a [34]. The selection criterion involves incorporating all the atoms that are located within a bond-path distance of 4 from the central atom, resulting in a distinctive portrayal of the atoms arrangement within the unit cell of the compound. The crystal structure is represented by enumerating all the subgraphs of non-equivalent atoms in the compound and merging them together. In specific scenarios, the central node can also be a collection of several atoms [35].

Incorporating chemical information into the structure graph is crucial for enhancing the scientific validity and interpretability of the machine learning model. Wang et al. (Fig. 2b) utilized a crystal graph multilayer descriptor to represent the material structure for the prediction of thermodynamic stability, magnetic ground state, and band gap [14]. Specifically, in addition to the connectivity matrix, seven element feature matrices were created for endowing the feature physical meaning. According to the different predictive properties, three feature matrices are selected to construct the corresponding multilayer descriptor. The process of feature engineering not only offers a pathway to obtain physical and chemical insights into the relationship between descriptors and properties but also allows for customization of descriptors for various material properties, thereby establishing an adaptable machine learning framework for future research.

In the above mentioned structure graphs, edges generally represent chemical bonds between atoms. This can effectively capture the main chemical interactions in the structure, but it will ignore the possible weak interactions, such as van der Waals interactions, which are especially important in molecular systems. Pan et al. reported the element-specific multiscale weighted colored graph representations for the consideration of weak interactions in molecular systems to predict molecular toxicity [36]. For a molecule, it firstly determines the set of edges between particular pairs of atoms to obtain the corresponding subgraphs and the corresponding algebraic representation of each subgraph, such as the associated Laplacian and adjacency matrices (Fig. 3c and d); the strength of interatomic interactions is then evaluated by radial distribution

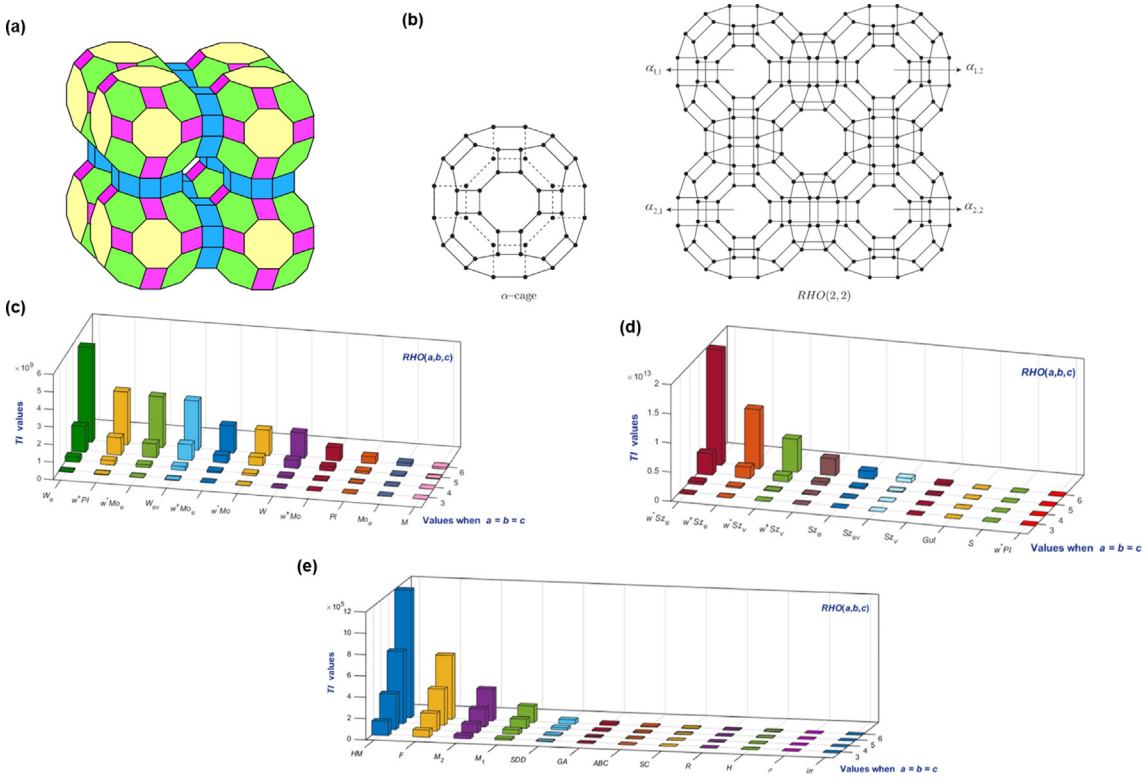

Fig. 1. (a) Zeolite RHO structure. (b) Primitive unit cell of zeolite RHO and its single layer. (c)–(e) Comparison of *TI* values for *RHO*(*a*, *b*, *c*) (Reprinted with permission from Ref. [28], Copyright 2022 Elsevier).

