

# PATH COMPLEX MESSAGE PASSING FOR MOLECULAR PROPERTY PREDICTION

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Geometric deep learning (GDL) has demonstrated enormous power in molecular data analysis. However, GDL faces challenges in achieving high efficiency and expressivity in molecular representations when high-order terms of the atomic force fields are not sufficiently learned. In this work, we introduce message passing on path complexes, called the Path Complex Message Passing, for molecular prediction. Path complexes represent the geometry of paths and can model the chemical and non-chemical interactions of atoms in a molecule across various dimensions. Our model defines messages on path complexes and employs neural message passing to learn simplex features, enabling feature communication within and between different dimensions. Since messages on high-order and low-order path complexes reflect different aspects of molecular energy, they are updated sequentially according to their order. The higher the order of the path complex, the richer the information it contains, and the higher its priority during inference. It can thus characterize various types of molecular interactions specified in molecular dynamics (MD) force fields. Our model has been extensively validated on benchmark datasets and achieves state-of-the-art results. The code is available at <https://anonymous.4open.science/r/Path-Complex-Neural-Network-32D6>

## 1 INTRODUCTION

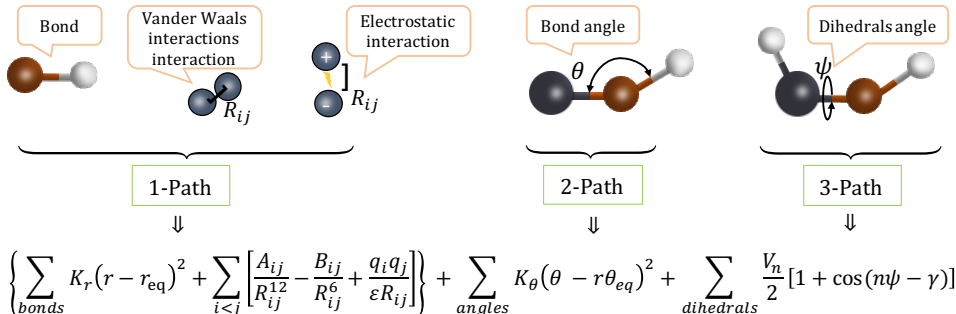
Accurate prediction of molecular properties is crucial in fields such as drug design Zhang et al. (2017); Chen et al. (2018); Mak & Pichika (2019); Chan et al. (2019), biology Townshend et al. (2021); Jamasb et al. (2022), chemistry Qiao et al. (2022), and materials science Vlassis et al. (2020). Geometric Deep Learning (GDL) has demonstrated significant potential in molecular sciences, leading to a surge in studies employing GDL models for effective molecular representation learning Bronstein et al. (2017); Atz et al. (2021); Ingraham et al. (2023). Among the three types of representations used in GDL models—topological, geometric, and functional—the molecular graph has become the most popular due to its simplicity, flexibility, and efficiency Wieder et al. (2020); Yu & Gao (2022); Atz et al. (2021); Li et al. (2022); Wang et al. (2022b). However, relying solely on graph representations fails to capture the many-body interactions inherent in complex systems, thereby limiting the expressiveness and predictive power of this approach Bodnar et al. (2021b). This paper develops a path complex-based neural message passing for molecule prediction, where the molecular energy of force field can be well represented.

In Graph Neural Networks (GNNs), the molecular graph is typically constructed based on covalent bonds. Node features are usually derived from atomic properties and are updated by aggregating information from neighboring nodes Huang et al. (2020); Shindo & Matsumoto (2019); Shui & Karypis (2020a); Schütt et al. (2017); Unke & Meuwly (2019). To enhance GNN performance, researchers have proposed several approaches. One major strategy is to design more complex molecular graphs that incorporate non-covalent interactions. The most common method involves introducing edges between any two atoms within a specified cutoff distance, effectively capturing non-covalent interactions. Additionally, molecule-based line graph models have been developed, where nodes represent atomic bonds and edges represent bond angles Choudhary & DeCost (2021).

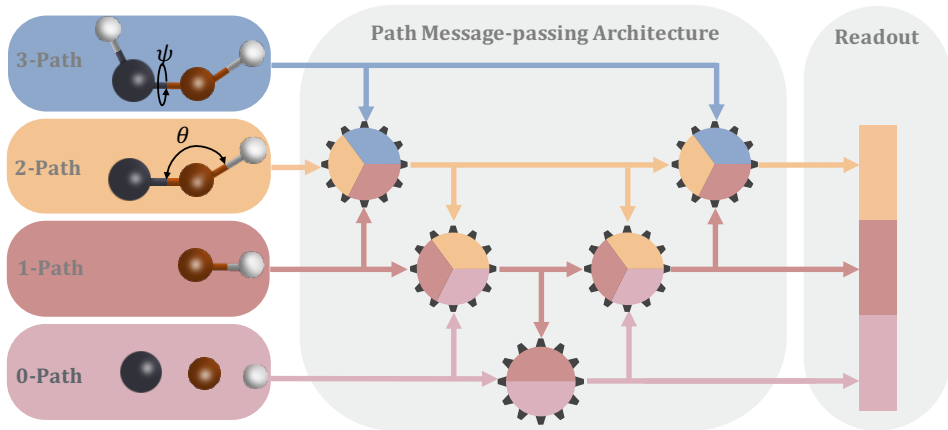
The second approach focuses on incorporating global physical features and local geometric information into GNN models. Global physical attributes such as temperature, pressure, and entropy

054 have been added to GNN architectures to better characterize molecular states and environments,  
 055 as demonstrated in MEGNet Chen et al. (2019) and SphereNet Liu et al. (2022). Local geometric  
 056 features—particularly bond lengths, bond angles Schütt et al. (2018); Flam-Shepherd et al. (2021),  
 057 dihedral angles Wang et al. (2022a), and torsion angles, which are crucial to molecular proper-  
 058 ties—have been extensively considered in models such as DimeNet Gasteiger et al. (2020), GemNet  
 059 Gasteiger et al. (2021), ALIGNN Choudhary & DeCost (2021), and GEM Fang et al. (2022).

060 Another approach involves designing efficient message-passing modules for invariant features,  
 061 equivariant properties, and higher-order tensors. The expressivity of GNNs is closely related to  
 062 the message-passing mechanisms used in layers that process invariant, equivariant, or higher-order  
 063 tensor features. These three approaches are often synergistically integrated to enhance model per-  
 064 formance.



077 Figure 1: Terms of the approximate equation to molecular dynamics force field correspond to path  
 078 complices of order one to three, which have been used in path complex message passing.



094 Figure 2: The architecture of PCMP utilizing path complexes up to order 3 is depicted. At each  
 095 layer  $l$ , each path complex message of a given order updates its features using messages from path  
 096 complexes of the same order and adjacent orders from the previous layer  $l - 1$ . Higher-order path  
 097 complex messages are updated before lower-order ones because the former encompass the paths of  
 098 the latter. Additionally, the interplay between high-order and low-order path complexes is learned  
 099 through message passing.

100

101 In this work, we develop path complex-based molecular representation and path complex message  
 102 passing (PCNN) model for molecular property analysis. PCNN is a neural message passing Gilmer  
 103 et al. (2017) on path complexes. A path is a sequence of points, and a path complex is a subset of all  
 104 possible paths. In the context of a molecule or molecular graph, a path corresponds to the geometry  
 105 defined by chemical or non-chemical bonds. Our path complexes are specifically designed — based  
 106 on molecular graphs that include both covalent and non-covalent bonds — to characterize different  
 107 types of energy specified in molecular dynamics (MD) force fields, as shown in Figure 1. The MD  
 potential energy Mayo et al. (1990); González (2011); Leach (2001) comprises bond terms ( $E_B$ ,

two-body interactions), bond-angle terms ( $E_A$ , three-body interactions), and dihedral-angle terms ( $E_T$ , four-body interactions), which are effectively characterized by our 1-path, 2-path, and 3-path features, respectively.

Each path complex is assigned a message. Similar to classical neural message passing on graphs Gilmer et al. (2017) and simplicial complexes Bodnar et al. (2021a;b), the propagation of messages in a path complex is influenced by the messages of its “adjacent” path complexes at different orders. A higher-order path complex contains longer paths and includes the shorter paths of lower-order path complexes. Therefore, we need to update messages according to the order of the path complexes: messages in higher-order path complexes have priority in being updated. However, since path complexes of adjacent orders are interconnected, we incorporate interactions between higher-order and lower-order path complex information during message passing. Specifically, the higher-order message first updates the lower-order one, and then the updated lower-order message exerts a reverse effect on the higher-order information. Figure 2 illustrates a neural message passing process among path complexes of different orders designed based on this principle.

PCNN thus enables information passing between path complex features, using the aggregated information to predict molecular properties. Testing on benchmark datasets demonstrates promising performance. Our contributions are as follows:

1. We have developed a path complex-based molecular representation that explicitly characterizes different terms in the molecular dynamics (MD) force field for molecular property prediction.
2. We propose constructing path complexes using the unique topologies of graphs, simplicial complexes, and hypergraphs. This method, based on path complex message passing, enables the exchange of information across multiple dimensions.
3. Our PCMP model has been rigorously tested and validated on benchmark molecular tasks, consistently achieving state-of-the-art results.

## 2 RELATED WORK

**Graph Neural Networks for Molecular Property Prediction** Graph neural network models have played a pivotal role in molecular data analysis. Traditional GNN models represent molecules as the de facto covalent-bond-based molecular graphs, and use major GNN architectures, such as GIN Xu et al. (2018), GAT Velickovic et al. (2017), GCN Kipf & Welling (2016a), SGCN Danel et al. (2020) and GTransformer Rong et al. (2020), to learn molecular properties Yang et al. (2019); Xiong et al. (2019); Choudhary & DeCost (2021); Fang et al. (2022). With the importance of non-covalent bonds, cutoff-distance-based molecular graph representations have been widely employed in GNN models, such as DimeNet Gasteiger et al. (2020), HMGNN Shui & Karypis (2020b), GeoGNN Fang et al. (2022), Mol-GDL Shen et al. (2023), etc. Further, higher-order interactions (beyond pair-wise forces) has been explicitly incorporated into GNN models, including ALIGNN Choudhary & DeCost (2021), GEM Fang et al. (2022), DimeNet Gasteiger et al. (2020), GemNet Gasteiger et al. (2021), etc, by the consideration of bond angles, dihedral angles, torsion angles, and other local geometric information. In particular, these higher-order terms can be directly related to MD force field information Halgren (1996); Choudhary et al. (2018). Finally, pre-training process has been adopted to further improve the accuracy of GNN models, such as N-Gram Liu et al. (2019), PretrainGNN Hu et al. (2019), GEM Fang et al. (2022), MolCLR Wang et al. (2022b), DMP Zhu et al. (2023), etc.

**Topological Deep Learning (TDL)** Topological Deep Learning (TDL) Hajij et al. (2022); Bodnar (2022) leverages novel topological tools to characterize data with complicated higher-order structures. Different from graph-based data representation, TDL uses topological representations from algebraic topology, including simplicial complexes Bodnar (2022); Schaub et al. (2022), cell complexes Hajij et al. (2020); Roddenberry et al. (2022); Giusti et al. (2023), sheaves Hansen & Ghrist (2019); Bodnar et al. (2021b), hypergraphs Feng et al. (2019); Kim et al. (2020); Bai et al. (2021), and combinatorial complexes Hajij et al. (2022) to model not only pair-wise interactions (as in graphs), but also higher-order interactions among three or more elements. In fact, these algebraic topology-based molecular representations have already achieved great success in molecular data

analysis, including protein flexibility and dynamic analysis Xia & Wei (2014); Sverrisson et al. (2021), drug design Cang & Wei (2017), virus analysis Chen et al. (2022), materials property analysis Reiser et al. (2022); Townsend et al. (2020). Further, TDL uses a generalized message-passing mechanism thus enables the communication of information from simplices of different dimensions. In contrast to GNNs, where information is passing among nodes or edges, TDL allows information to propagate through any neighborhood relation Roddenberry et al. (2021).

Recently, path complex and its related models, including path homology Grigor’yan et al. (2018), persistent path homology Chowdhury & Mémoli (2018); Liu et al. (2023); Chen et al. (2023), path Laplacian Wang & Wei (2023), a special path-complex-based topological message passing model Truong & Chin (2024) has been developed and demonstrated great potential for the analysis of molecular structures.

**Geometric Deep Learning and Molecular Representation** Generally speaking, molecules in GDL models are characterized by three types of molecular representations, including topological representations (such as molecular graphs), geometric representation (such as molecular surfaces), and function representation (such as molecular density). Deep learning models including (3D) convolutional neural networks, graph neural networks (GNNs), recurrent neural networks, and others, have been constructed based on these representations Wieder et al. (2020); Yu & Gao (2022); Atz et al. (2021); Li et al. (2022); Wang et al. (2022b). With its simplicity, flexibility and efficiency, molecular graphs are the most popular of various types of GNN models have been proposed, including graph recurrent neural networks (GraphRNN) You et al. (2018), graph convolutional networks (GCN) Welling & Kipf (2016), graph autoencoders Kipf & Welling (2016b), graph transformers Yun et al. (2019), etc. These GNN models have been widely used in molecular data analysis.

### 3 PATH COMPLEX MESSAGE PASSING

Path complex was originally developed on directed graph (or digraph) and set, by Grigoryan, Lin, Muranov and Yau in 2012 Grigor’yan et al. (2012). They also proposed a new homology theory for path complex, called path homology, and use it to explore topological invariant information of digraphs Grigor’yan et al. (2014). Mathematically, path homology provides a novel framework to systematically explore intrinsic topological information of more general structures Grigor’yan et al. (2019); Grigor’yan et al. (2020). Details of path complex and path homology can be found in the Appendix B.

Here we propose a generalized way to construct path complex based on undirected graph, simplicial complex, and hypergraph. On undirected graph, we propose graph collapse and expansion operations, and use them to systematically study graph isomorphism by their path complex homology groups. We found that the path complex homology is a graph weak isomorphism invariant. For simplicial complex and hypergraph, we propose simplex- and hyperedge- based path complex.

#### 3.1 GENERALIZED PATH COMPLEX

**Path complex for undirected-graph** Firstly, we give the construction of path complex for undirected graphs. Secondly, we introduce the graph weak isomorphism and related mathematical properties. Finally, we states the weak isomorphism invariance of path complex homology for graphs.

**Definition 3.1** (Path). Given a simple undirected graph  $G = (V, E)$  over the verset set  $V$ , an  $n$ -path  $\sigma_n$  of  $G$  is defined as any sequence of  $n + 1$  vertices  $v_0 v_1 \cdots v_n (v_i \in V)$  such that every two vertices are distinct and every two adjacent vertices form an edge.

Note that for each  $n$ -path  $\sigma_n = v_0 v_1 \cdots v_n$ ,  $\sigma'_n = v_n \cdots v_1 v_0$  is also an  $n$ -path, we identify these two paths as the same one. For an  $n$ -path  $\sigma_n = v_0 \cdots v_n$ , the  $(n - 1)$ -paths by removing the first or last vertex, denoted by  $\partial^L_{\sigma_n}$  and  $\partial^R_{\sigma_n}$  respectively, are called the faces of  $\sigma_n$ . Two  $n$ -paths are neighbors if they are faces of a common  $(n + 1)$ -path. Let  $\mathcal{N}(\sigma_n)$  be the set of neighbors of  $\sigma_n$ .

**Definition 3.2** (Path complex from undirected graphs). Given a simple undirected graph  $G = (V, E)$ , all paths of  $G$  form a path complex  $P_G$ . We call  $P_G$  the path complex derived from  $G$ .

**Theorem 3.3** (Path Complex Invariance). *The graph neural network is invariant to the permutation of the simplexes in the path complex  $P_G$ , meaning the output of the network remains unchanged under any permutation  $\pi$  of the vertices.*

**Definition 3.4** (Graph collapse and expansion). Given a graph  $G = (V, E)$ , take an edge  $(v_1, v_2) \in E$  such that  $\deg(v_1) = 1$ . Let  $V' = V \setminus \{v_1\}$ ,  $E' = E \setminus \{(v_1, v_2)\}$ , then  $G' = (V', E')$  is a new graph. We say that  $G'$  is derived from  $G$  by a graph collapse and  $G$  is derived from  $G'$  by a graph expansion.