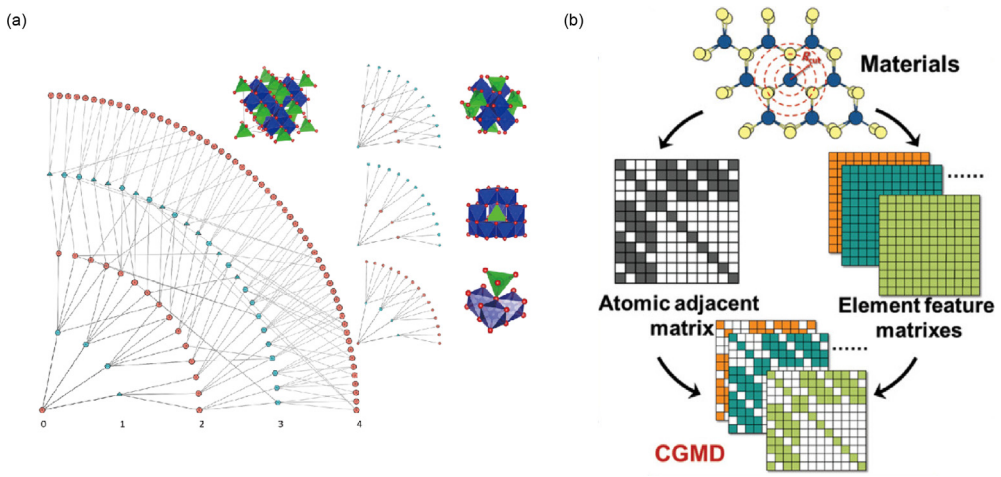

Fig. 2. (a) The subgraphs of spinel Co_3O_4 (Reprinted with permission from Ref. [34], Copyright 2019 Springer Nature). (b) Construction of element feature matrices for 2D materials (Reprinted with permission from Ref. [14], Copyright 2021 Wiley).

functions to distinguish between covalent and non-covalent interactions, and the corresponding function values are assigned to the Laplace and adjacency matrices (Fig. 3e); by performing statistical operations on the matrixes eigenvalues, a numerical representation of the molecules can be obtained as an input for machine learning. This feature is independent of molecular size, but can effectively represent the effect of weak interactions on the system properties. Based on a similar idea, a multiscale weighted spectral was developed to predict the structure of lithium clusters [37].

Although the structure graph has made great progresses, there are still problems to be solved, such as how to consider the bond angle, molecular orientation and other information that characterizes the three-dimensional structure in space, which may be important in some subfields of materials science (e.g. heterocatalysis and pharmacy).

3. Algebraic topology

Algebraic topology is a branch of mathematics that uses tools from abstract algebra to study topological spaces. Algebraic topology-based structural features can offer a feasible approach to encode compound structures using independent entities, rings, and topological faces of high dimensions in space [38–40]. Persistent homology is a powerful tool in algebraic topology to capture topological invariants in a changing scale, while more geometric information is preserved. By treating the atoms in a solid as a 3D point cloud in space, persistent homology can be used to construct structural features. Specifically, a family of complexes with different connectivity properties can be constructed by varying the filtration parameters, so that each moment of the complex may have a different topology. Thus, persistent homology can generate topological fingerprints that describe the presence of connected components, rings and voids in the structure during the change [41]. Connected components encode bond lengths and pairwise interactions, while holes and voids provide insights into many-body interactions and the spatial information. This approach offers a valuable perspective for materials science and can be directly used to characterize the difference between material structures [42–45].

Pan et al. developed a persistent-homology-based machine learning algorithm to accelerate the search of the globally stable structure of clusters [46] (Fig. 4a). They introduced the idea of persistent pairwise independence, which involves counting the number of independent atom pairs throughout the filtration process. The change in persistent pairwise independence throughout filtration provides information about the