**Definition 3.5** (Weak isomorphism). Given two graphs  $G_1, G_2$ ,  $G_1$  and  $G_2$  are called weak isomorphic if  $G_1$  can be derive from  $G_2$  by a sequence of graph collapse and expansion operations.

It can be seen that two graphs are weak isomorphic if they are isomorphic.

**Theorem 3.6.** *If two graphs  $G_1$  and  $G_2$  are weak isomorphic, then, 1) The number of connected components of  $G_1$  and  $G_2$  are same; 2) The number of cycles of  $G_1$  and  $G_2$  are same.*

**Theorem 3.7.** *Given two graphs  $G_1, G_2$ , let  $P_{G_1}, P_{G_2}$  be the path complexes derived from  $G_1$  and  $G_2$  respectively. If  $G_1$  and  $G_2$  are weak isomorphic, then*

$$H_k(P_{G_1}) \cong H_k(P_{G_2}) \quad (k \geq 0)$$

Theorem 3.7 means the path complex homology is a graph weak isomorphism invariant. Consequently, for two graphs  $G_1$  and  $G_2$ , if there exists  $k$  such that  $H_k(P_{G_1}) \not\cong H_k(P_{G_2})$ , then  $G_1$  and  $G_2$  are not weak isomorphic and not isomorphic.

The profound theoretical relationship between the Weisfeiler-Lehman (WL) graph isomorphism test and message-passing graph neural networks (GNNs) has been extensively documented Xu et al. (2018).

**Definition 3.8** (PWL). The steps of general PWL are as follows:

1. Given a path complex  $P$ , all the paths of  $P$  are initialized with the same color.
2. For the color  $c_\sigma^t$  of path  $\sigma$  at iteration  $t$ , the color  $c_\sigma^{t+1}$  of  $\sigma$  at the next iteration is computed by perfectly hashing the color multi-set of the neighbors of  $\sigma$ .
3. The algorithm stops once a stable coloring is reached. Two path complexes are considered non-isomorphic if their color histograms are different at some dimensions.

**Theorem 3.9.** *PWL is strictly more powerful than WL.*

Figure 8 shows two graphs that cannot be distinguished by the WL test, but their derived path complexes can be distinguished by PWL.

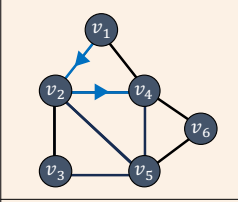
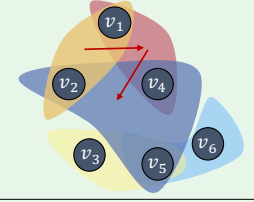
Path complex			Path types in Simplicial complex		
	Graph path complex	Hypergraph path complex	Vertex-path	Edge-path	Triangle-path
	{0-path}	$\{v_1\}$	$\{v_1 v_2\}$	$\{v_1\}$	$\{v_1 v_2\}$
{1-path}	$\{v_1 - v_2\}$	$\{v_1 v_2 - v_1 v_4\}$	$\{v_1 - v_2\}$	$\{v_1 v_2 - v_2 v_4\}$	$\{v_1 v_2 v_4 - v_2 v_4 v_5\}$
{2-path}	$\{v_1 - v_2 - v_4\}$	$\{v_1 v_2 - v_1 v_4 - v_2 v_4 v_5\}$	$\{v_1 - v_2 - v_4\}$	$\{v_1 v_2 - v_2 v_4 - v_4 v_5\}$	$\{v_1 v_2 v_4 - v_2 v_4 v_5 - v_2 v_3 v_5\}$

Figure 3: The path complexes of graphs, simplicial complexes, and hypergraphs. The table lists the 0-paths, 1-paths, and 2-paths, where the red arrows indicate the selected 2-paths. Specifically, for simplicial complexes, we enumerate the path complexes from vertexes (0-simplices), edges (1-simplices) and triangles (2-simplices), respectively.

**Path Complex for simplicial complex and hypergraph** Generally speaking, a path complex is a set of paths that is closed under the removing of the first or last vertices of each path. So we can also construct path complex from simplicial complex and hypergraph by defining simplex-paths and

hyperedge-paths. Various kinds of paths can be defined by considering the lower adjacent, upper adjacent, face and coface relations among simplices and hyperedges. Figure 3 shows examples of path complexes constructed from graph, simplicial complex and hypergraph. Details can be found in Appendix B.1.

### 3.2 MOLECULAR PATH COMPLEX REPRESENTATION AND PATH FEATURES

**Molecular Path Complex Representation** Currently, covalent-bond molecular graphs serve as the standard for molecular topological representations. These graphs underpin the molecular force fields used in molecular dynamics simulations, incorporating terms for both covalent bonds—such as bond lengths, angles, and dihedral angles—and non-covalent interactions like electrostatic and van der Waals forces. To enhance molecular representations with comprehensive force field data, we introduce the molecular path complex. This model utilizes path simplices across different dimensions to distinctly represent both covalent and non-covalent bond terms. As depicted in Figure 5 (in Appendix A.1), the  $C_2H_6O$  molecule is illustrated alongside its corresponding path simplices. Specifically, our 1-path simplex captures bond lengths, the 2-path simplex details bond angles, and the 3-path simplex reflects dihedral angles.

**Path Features** Our path (simplex) features are meticulously designed to encapsulate the various atomic properties and interactions detailed in molecular dynamics (MD) force fields. Specifically, our 0-path features—comprising atomic number, radius, and electronegativity—are derived using Rdkit, akin to the approach in CGCNN Xie & Grossman (2018). Table 7 (in Appendix A.1) presents a comprehensive listing of our 1-path, 2-path, and 3-path features. Importantly, our model employs detailed local geometric properties of the path complex as path features. This method allows us to explicitly learn covalent bond terms defined in the MD force fields, while also implicitly capturing non-bond interactions.

### 3.3 MOLECULAR PCMP MODEL

Path Complex Message Passing (PCMP) introduces a novel method for message passing in graphs by leveraging path complexes, which are composed of paths of varying lengths. In contrast to traditional graph neural networks (GNNs) that primarily aggregate information from local neighbors, PCMP prioritizes message propagation along higher-order path complexes. By incorporating longer paths, PCMP effectively captures long-range dependencies within the graph, enhancing its ability to model complex relationships.

A key feature of PCMP is the hierarchical message passing mechanism between different orders of path complexes. First, messages in higher-order paths are updated, reflecting the broader structure of the graph. These updated messages are then propagated to lower-order paths, ensuring that global information from longer paths informs to lower-order paths. After this, a feedback mechanism is employed, where updated messages from lower-order paths influence the higher-order paths, thus refining the representation at all levels. This bidirectional interaction between higher- and lower-order path allows PCMP to seamlessly integrate both global and local information effectively. The Path Message Passing Module can be found in Appendix A.2.

**Theorem 3.10.** *A Path Complex Message Passing (PCMP) with sufficient layers and injective neighborhood aggregators achieves the same expressive power as the PWL.*

## 4 EXPERIMENTS

### 4.1 BENCHMARK DATASETS AND MODELS

To thoroughly validate our PCMP model, we use three widely recognized benchmark datasets from MoleculeNet Wu et al. (2018) and MolBench Jiang et al. (2023). During data preprocessing, we employ the Merck molecular force field (MMFF94) function from RDKit to generate 3D molecular structures. The datasets are split into training, validation, and test sets using the scaffold splitting method, with a ratio of 8:1:1. Detailed descriptions of the datasets, preprocessing steps, and splitting method are provided in Appendix A.3.

We compare the performance of our PCMP model against state-of-the-art GNN models, both with and without pre-training. The non-pre-trained GNN models include (1) widely-used architectures such as GIN Xu et al. (2018), GAT Velickovic et al. (2017), and GCN Kipf & Welling (2016a); (2) recent models incorporating 3D molecular geometry, including SGCN Danel et al. (2020), DimeNet Gasteiger et al. (2020), and HMGNN Shui & Karypis (2020b); and (3) architectures specifically designed for molecular representation, such as D-MPNN Yang et al. (2019), AttentiveFP Xiong et al. (2019), and Mol-GDL Shen et al. (2023). For pre-trained models, we compare against N-Gram Liu et al. (2019), PretrainGNN Hu et al. (2019), GROVER Rong et al. (2020), GEM Fang et al. (2022), DMP Zhu et al. (2023), SMPT Li et al. (2024) and DGCL Jiang et al. (2024).

Table 1: Comparison with GNN architectures. The best performance is indicated as **bold**, and the subindex indicates standard deviation values. \* indicates that the result is not available for the model.

	Method	QM7	QM9	Tox21	HIV	MUV
GNN	GIN	110.3 <sub>(7.2)</sub>	0.00886 <sub>(0.00005)</sub>	0.740 <sub>(0.008)</sub>	0.753 <sub>(0.019)</sub>	0.718 <sub>(0.003)</sub>
	GAT	103.0 <sub>(4.4)</sub>	0.01117 <sub>(0.00018)</sub>	0.745 <sub>(0.006)</sub>	0.724 <sub>(0.008)</sub>	0.671 <sub>(0.011)</sub>
	GCN	100.0 <sub>(3.8)</sub>	0.00923 <sub>(0.00019)</sub>	0.709 <sub>(0.003)</sub>	0.740 <sub>(0.003)</sub>	0.716 <sub>(0.004)</sub>
	D-MPNN	103.5 <sub>(8.6)</sub>	0.00812 <sub>(0.00009)</sub>	0.759 <sub>(0.007)</sub>	0.771 <sub>(0.005)</sub>	0.786 <sub>(0.014)</sub>
	Attentive FP	72.0 <sub>(2.7)</sub>	0.00812 <sub>(0.00001)</sub>	0.761 <sub>(0.005)</sub>	0.757 <sub>(0.014)</sub>	0.766 <sub>(0.015)</sub>
	GTransformer	161.3 <sub>(7.1)</sub>	0.00923 <sub>(0.00019)</sub>	*	*	*
	SGCN	131.3 <sub>(11.6)</sub>	0.01459 <sub>(0.00055)</sub>	*	*	*
	DimNet	95.6 <sub>(4.1)</sub>	0.01031 <sub>(0.00076)</sub>	*	*	*
	HMGNN	101.6 <sub>(3.2)</sub>	0.01239 <sub>(0.00001)</sub>	*	*	*
	Mol-GDL	62.2 <sub>(0.4)</sub>	0.00952 <sub>(0.00013)</sub>	0.791 <sub>(0.005)</sub>	0.808 <sub>(0.007)</sub>	0.675 <sub>(0.014)</sub>
	Pretrain_GNN	N-Gram <sub>RF</sub>	92.8 <sub>(4.0)</sub>	0.01037 <sub>(0.00016)</sub>	0.743 <sub>(0.004)</sub>	0.772 <sub>(0.001)</sub>
N-Gram <sub>XGB</sub>		81.9 <sub>(1.9)</sub>	0.00964 <sub>(0.00031)</sub>	0.758 <sub>(0.009)</sub>	0.787 <sub>(0.004)</sub>	0.748 <sub>(0.002)</sub>
PretrainGNN		113.2 <sub>(0.6)</sub>	0.00922 <sub>(0.00004)</sub>	0.781 <sub>(0.006)</sub>	0.799 <sub>(0.007)</sub>	0.813 <sub>(0.021)</sub>
GROVER <sub>base</sub>		94.5 <sub>(3.8)</sub>	0.00986 <sub>(0.00055)</sub>	0.743 <sub>(0.001)</sub>	0.625 <sub>(0.009)</sub>	0.673 <sub>(0.018)</sub>
GROVER <sub>large</sub>		92.0 <sub>(0.9)</sub>	0.00986 <sub>(0.00025)</sub>	0.735 <sub>(0.001)</sub>	0.682 <sub>(0.011)</sub>	0.673 <sub>(0.018)</sub>
MolCLR		66.8 <sub>(2.3)</sub>	*	0.750 <sub>(0.002)</sub>	0.781 <sub>(0.005)</sub>	0.796 <sub>(0.019)</sub>
GEM		58.9 <sub>(0.8)</sub>	0.00746 <sub>(0.00001)</sub>	0.781 <sub>(0.001)</sub>	0.806 <sub>(0.009)</sub>	0.817 <sub>(0.005)</sub>
DMP		74.4 <sub>(1.2)</sub>	*	0.791 <sub>(0.004)</sub>	0.814 <sub>(0.004)</sub>	*
SMPT		*	*	0.797 <sub>(0.001)</sub>	0.812 <sub>(0.001)</sub>	0.822 <sub>(0.008)</sub>
DGCL		100.9 <sub>(2.5)</sub>	*	0.772 <sub>(0.310)</sub>	0.815 <sub>(1.100)</sub>	*
<b>PCMP</b>		<b>53.6</b> <sub>(2.1)</sub>	<b>0.00683</b> <sub>(0.00005)</sub>	<b>0.801</b> <sub>(0.002)</sub>	<b>0.823</b> <sub>(0.004)</sub>	<b>0.827</b> <sub>(0.015)</sub>

## 4.2 RESULTS

The comparison of our PCMP model with existing models on benchmark datasets is presented in Table 1. Detailed parameter settings for PCMP can be found in Section A.4 (Appendix A). Our PCMP model shows a significant performance advantage across datasets, mainly due to its advanced feature extraction capabilities and superior recognition of complex molecular structures. The message-passing mechanism in PCMP is organized into two distinct layers: the upper embedding, which considers upper adjacent neighbors, and the lower embedding, which incorporates both upper adjacent and face neighbors. This dual-layered approach integrates path information from multiple perspectives, enhancing the model’s ability to capture both local and global graph structures. Each path is updated not only based on the features of its constituent nodes but also by incorporating information from both higher-order and lower-order connected paths. This sophisticated mechanism enables the model to detect subtle structural variations within molecules that are often difficult to distinguish. Compared to models that rely heavily on traditional pre-training, PCMP reduces computational demands by bypassing extensive pre-training phases while achieving excellent results, making it a promising solution for molecular property prediction.

## 4.3 ABLATION STUDY OF PCMP

**Impact of Message Passing Mechanisms** In the PCMP model, high-order features are first updated and then used as inputs to update low-order features, with the updated low-order features subsequently used to update high-order features. To evaluate the significance of this interaction, we introduced three variant models: PCMP-PARAL, PCMP-HL, and PCMP-LH, each limiting the

information exchange between different feature levels. Specifically, PCMP-PARAL restricts both high-to-low and low-to-high feature updates; PCMP-HL limits message passing from high-order to low-order features; and PCMP-LH restricts message passing from low-order to high-order features. Table 2 compares the performance of these models—PCMP-PARAL, PCMP-HL, PCMP-LH—against the standard PCMP on benchmark datasets.

Table 2: Results on benchmark datasets with different message passing mechanisms.

Method	QM7	QM9	Tox21	HIV	MUV
PCMP-PARAL	56.9 <sub>(1.5)</sub>	0.00751 <sub>(0.00015)</sub>	0.779 <sub>(0.008)</sub>	0.794 <sub>(0.016)</sub>	0.808 <sub>(0.006)</sub>
PCMP-HL	54.8 <sub>(1.0)</sub>	0.00727 <sub>(0.00006)</sub>	0.796 <sub>(0.002)</sub>	0.803 <sub>(0.020)</sub>	0.814 <sub>(0.012)</sub>
PCMP-LH	55.3 <sub>(1.9)</sub>	0.00764 <sub>(0.00005)</sub>	0.793 <sub>(0.004)</sub>	0.793 <sub>(0.009)</sub>	0.806 <sub>(0.004)</sub>
PCMP	<b>53.6</b> <sub>(2.1)</sub>	<b>0.00683</b> <sub>(0.00005)</sub>	<b>0.801</b> <sub>(0.004)</sub>	<b>0.823</b> <sub>(0.004)</sub>	<b>0.827</b> <sub>(0.015)</sub>

**Impact of Input Path** The PCMP model integrates various path-based features to improve its ability to accurately capture molecular structures. These paths, ranging from 0-path to 3-path, represent different levels of molecular interaction complexity, from the simplest to the most intricate. Table 3 shows the performance results for different path inputs on the benchmark datasets. As indicated, incorporating higher-order paths (2-path and 3-path) generally enhances the model’s performance, with the inclusion of the 3-path yielding the best results across all datasets. This underscores its effectiveness in capturing complex molecular interactions. However, the slight performance degradation when excluding the 2-path and 3-path elements suggests that lower-order information remains crucial, especially for the QM7 dataset, where simpler molecular representations are sufficient.