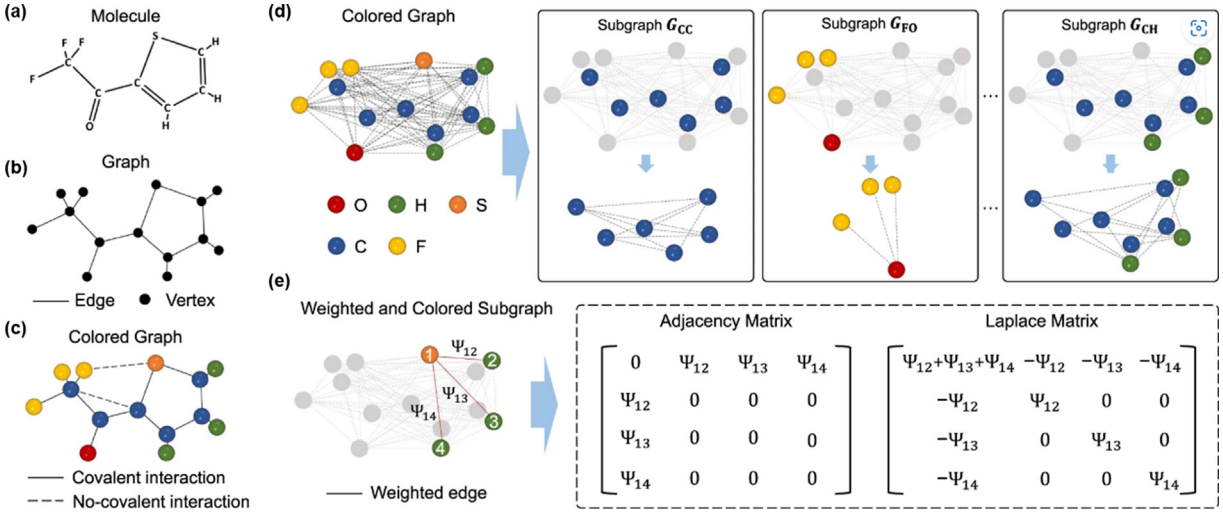

Fig. 3. Illustration of weighted colored element-specific algebraic graphs. (a) The molecular structure of 2-trifluoroacetyl. (b) A traditional graph representation and (c) a colored graph representation. (d) Illustration of the process of decomposing a colored graph into element-specific CC, FO, and CH subgroups, where element refers to the chemical element in this study, e.g., H, C, N. (e) Illustration of weighted colored element-specific subgraph G_{SH}, its adjacency matrix, and Laplacian matrix, where Ψ refers to the weight of the edge in subgraph (Reprinted with permission from Ref. [36], Copyright 2022 Springer Nature).