Table 3: The results for input different path of the benchmark datasets.

Input-Path	QM7	QM9	Tox21	HIV	MUV
{0,1}-path	57.0 <sub>(1.4)</sub>	0.00898 <sub>(0.00012)</sub>	0.786 <sub>(0.006)</sub>	0.782 <sub>(0.007)</sub>	0.767 <sub>(0.003)</sub>
{0,1,2}-path	56.9 <sub>(1.1)</sub>	0.00700 <sub>(0.00006)</sub>	0.792 <sub>(0.002)</sub>	0.803 <sub>(0.005)</sub>	0.815 <sub>(0.012)</sub>
{0,1,2,3}-path	<b>53.6</b> <sub>(2.1)</sub>	<b>0.00683</b> <sub>(0.00005)</sub>	<b>0.801</b> <sub>(0.002)</sub>	<b>0.823</b> <sub>(0.004)</sub>	<b>0.827</b> <sub>(0.015)</sub>

**Impact of Readout Path** To investigate whether feature outputs at different levels can improve model performance, we designed several output strategies. As shown in Table 4, utilizing outputs from multiple levels allows the PCMP model to capture both low-order and high-order molecular features, significantly enhancing the model’s ability to represent complex structures and improving its overall expressiveness.

Table 4: The results for Readout different path of the benchmark datasets.

Output-Path	QM7	QM9	Tox21	HIV	MUV
{0}-path	56.8 <sub>(1.2)</sub>	0.00797 <sub>(0.00004)</sub>	0.789 <sub>(0.008)</sub>	0.784 <sub>(0.010)</sub>	0.806 <sub>(0.018)</sub>
{0,1}-path	56.4 <sub>(1.5)</sub>	0.00698 <sub>(0.00010)</sub>	0.793 <sub>(0.009)</sub>	0.800 <sub>(0.007)</sub>	0.808 <sub>(0.007)</sub>
{0,1,2}-path	<b>53.6</b> <sub>(2.1)</sub>	<b>0.00683</b> <sub>(0.00005)</sub>	<b>0.801</b> <sub>(0.002)</sub>	<b>0.823</b> <sub>(0.004)</sub>	<b>0.827</b> <sub>(0.015)</sub>

**Visualization of the High-order Path Contribute** To clearly quantify the impact of different high-order paths in the PCMP model on both message passing and Readout processes, we employ various order paths in regression tasks (QM7 and QM9 datasets) and classification tasks (Tox21, HIV, and MUV datasets). Figure 4 displays the performance improvement. It is important to note that the PCMP model utilizes face and coffee neighbors for information dissemination and update processes. Therefore, the paths we have selected to analyze are the 0,1-order path, 0,1,2-order path, and 0,1,2,3-order path. This delineation helps in understanding how path complexity influences the effectiveness of the model across different types of tasks and datasets.

**Impact of Geometric Feature in Path Complex** Since the path complex uses different geometric features for different order paths in addition to the bond angle and dihedral angle in MD in characterizing the characteristics of molecules, the effectiveness of these added geometric features is verified



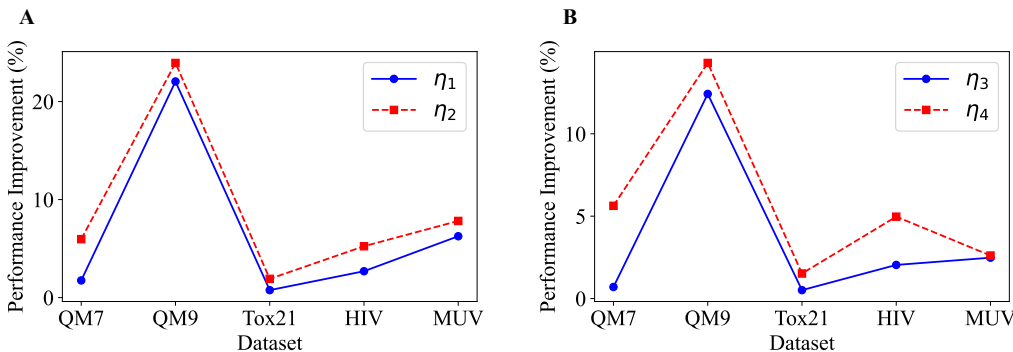


Figure 4: Visualization of High-order Path Contributions Across Datasets. A: Input high-order paths showing model performance improvement, where  $\eta_1$  and  $\eta_2$  represent the improvements with input 0,1,2-order paths and 0,1,2,3-order 0,1,2-order paths relative to input 0,1-order paths. B: Output high-order paths and model performance improvement, where  $\eta_3$  and  $\eta_4$  represent the improvements with output paths 0,1-order paths and 0,1,2-order paths relative to output 0-order path.

in Table 5. By comparison, we found that after using these geometric features, QM7 increased by 4.97%, QM9 increased by 5.92%, Tox21 increased by 3.49%, HIV increased by 2.88%, and MUV increased by 6.71%.

Table 5: Comparison of features. Angle and Dihedral represent that 2-path and 3-path only use angle and Dihedral, and Path feature represents the encoding method in Table7.

	Method	QM7	QM9	Tox21	HIV	MUV
PCMP	Angle and Dihedral	56.4 <sub>(1.7)</sub>	0.00726 <sub>(0.00012)</sub>	0.774 <sub>(0.009)</sub>	0.800 <sub>(0.006)</sub>	0.775 <sub>(0.023)</sub>
	Path Features	<b>53.6</b> <sub>(2.1)</sub>	<b>0.00683</b> <sub>(0.00005)</sub>	<b>0.801</b> <sub>(0.002)</sub>	<b>0.823</b> <sub>(0.004)</sub>	<b>0.827</b> <sub>(0.015)</sub>

**Sensitivity of Hyperparameters** We explored the model’s sensitivity to hyperparameters and the experimental results are displayed in Table 6. According to the results, the model’s performance metrics are generally stable across different hyperparameters settings.

Table 6: Sensitivity of hyperparameters for benchmark datasets.

Hyperparameters	QM7	QM9	Tox21	HIV	MUV	
Head	1	<b>53.6</b> <sub>(2.1)</sub>	0.00931 <sub>(0.00007)</sub>	0.748 <sub>(0.006)</sub>	0.793 <sub>(0.004)</sub>	<b>0.827</b> <sub>(0.015)</sub>
	2	54.7 <sub>(1.8)</sub>	0.00747 <sub>(0.00006)</sub>	0.754 <sub>(0.007)</sub>	<b>0.823</b> <sub>(0.004)</sub>	0.787 <sub>(0.019)</sub>
	4	55.9 <sub>(1.0)</sub>	0.00721 <sub>(0.00006)</sub>	0.759 <sub>(0.005)</sub>	0.799 <sub>(0.013)</sub>	0.801 <sub>(0.017)</sub>
	6	58.3 <sub>(2.5)</sub>	<b>0.00683</b> <sub>(0.00005)</sub>	<b>0.801</b> <sub>(0.004)</sub>	0.807 <sub>(0.016)</sub>	0.814 <sub>(0.013)</sub>
	8	58.5 <sub>(1.2)</sub>	0.00826 <sub>(0.00008)</sub>	0.765 <sub>(0.004)</sub>	0.808 <sub>(0.004)</sub>	0.783 <sub>(0.008)</sub>
Batch Size	64	59.6 <sub>(1.7)</sub>	0.00868 <sub>(0.00012)</sub>	0.784 <sub>(0.004)</sub>	0.796 <sub>(0.003)</sub>	0.785 <sub>(0.007)</sub>
	128	57.5 <sub>(1.0)</sub>	0.01040 <sub>(0.00007)</sub>	0.778 <sub>(0.003)</sub>	0.803 <sub>(0.007)</sub>	0.819 <sub>(0.016)</sub>
	256	54.3 <sub>(0.9)</sub>	0.00756 <sub>(0.00014)</sub>	0.778 <sub>(0.003)</sub>	0.807 <sub>(0.015)</sub>	0.789 <sub>(0.022)</sub>
	512	<b>53.6</b> <sub>(2.1)</sub>	<b>0.00683</b> <sub>(0.00004)</sub>	<b>0.801</b> <sub>(0.004)</sub>	<b>0.823</b> <sub>(0.004)</sub>	<b>0.827</b> <sub>(0.015)</sub>
LR	5e-3	58.5 <sub>(1.2)</sub>	0.00784 <sub>(0.00004)</sub>	0.782 <sub>(0.004)</sub>	0.805 <sub>(0.009)</sub>	0.808 <sub>(0.014)</sub>
	1e-3	56.2 <sub>(1.4)</sub>	<b>0.00683</b> <sub>(0.00005)</sub>	<b>0.791</b> <sub>(0.009)</sub>	0.811 <sub>(0.013)</sub>	<b>0.827</b> <sub>(0.015)</sub>
	5e-4	54.7 <sub>(1.3)</sub>	0.00880 <sub>(0.00006)</sub>	0.781 <sub>(0.012)</sub>	<b>0.823</b> <sub>(0.005)</sub>	0.814 <sub>(0.012)</sub>
	1e-4	<b>53.6</b> <sub>(2.1)</sub>	0.00962 <sub>(0.00012)</sub>	0.789 <sub>(0.006)</sub>	0.799 <sub>(0.018)</sub>	0.822 <sub>(0.013)</sub>
	5e-5	64.9 <sub>(3.1)</sub>	0.00784 <sub>(0.00011)</sub>	0.724 <sub>(0.007)</sub>	0.788 <sub>(0.024)</sub>	0.818 <sub>(0.019)</sub>

486 5 CONCLUSION  
487

488 In this study, we introduced the path complex message passing, a novel model for molecular struc-  
489 ture representation based on path complexes, designed to predict molecular properties. By inte-  
490 grating force fields with path complexes, the model enhances our understanding of the relationship  
491 between molecular structure and function, offering valuable insights for both theoretical research  
492 and practical applications in molecular design and materials science. The PCMP model employs  
493 0-paths for atomic properties, 1-paths for pairwise interactions, 2-paths for bond angle terms, and  
494 3-paths for dihedral angle information. These paths are used to compute attention scores, enabling  
495 efficient message propagation and feature integration across various levels of molecular information.  
496 Validation on five benchmark datasets has demonstrated the PCMP’s superior predictive capabilities.  
497 Ablation studies further confirm that incorporating higher-order features significantly improves per-  
498 formance, pointing to promising directions for future research in molecular simulation and design.

499 REFERENCES  
500

- 501 Kenneth Atz, Francesca Grisoni, and Gisbert Schneider. Geometric deep learning on molecular  
502 representations. *Nature Machine Intelligence*, 3(12):1023–1032, 2021.
- 503 Song Bai, Feihu Zhang, and Philip HS Torr. Hypergraph convolution and hypergraph attention.  
504 *Pattern Recognition*, 110:107637, 2021.
- 505 Lorenz C Blum and Jean-Louis Reymond. 970 million druglike small molecules for virtual screening  
506 in the chemical universe database gdb-13. *Journal of the American Chemical Society*, 131(25):  
507 8732–8733, 2009.
- 508 Cristian Bodnar. *Topological Deep Learning: Graphs, Complexes, Sheaves*. PhD thesis, Apollo  
509 - University of Cambridge Repository, 2022. URL <https://www.repository.cam.ac.uk/handle/1810/350982>.
- 510 Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Lio, Guido F Montufar, and  
511 Michael Bronstein. Weisfeiler and lehman go cellular: Cw networks. *Advances in neural infor-*  
512 *mation processing systems*, 34:2625–2640, 2021a.
- 513 Cristian Bodnar, Fabrizio Frasca, Yuguang Wang, Nina Otter, Guido F Montufar, Pietro Lio, and  
514 Michael Bronstein. Weisfeiler and lehman go topological: Message passing simplicial networks.  
515 In *International Conference on Machine Learning*, pp. 1026–1037. PMLR, 2021b.
- 516 Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geomet-  
517 ric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42,  
518 2017.
- 519 Zixuan Cang and Guo-Wei Wei. Topologynet: Topology based deep convolutional and multi-  
520 task neural networks for biomolecular property predictions. *PLoS computational biology*, 13  
521 (7):e1005690, 2017.
- 522 HC Stephen Chan, Hanbin Shan, Thamani Dahoun, Horst Vogel, and Shuguang Yuan. Advancing  
523 drug discovery via artificial intelligence. *Trends in pharmacological sciences*, 40(8):592–604,  
524 2019.
- 525 Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Graph networks as a universal  
526 machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9):3564–  
527 3572, 2019.
- 528 Dong Chen, Jian Liu, Jie Wu, Guo-Wei Wei, Feng Pan, and Shing-Tung Yau. Path topology in  
529 molecular and materials sciences. *The journal of physical chemistry letters*, 14(4):954–964, 2023.
- 530 Hongming Chen, Ola Engkvist, Yinhai Wang, Marcus Olivecrona, and Thomas Blaschke. The rise  
531 of deep learning in drug discovery. *Drug discovery today*, 23(6):1241–1250, 2018.
- 532 Jiahui Chen, Yuchi Qiu, Rui Wang, and Guo-Wei Wei. Persistent laplacian projected omicron ba. 4  
533 and ba. 5 to become new dominating variants. *Computers in biology and medicine*, 151:106262,  
534 2022.

- 540 Kamal Choudhary and Brian DeCost. Atomistic line graph neural network for improved materials  
541 property predictions. *npj Computational Materials*, 7(1):185, 2021.
- 542
- 543 Kamal Choudhary, Brian DeCost, and Francesca Tavazza. Machine learning with force-field-  
544 inspired descriptors for materials: Fast screening and mapping energy landscape. *Physical review*  
545 *materials*, 2(8):083801, 2018.
- 546 Samir Chowdhury and Facundo Mémoli. Persistent path homology of directed networks. In *Pro-*  
547 *ceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1152–  
548 1169. SIAM, 2018.
- 549
- 550 Samir Chowdhury, Steve Huntsman, and Matvey Yutin. Path homologies of motifs and temporal  
551 network representations. *Applied Network Science*, 7(1):4, 2022.
- 552
- 553 Tomasz Danel, Przemysław Spurek, Jacek Tabor, Marek Śmieja, Łukasz Struski, Agnieszka Słowik,  
554 and Łukasz Maziarka. Spatial graph convolutional networks. In *International Conference on*  
555 *Neural Information Processing*, pp. 668–675. Springer, 2020.
- 556 Xiaomin Fang, Lihang Liu, Jieqiong Lei, Donglong He, Shanzhuo Zhang, Jingbo Zhou, Fan Wang,  
557 Hua Wu, and Haifeng Wang. Geometry-enhanced molecular representation learning for property  
558 prediction. *Nature Machine Intelligence*, 4(2):127–134, 2022.
- 559
- 560 Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks.  
561 In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 3558–3565, 2019.
- 562 Daniel Flam-Shepherd, Tony C Wu, Pascal Friederich, and Alan Aspuru-Guzik. Neural message  
563 passing on high order paths. *Machine Learning: Science and Technology*, 2(4):045009, 2021.
- 564
- 565 Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molec-  
566 ular graphs. *arXiv preprint arXiv:2003.03123*, 2020.
- 567 Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph  
568 neural networks for molecules. *Advances in Neural Information Processing Systems*, 34:6790–  
569 6802, 2021.
- 570
- 571 Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural  
572 message passing for quantum chemistry. In *International conference on machine learning*, pp.  
573 1263–1272. PMLR, 2017.
- 574
- 575 Lorenzo Giusti, Claudio Battiloro, Lucia Testa, Paolo Di Lorenzo, Stefania Sardellitti, and Sergio  
576 Barbarossa. Cell attention networks. In *2023 International Joint Conference on Neural Networks*  
577 *(IJCNN)*, pp. 1–8. IEEE, 2023.
- 578 Miguel A González. Force fields and molecular dynamics simulations. *École thématique de la*  
579 *Société Française de la Neutronique*, 12:169–200, 2011.
- 580
- 581 Alexander Grigor’yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Homologies of path com-  
582 plexes and digraphs. *arXiv preprint arXiv:1207.2834*, 2012.
- 583
- 584 Alexander Grigor’yan, Rolando Jimenez, Yuri Muranov, and Shing-Tung Yau. Homology of path  
585 complexes and hypergraphs. *Topology and its Applications*, 267:106877, 2019.
- 586
- 587 A.A. Grigor’yan, Y. Lin, Y.V. Muranov, et al. Path complexes and their homologies. *Journal of*  
588 *Mathematical Sciences*, 248:564–599, 2020. doi: 10.1007/s10958-020-04897-9. URL <https://doi.org/10.1007/s10958-020-04897-9>.
- 589
- 590 Alexander Grigor’yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Homotopy theory for di-  
591 graphs. *Pure and Applied Mathematics Quarterly*, 10(4):619–674, 2014.
- 592
- 593 Alexander Grigor’yan, Rolando Jimenez, Yuri Muranov, and Shing-Tung Yau. On the path homol-  
ogy theory of digraphs and eilenberg–steenrod axioms. *Homology, Homotopy and Applications*,  
20(2):179–205, 2018.