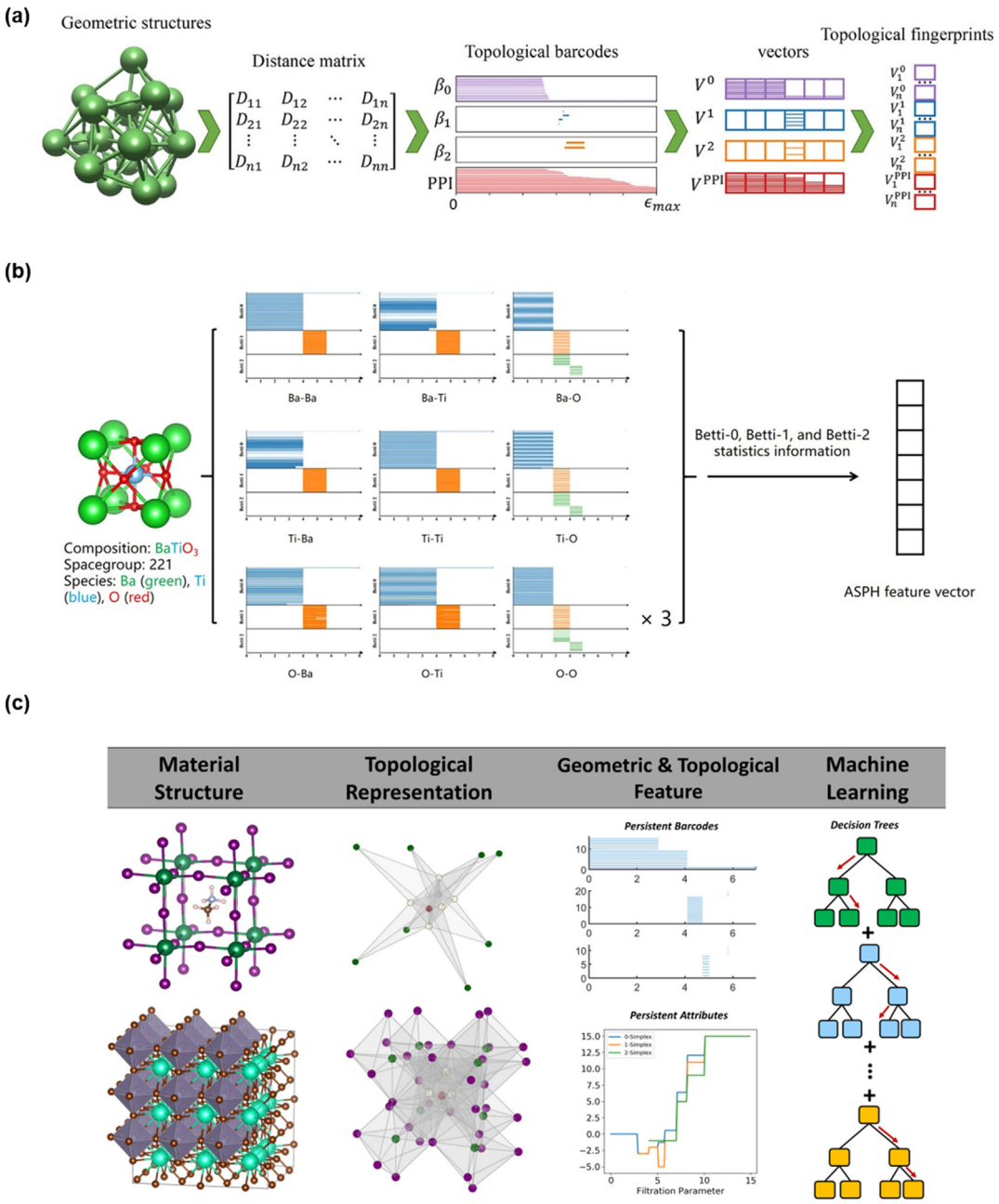

Fig. 4. (a) The flowchart for the construction of topological fingerprint of a Li cluster (Reprinted with permission from Ref. [46], Copyright 2020 American Chemical Society). (b) The construction of element-specific topological descriptor for $BaTiO_3$ (Reprinted with permission from Ref. [48], Copyright 2021 Springer Nature). (c) The flowchart of our persistent homology-based machine learning models for the prediction of formation energy and bandgap of organic-inorganic halide perovskite (Reprinted with permission from Ref. [49], Copyright 2021 Springer Nature).

interatomic distances between atom pairs. The resulting persistent barcodes and pairwise independence measures are then merged and transformed into 1D vectors that are invariant to translations and rotations. These vectors can serve as structural features for machine learning applications. To demonstrate the effectiveness of this approach, the researchers used Li_n ($3 \leq n \leq 10$) clusters as examples. The resulting descriptors are able to capture the existence of small cages inside the clusters and encode the corresponding geometric information. This success indicates that persistent homology has the potential to extract important topological information related to interstitial voids and macroscopic pores in materials, which can improve the accuracy of structure prediction in large clusters (Li_{20} and Li_{40}).

Despite the capability of persistent homology to capture both local and global structural information at once, it is not a common practice to include the element information in the structural features created using this method. To overcome this challenge, one approach is to generate persistent barcodes for different atom subsets in the compound, incorporate the element information, and finally merge all the element-specific barcodes together [47].

An example that predicts the formation energy of crystalline compounds has been demonstrated by Pan et al. [48] (Fig. 4b). To account for periodicity in the construction of persistent barcodes for the central atom, a cutoff radius is introduced to determine the range of constituent atoms involved. The $BaTiO_3$ is taken as a case, where nine different barcodes are generated by considering all possible combinations of atom pairs from Ba, Ti, and O. The statistics information of these barcodes is merged with the element information to create a statistical representation for $BaTiO_3$. The element-specific persistent homology approach

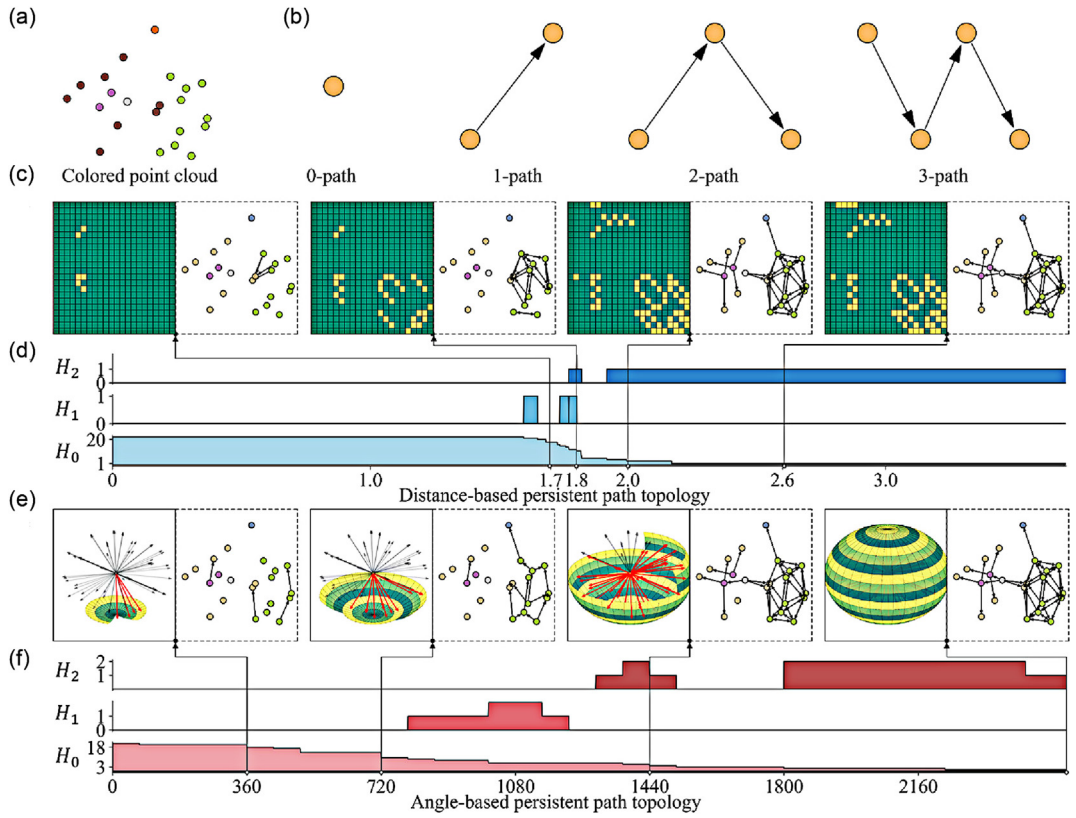

Fig. 5. Illustration of persistent path topology. (a) The weight function-based matrix is constructed from a molecular structure. (b) Illustration of the basic component that makes up the path complex, p-path, where p = 0, 1, 2, and 3. (c) Illustration of the distance-based filtration. As the filtration parameter increases, the path complex based on the weight matrix expands accordingly. In the left figure, the *x*- and *y*-axes 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. (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 (Reprinted with permission from Ref. [53], Copyright 2023 American Chemical Society).