- 594 Alexander Grigor'yan, Yong Lin, and Shing-Tung Yau. Analytic and reidemeister torsions of di-  
595 graphs and path complexes. *Pure and Applied Mathematics Quarterly*, 20(2):703–755, 2024.
- 596  
597 Mustafa Hajij, Kyle Istvan, and Ghada Zamzmi. Cell complex neural networks. In *TDA \&  
598 Beyond*, 2020.
- 599 Mustafa Hajij, Ghada Zamzmi, Theodore Papamarkou, Nina Miolane, Aldo Guzmán-Sáenz,  
600 Karthikeyan Natesan Ramamurthy, Tolga Birdal, Tamal K Dey, Soham Mukherjee, Shreyas N  
601 Samaga, et al. Topological deep learning: Going beyond graph data. *arXiv e-prints*, pp. arXiv-  
602 2206, 2022.
- 603  
604 Thomas A Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and per-  
605 formance of mmff94. *Journal of computational chemistry*, 17(5-6):490–519, 1996.
- 606 Jakob Hansen and Robert Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied  
607 and Computational Topology*, 3(4):315–358, 2019.
- 608  
609 Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure  
610 Leskovec. Strategies for pre-training graph neural networks. In *International Conference on  
611 Learning Representations*, 2019.
- 612 Kexin Huang, Tianfan Fu, Lucas Glass, Marinka Zitnik, Cao Xiao, and Jimeng Sun. Deeppurpose:  
613 a deep learning library for drug–target interaction prediction. *Bioinformatics*, 36:5545 – 5547,  
614 2020. URL <https://api.semanticscholar.org/CorpusID:220496219>.
- 615  
616 John B Ingraham, Max Baranov, Zak Costello, Karl W Barber, Wujie Wang, Ahmed Ismail, Vincent  
617 Frappier, Dana M Lord, Christopher Ng-Thow-Hing, Erik R Van Vlack, et al. Illuminating protein  
618 space with a programmable generative model. *Nature*, 623(7989):1070–1078, 2023.
- 619 Arian Jamasb, Ramon Viñas Torné, Eric Ma, Yuanqi Du, Charles Harris, Kexin Huang, Dominic  
620 Hall, Pietro Lió, and Tom Blundell. Graphein—a python library for geometric deep learning and  
621 network analysis on biomolecular structures and interaction networks. *Advances in Neural Infor-  
622 mation Processing Systems*, 35:27153–27167, 2022.
- 623 Xiuyu Jiang, Liqin Tan, Jianhuan Cen, and Qingsong Zou. Molbench: A benchmark of ai models  
624 for molecular property prediction. In *International Symposium on Benchmarking, Measuring and  
625 Optimization*, pp. 53–70. Springer, 2023.
- 626  
627 Xiuyu Jiang, Liqin Tan, and Qingsong Zou. Dgcl: dual-graph neural networks contrastive learning  
628 for molecular property prediction. *Briefings in Bioinformatics*, 25(6):bbae474, 2024.
- 629 Eun-Sol Kim, Woo Young Kang, Kyoung-Woon On, Yu-Jung Heo, and Byoung-Tak Zhang. Hyper-  
630 graph attention networks for multimodal learning. In *Proceedings of the IEEE/CVF conference  
631 on computer vision and pattern recognition*, pp. 14581–14590, 2020.
- 632  
633 Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional net-  
634 works. *arXiv preprint arXiv:1609.02907*, 2016a.
- 635 Thomas N Kipf and Max Welling. Variational graph auto-encoders. *Advances in neural information  
636 processing systems*, 2016b.
- 637  
638 Andrew R Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.
- 639 Shuangli Li, Jingbo Zhou, Tong Xu, Dejing Dou, and Hui Xiong. GeomGCL: geometric graph  
640 contrastive learning for molecular property prediction. In *Proceedings of the AAAI Conference  
641 on Artificial Intelligence*, volume 36, pp. 4541–4549, 2022.
- 642  
643 Yishui Li, Wei Wang, Jie Liu, and Chengkun Wu. Pre-training molecular representation model with  
644 spatial geometry for property prediction. *Computational Biology and Chemistry*, 109:108023,  
645 2024.
- 646 Jian Liu, Dong Chen, Feng Pan, and Jie Wu. Neighborhood path complex for the quantitative  
647 analysis of the structure and stability of carboranes. *Journal of Computational Biophysics and  
Chemistry*, 22(04):503–511, 2023.