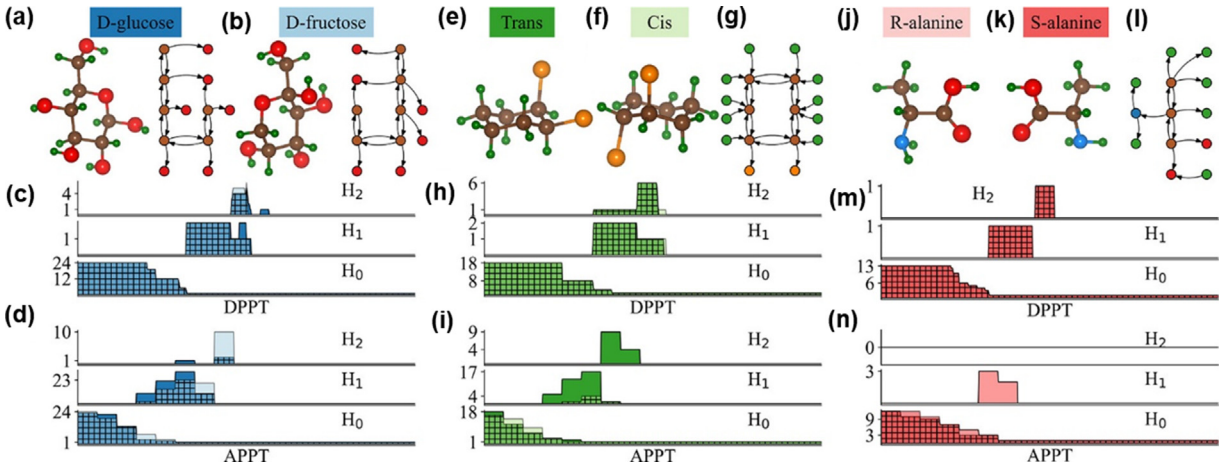

Fig. 6. Illustration of the DPPT and APPT analysis of spatial isomers. (a) The molecular structures of D-glucose (left) and associated digraphs (right). (b) The molecular structures of D-fructose (left) and associated digraphs (right). (c) The DPPT analysis of D-glucose and D-fructose. Shared parts are 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- and *cis*-1,2-dichlorocyclohexane. Shared parts are plotted in mesh. (i) The APPT analysis of *trans*-1,2- and *cis*-1,2-dichlorocyclohexane. Shared parts are depicted in mesh. (j) The structure of R-alanine. (k) The structure of Salanine. (l) Shared digraph representation of R- and S-alanine. (m) The DPPT analysis of R- and S-alanine. Shared parts are plotted in mesh. (n) The APPT analysis of R- and S-alanine. Shared parts are presented in mesh (Reprinted with permission from Ref. [53], Copyright 2023 American Chemical Society).

preserves the physical and chemical information of the compound during topological abstraction. However, it is worth noting that the derived feature depends on the selected cutoff radius to handle periodicity, which may lead to biases and affect the predictive ability of machine learning models. Sum et al. [49] (Fig. 4c) extended the element-specific persistent homology representation to the prediction of formation energy and bandgap of organic-inorganic halide perovskite, which achieved significant accuracy advantages over traditional descriptor based on machine learning models.

Through the above operations, the information of the elements can be considered. However, persistent-homology is insensitive to asymmetric or directional relations in essence. For the material structure, the chemical bonds are usually spatial directional and polar, which potentially has impacts on the properties of the material. In mathematics, path homology (PH) proposed by Grigor'yan et al. can be used for the directed networks [50,51]. Persistent path topology (PPT) was introduced to further empower path homology by multiscale filtration [52]. Pan et al. introduced this method into the field of materials [53]. The point cloud that labels different atom types by colors is taken as an example (Fig. 5a). Unlike persistent homology, which considers all atoms equally, PPT focuses on paths that can be identified by bond polarity based on electronegativities of the atoms. The four simplest paths, i.e., 0-, 1-, 2-, and 3-path, are shown in Fig. 5b and serve as the fundamental elements for constructing the p-path complex, which is a topological space that encompasses all possible paths in the system and characterized by the longest path length p. Similar to the operation in persistent-homology, distance can be used as a criterion to obtain topological invariants via the filtration in PPT, namely distance-based persistent path topology (DPPT) (Fig. 5c and d). The difference is that in PPT, the angle can also be used as a filtering standard to characterize the structure of the material, namely angle-based persistent path topology (APPT) (Fig. 5e and f). The authors showcase the effectiveness of two PPT techniques in distinguishing and identifying isomers in molecular systems, including D-fructose and D-glucose isomers, *cis-trans* isomers and chiral molecules (Fig. 6). Additionally, PPT is applied to study and characterize the Jahn-Teller effect and differentiate isomers of high entropy alloy catalysts, highlighting its ability to reveal the underlying physical and chemical properties induced by atomic ordering. Although this encoding scheme was originally demonstrated by characterizing the difference of material structure, its utilization in machine learning models is also envisaged.

4. Summary and perspective

This review provides a brief overview of two topology-based structural features: structure graph and algebraic topology. They are capable of abstracting material structures and extracting essential feature information, such as chemical bond connectivity. As they can effectively capture interactions among nearest neighbors, they are well-suited for investigating chemical reactions. Additional factors such as element information, spatial information, and weak interactions can also be incorporated into these features. While richer information will help improve the interpretability and accuracy of the model, this will inevitably incur a corresponding computational cost. Therefore, a balance between the efficiency and interpretability of machine learning models should be considered in practical application scenarios.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

The authors acknowledge financial support from the Guangdong Basic and Applied Basic Research Foundation (2020A1515110843), Young S&T Talent Training Program of Guangdong Provincial Association for S&T (SKXRC202211), Chemistry and Chemical Engineering Guangdong Laboratory (1922018), Soft Science Research Project of Guangdong Province (2017B030301013), National Natural Science Foundation of China (22109003), Natural Science Foundation of Shenzhen (JCYJ20190813110605381) and the Major Science and Technology Infrastructure Project of Material Genome Big-science Facilities Platform supported by Municipal Development and Reform Commission of Shenzhen.

References

[1] A. Gaulton, A. Hersey, M. Nowotka, A.P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L.J. Bellis, E. Cibrian-Uhalte, M. Davies, N. Dedman, A. Karlsson, M.P. Magarinos, J.P. Overington, G. Papadatos, I. Smit, A.R. Leach, The ChEMBL database in 2017, Nucleic Acids Res. 45 (2017) D945–D954, https://doi.org/10.1093/nar/gkw1074.
[2] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, J. Wang, B. Yu, J. Zhang, S.H. Bryant, PubChem substance and compound databases, Nucleic Acids Res. 44 (2016) D1202–D1213, https://doi.org/10.1093/nar/gkv951.
[3] Inorganic Crystal Structure Database (ICSD). https://icsd.fiz-karlsruhe.de/index.xhtml.http://doi.org/10.14102/j.cnki.0254-5861.2011-2011-2756.
[4] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature 559 (2018) 547–555, https://doi.org/10.1038/s41586-018-0337-2.
[5] M.I. Jordan, T.M. Mitchell, Machine learning: trends, perspectives, and prospects, Science 349 (2015) 255–260, https://doi.org/10.1126/science.aaa8415.
[6] E. Chemali, P.J. Kollmeyer, M. Preindl, A. Emadi, State-of-charge estimation of Li-ion batteries using deep neural networks: a machine learning approach, J. Power Sources 400 (2018) 242–255, https://doi.org/10.1016/j.jpowsour.2018.06.104.
[7] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, K. Dietmayer, Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods, J. Power Sources 239 (2013) 680–688, https://doi.org/10.1016/j.jpowsour.2012.11.146.
[8] M. Andersen, K. Reuter, Adsorption enthalpies for catalysis modeling through machine-learned descriptors, Acc. Chem. Res. 54 (2021) 2741–2749, https://doi.org/10.1021/acs.accounts.1c00153.
[9] S. Ma, Z.-P. Liu, Machine learning for atomic simulation and activity prediction in heterogeneous catalysis: current status and future, ACS Catal. (2020) 13213–13226, https://doi.org/10.1021/acscatal.0c03472.
[10] J.A. Esterhuizen, B.R. Goldsmith, S. Linic, Interpretable machine learning for knowledge generation in heterogeneous catalysis, Nat. Catal. 5 (2022) 175–184, https://doi.org/10.1038/s41929-022-00744-z.
[11] H. Mai, T.C. Le, D. Chen, D.A. Winkler, R.A. Caruso, Machine learning for electrocatalyst and photocatalyst design and discovery, Chem. Rev. 122 (2022) 13478–13515, https://doi.org/10.1021/acs.chemrev.2c00061.
[12] J. Xu, X.M. Cao, P. Hu, Perspective on computational reaction prediction using machine learning methods in heterogeneous catalysis, Phys. Chem. Chem. Phys. 23 (2021) 11155–11179, https://doi.org/10.1039/d1cp01349a.
[13] S. Lu, Q. Zhou, Y. Guo, J. Wang, On-the-fly Interpretable machine learning for rapid discovery of two-dimensional ferromagnets with high curie temperature, Chem 8 (2021) 769–783, https://doi.org/10.1016/j.chempr.2021.11.009.
[14] S. Lu, Q. Zhou, Y. Guo, Y. Zhang, Y. Wu, J. Wang, Coupling a crystal graph multilayer descriptor to active learning for rapid discovery of 2D ferromagnetic semiconductors/half-metals/metals, Adv. Mater. 32 (2020) 2002658, https://doi.org/10.1002/adma.202002658.
[15] Y.J. Chen, H.Y. Lan, Z.D. Du, S.L. Liu, J.H. Tao, D. Han, T. Luo, Q. Guo, L. Li, Y. Xie, T.S. Chen, An instruction set architecture for machine learning, ACM Trans. Comput. Syst. 36 (2019) 9.
[16] S. Li, Y. Liu, D. Chen, Y. Jiang, Z. Nie, F. Pan, Encoding the atomic structure for machine learning in materials science, WIREs Comput. Mol. Sci. 12 (2021) e1558, https://doi.org/10.1002/wcms.1558.
[17] D.B. West, Introduction to Graph Theory, Prentice hall Upper Saddle River, 2001.
[18] B. Bollobás, Modern Graph Theory, Springer Science & Business, Media, 1998.
[19] W.T. Tutte, W.T. Tutte, Graph Theory, Cambridge university press, 2001.
[20] R. Garcia-Domenech, J. Galvez, J.V. de Julian-Ortiz, L. Pogliani, Some new trends in chemical graph theory, Chem. Rev. 108 (2008) 1127–1169, https://doi.org/10.1021/cr0780006.
[21] J.R. Boes, O. Mamun, K. Winther, T. Bligaard, Graph theory approach to high-throughput surface adsorption structure generation, J. Phys. Chem. A 123 (2019) 2281–2285, https://doi.org/10.1021/acs.jpca.9b00311.
[22] S. Deshpande, T. Maxson, J. Greeley, Graph theory approach to determine configurations of multidentate and high coverage adsorbates for heterogeneous catalysis, NPJ Comput. Mater. 6 (2020) 79, https://doi.org/10.1038/s41524-020-0345-2.
[23] E.A. Walker, M.M. Mohammadi, M.T. Swihart, Graph theory model of dry reforming of methane using Rh(111), J. Phys. Chem. Lett. 11 (2020) 4917–4922, https://doi.org/10.1021/acs.jpclett.0c01038.
[24] S. Kozuch, Steady state kinetics of any catalytic network: graph theory, the energy span model, the analogy between catalysis and electrical circuits, and the meaning of “mechanism”, ACS Catal. 5 (2015) 5242–5255, https://doi.org/10.1021/acscatal.5b00694.

[25] L. Kollias, R. Rousseau, V.A. Glezakou, M. Salvalaglio, Understanding metal-organic framework nucleation from a solution with evolving graphs, J. Am. Chem. Soc. 144 (2022) 11099–11109, https://doi.org/10.1021/jacs.1c13508.
[26] S. Li, Z. Chen, Z. Wang, M. Weng, J. Li, M. Zhang, J. Lu, K. Xu, F. Pan, Graph-based discovery and analysis of atomic-scale one-dimensional materials, Natl. Sci. Rev. 9 (2022) nwac028, https://doi.org/10.1093/nsr/nwac028.
[27] M. Arockiaraj, J. Clement, D. Paul, K. Balasubramanian, Relativistic distance-based topological descriptors of Linde type A zeolites and their doped structures with very heavy elements, Mol. Phys. 119 (2020) e1798529, https://doi.org/10.1080/00268976.2020.1798529.
[28] M. Arockiaraj, D. Paul, S. Klavžar, J. Clement, S. Tigga, K. Balasubramanian, Relativistic distance based and bond additive topological descriptors of zeolite RHO materials, J. Mol. Struct. 1250 (2022) 131798, https://doi.org/10.1016/j.molstruc.2021.131798.
[29] D. Paul, M. Arockiaraj, K. Jacob, J. Clement, Multiplicative versus scalar multiplicative degree based descriptors in QSAR/QSPR studies and their comparative analysis in entropy measures, Eur. Phys. J. Plus 138 (2023) 323, https://doi.org/10.1140/epjp/s13360-023-03920-7.
[30] M. Arockiaraj, J. Clement, D. Paul, K. Balasubramanian, Quantitative structural descriptors of sodalite materials, J. Mol. Struct. 1223 (2021) 128766, https://doi.org/10.1016/j.molstruc.2020.128766.
[31] M. Arockiaraj, D. Paul, S. Klavžar, J. Clement, S. Tigga, K. Balasubramanian, Relativistic topological and spectral characteristics of zeolite SAS structures, J. Mol. Struct. 1270 (2022) 133854, https://doi.org/10.1016/j.molstruc.2022.133854.
[32] K. Jacob, J. Clement, M. Arockiaraj, D. Paul, K. Balasubramanian, Topological characterization and entropy measures of tetragonal zeolite merlinoites, J. Mol. Struct. 1277 (2023) 134786, https://doi.org/10.1016/j.molstruc.2022.134786.
[33] Z.W. Ulissi, A.J. Medford, T. Bligaard, J.K. Norskov, To address surface reaction network complexity using scaling relations machine learning and DFT calculations, Nat. Commun. 8 (2017) 14621, https://doi.org/10.1038/ncomms14621.
[34] M. Weng, Z. Wang, G. Qian, Y. Ye, Z. Chen, X. Chen, S. Zheng, F. Pan, Identify crystal structures by a new paradigm based on graph theory for building materials big data, Sci. China Chem. 62 (2019) 982–986, https://doi.org/10.1007/s11426-019-9502-5.
[35] P.G. Ghanekar, S. Deshpande, J. Greeley, Adsorbate chemical environment-based machine learning framework for heterogeneous catalysis, Nat. Commun. 13 (2022) 5788, https://doi.org/10.1038/s41467-022-33256-2.
[36] D. Chen, K. Gao, D.D. Nguyen, X. Chen, Y. Jiang, G.W. Wei, F. Pan, Algebraic graph-assisted bidirectional transformers for molecular property prediction, Nat. Commun. 12 (2021) 3521, https://doi.org/10.1038/s41467-021-23720-w.
[37] S. Ma, S. Zheng, W. Zhang, D. Chen, F. Pan, Algebraic graph-based machine learning model for Li-cluster prediction, J. Phys. Chem. A 127 (2023) 2051–2059, https://doi.org/10.1021/acs.jpca.3c00272.
[38] S. Lefschetz, Algebraic Topology, American Mathematical Soc, 1942.
[39] E.H. Spanier, Algebraic Topology, Springer Science & Business, Media, 1989.
[40] T. tom Dieck, Algebraic Topology, European Mathematical Society, 2008.
[41] G. Carlsson, Topology and data, Bull. Am. Math. Soc. 46 (2009) 255–308.
[42] D. Ushizima, D. Morozov, G.H. Weber, A.G. Bianchi, J.A. Sethian, E.W. Bethel, Augmented topological descriptors of pore networks for material science, IEEE Trans. Vis. Comput. Graph. 18 (2012) 2041–2050, https://doi.org/10.1109/TVCG.2012.200.
[43] J. Grbić, J. Wu, K. Xia, G.-W. Wei, Aspects of topological approaches for data science, Found. Data Sci. 4 (2022) 165–216, https://doi.org/10.3934/fods.2022002.
[44] K. Xia, X. Feng, Y. Tong, G.W. Wei, Persistent homology for the quantitative prediction of fullerene stability, J. Comput. Chem. 36 (2015) 408–422, https://doi.org/10.1002/jcc.23816.
[45] Y. Hiraoka, T. Nakamura, A. Hirata, E.G. Escolar, K. Matsue, Y. Nishiura, Hierarchical structures of amorphous solids characterized by persistent homology, Proc. Natl. Acad. Sci. USA 113 (2016) 7035–7040, https://doi.org/10.1073/pnas.1520877113.
[46] X. Chen, D. Chen, M. Weng, Y. Jiang, G.W. Wei, F. Pan, Topology-based machine learning strategy for cluster structure prediction, J. Phys. Chem. Lett. 11 (2020) 4392–4401, https://doi.org/10.1021/acs.jpclett.0c00974.
[47] K.D. Wu, Z.X. Zhao, R.X. Wang, G.W. Wei, Top*P–S*: persistent homology-based multi-task deep neural networks for simultaneous predictions of partition coefficient and aqueous solubility, J. Comput. Chem. 39 (2018) 1444–1454, https://doi.org/10.1002/jcc.25213.
[48] Y. Jiang, D. Chen, X. Chen, T. Li, G.W. Wei, F. Pan, Topological representations of crystalline compounds for the machine-learning prediction of materials properties, NPJ Comput. Mater. 6 (2022) 45–52, https://doi.org/10.1038/s41524-021-00493-w.
[49] D.V. Anand, Q. Xu, J. Wee, K. Xia, T.C. Sum, Topological feature engineering for machine learning based halide perovskite materials design, NPJ Comput. Mater. 8 (2022) 203, https://doi.org/10.1038/s41524-022-00883-8.
[50] A. Grigor'yan, Y. Lin, Y.V. Muranov, S.-T. Yau, Path complexes and their homologies, J. Math. Sci. 248 (2020) 564–599, https://doi.org/10.1007/s10958-020-04897-9.
[51] A. Grigoryan, R. Jimenez, Y. Muranov, S.T. Yau, On the path homology theory of digraphs and eilenberg-steenrod axioms, Homol. Homotopy Appl. 20 (2018) 179–205, https://doi.org/10.4310/HHA.2018.v20.n2.a9.
[52] S. Chowdhury, F. Mémoli, Persistent Path Homology of Directed Networks, Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2018, pp. 1152–1169, https://doi.org/10.1137/1.9781611975031.7.
[53] D. Chen, J. Liu, J. Wu, G.W. Wei, F. Pan, S.T. Yau, Path topology in molecular and materials sciences, J. Phys. Chem. Lett. 14 (2023) 954–964, https://doi.org/10.1021/acs.jpclett.2c03706.