- 648 Shengchao Liu, Mehmet F Demirel, and Yingyu Liang. N-gram graph: Simple unsupervised repre-  
649 sentation for graphs, with applications to molecules. *Advances in neural information processing*  
650 *systems*, 32, 2019.
- 651
- 652 Yi Liu, Limei Wang, Meng Liu, Yuchao Lin, Xuan Zhang, Bora Oztekin, and Shuiwang Ji. Spherical  
653 message passing for 3d molecular graphs. In *International Conference on Learning Representa-*  
654 *tions (ICLR)*, 2022.
- 655
- 656 Kit-Kay Mak and Mallikarjuna Rao Pichika. Artificial intelligence in drug development: present  
657 status and future prospects. *Drug discovery today*, 24(3):773–780, 2019.
- 658
- 659 Stephen L Mayo, Barry D Olafson, and William A Goddard. Dreiding: a generic force field for  
660 molecular simulations. *Journal of Physical chemistry*, 94(26):8897–8909, 1990.
- 661
- 662 Zhuoran Qiao, Anders S Christensen, Matthew Welborn, Frederick R Manby, Anima Anandku-  
663 mar, and Thomas F Miller III. Informing geometric deep learning with electronic interactions  
664 to accelerate quantum chemistry. *Proceedings of the National Academy of Sciences*, 119(31):  
e2205221119, 2022.
- 665
- 666 Bharath Ramsundar, Peter Eastman, Pat Walters, and Vijay Pande. *Deep learning for the life sci-*  
667 *ences: applying deep learning to genomics, microscopy, drug discovery, and more.* ” O’Reilly  
668 Media, Inc.”, 2019.
- 669
- 670 Patrick Reiser, Marlen Neubert, André Eberhard, Luca Torresi, Chen Zhou, Chen Shao, Houssam  
671 Metni, Clint van Hoesel, Henrik Schopmans, Timo Sommer, et al. Graph neural networks for  
672 materials science and chemistry. *Communications Materials*, 3(1):93, 2022.
- 673
- 674 T Mitchell Roddenberry, Nicholas Glaze, and Santiago Segarra. Principled simplicial neural net-  
675 works for trajectory prediction. In *International Conference on Machine Learning*, pp. 9020–  
9029. PMLR, 2021.
- 676
- 677 T Mitchell Roddenberry, Michael T Schaub, and Mustafa Hajj. Signal processing on cell com-  
678 plexes. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal*  
*Processing (ICASSP)*, pp. 8852–8856. IEEE, 2022.
- 679
- 680 Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang.  
681 Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information*  
682 *Processing Systems*, 33:12559–12571, 2020.
- 683
- 684 Lars Ruddigkeit, Ruud Van Deursen, Lorenz C Blum, and Jean-Louis Reymond. Enumeration of 166  
685 billion organic small molecules in the chemical universe database gdb-17. *Journal of chemical*  
*information and modeling*, 52(11):2864–2875, 2012.
- 686
- 687 Michael T Schaub, Jean-Baptiste Seby, Florian Frantzen, T Mitchell Roddenberry, Yu Zhu, and  
688 Santiago Segarra. Signal processing on simplicial complexes. In *Higher-Order Systems*, pp.  
689 301–328. Springer, 2022.
- 690
- 691 Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Saucedo Felix, Stefan Chmiela, Alexandre  
692 Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network  
693 for modeling quantum interactions. *Advances in neural information processing systems*, 30, 2017.
- 694
- 695 Kristof T Schütt, Huziel E Saucedo, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller.  
696 Schnet—a deep learning architecture for molecules and materials. *The Journal of Chemical*  
*Physics*, 148(24), 2018.
- 697
- 698 Cong Shen, Jiawei Luo, and Kelin Xia. Molecular geometric deep learning. *Cell Reports Methods*,  
699 3(11), 2023.
- 700
- 701 Hiroyuki Shindo and Yuji Matsumoto. Gated graph recursive neural networks for molecular property  
prediction. *ArXiv*, abs/1909.00259, 2019. URL [https://api.semanticscholar.org/](https://api.semanticscholar.org/CorpusID:202541698)  
CorpusID:202541698.

- 702 Zeren Shui and George Karypis. Heterogeneous molecular graph neural networks for predicting  
703 molecule properties. *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 492–  
704 500, 2020a. URL <https://api.semanticscholar.org/CorpusID:221971188>.  
705
- 706 Zeren Shui and George Karypis. Heterogeneous molecular graph neural networks for predicting  
707 molecule properties. In *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 492–  
708 500. IEEE, 2020b.
- 709 Freyr Sverrisson, Jean Feydy, Bruno E Correia, and Michael M Bronstein. Fast end-to-end learning  
710 on protein surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern  
711 Recognition*, pp. 15272–15281, 2021.
- 712 Jacob Townsend, Cassie Putman Micucci, John H Hymel, Vasileios Maroulas, and Konstantinos D  
713 Vogiatzis. Representation of molecular structures with persistent homology for machine learning  
714 applications in chemistry. *Nature communications*, 11(1):3230, 2020.
- 715
- 716 Raphael JL Townshend, Stephan Eismann, Andrew M Watkins, Ramya Rangan, Masha Karelina,  
717 Rhiju Das, and Ron O Dror. Geometric deep learning of rna structure. *Science*, 373(6558):  
718 1047–1051, 2021.
- 719 Quang Truong and Peter Chin. Weisfeiler and lehman go paths: Learning topological features via  
720 path complexes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp.  
721 15382–15391, 2024.
- 722
- 723 Oliver T Unke and Markus Meuwly. Physnet: A neural network for predicting energies, forces,  
724 dipole moments, and partial charges. *Journal of chemical theory and computation*, 15(6):3678–  
725 3693, 2019.
- 726 Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Ben-  
727 gio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.
- 728
- 729 Nikolaos N Vlassis, Ran Ma, and WaiChing Sun. Geometric deep learning for computational me-  
730 chanics part i: Anisotropic hyperelasticity. *Computer Methods in Applied Mechanics and Engi-  
731 neering*, 371:113299, 2020.
- 732 Limei Wang, Yi Liu, Yuchao Lin, Haoran Liu, and Shuiwang Ji. Comenet: Towards complete and  
733 efficient message passing for 3d molecular graphs. *Advances in Neural Information Processing  
734 Systems*, 35:650–664, 2022a.
- 735
- 736 Rui Wang and Guo-Wei Wei. Persistent path laplacian. *Foundations of data science (Springfield,  
737 Mo.)*, 5(1):26–55, 2023.
- 738 Yuyang Wang, Jianren Wang, Zhonglin Cao, and Amir Barati Farimani. Molecular contrastive  
739 learning of representations via graph neural networks. *Nature Machine Intelligence*, 4(3):279–  
740 287, 2022b.
- 741
- 742 Max Welling and Thomas N Kipf. Semi-supervised classification with graph convolutional net-  
743 works. In *J. International Conference on Learning Representations (ICLR 2017)*, 2016.
- 744
- 745 Oliver Wieder, Stefan Kohlbacher, Méline Kuenemann, Arthur Garon, Pierre Ducrot, Thomas Sei-  
746 del, and Thierry Langer. A compact review of molecular property prediction with graph neural  
747 networks. *Drug Discovery Today: Technologies*, 37:1–12, 2020.
- 748
- 749 Shuang Wu, Xiang Liu, Ang Dong, Claudia Gragnoli, Christopher Griffin, Jie Wu, Shing-Tung Yau,  
750 and Rongling Wu. The metabolomic physics of complex diseases. *Proceedings of the National  
751 Academy of Sciences*, 120(42):e2308496120, 2023.
- 752
- 753 Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S  
754 Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learn-  
755 ing. *Chemical science*, 9(2):513–530, 2018.
- 756
- 757 Kelin Xia and Guo-Wei Wei. Persistent homology analysis of protein structure, flexibility, and  
758 folding. *International journal for numerical methods in biomedical engineering*, 30(8):814–844,  
759 2014.

- 756 Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and  
757 interpretable prediction of material properties. *Physical review letters*, 120(14):145301, 2018.  
758
- 759 Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, Xiaozhe Wan, Xutong Li, Zhao-  
760 jun Li, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, et al. Pushing the boundaries of molecular  
761 representation for drug discovery with the graph attention mechanism. *Journal of medicinal  
762 chemistry*, 63(16):8749–8760, 2019.
- 763 Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural  
764 networks? *arXiv preprint arXiv:1810.00826*, 2018.  
765
- 766 Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, Angel Guzman-  
767 Perez, Timothy Hopper, Brian Kelley, Miriam Mathea, et al. Analyzing learned molecular repre-  
768 sentations for property prediction. *Journal of chemical information and modeling*, 59(8):3370–  
769 3388, 2019.
- 770 Jiaxuan You, Rex Ying, Xiang Ren, William Hamilton, and Jure Leskovec. GraphRNN: Gener-  
771 ating realistic graphs with deep auto-regressive models. In Jennifer Dy and Andreas Krause  
772 (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of  
773 *Proceedings of Machine Learning Research*, pp. 5708–5717. PMLR, 10–15 Jul 2018. URL  
774 <https://proceedings.mlr.press/v80/you18a.html>.
- 775 Zhaoning Yu and Hongyang Gao. Molecular graph representation learning via heterogeneous motif  
776 graph construction. *arXiv preprint arXiv:2202.00529*, 2022.  
777
- 778 Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. Graph trans-  
779 former networks. *Advances in neural information processing systems*, 32, 2019.
- 780 Lu Zhang, Jianjun Tan, Dan Han, and Hao Zhu. From machine learning to deep learning: progress  
781 in machine intelligence for rational drug discovery. *Drug discovery today*, 22(11):1680–1685,  
782 2017.
- 783 Jinhua Zhu, Yingce Xia, Lijun Wu, Shufang Xie, Wengang Zhou, Tao Qin, Houqiang Li, and Tie-  
784 Yan Liu. Dual-view molecular pre-training. In *Proceedings of the 29th ACM SIGKDD Conference  
785 on Knowledge Discovery and Data Mining*, pp. 3615–3627, 2023.  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809

## A APPENDIX / SUPPLEMENTAL MATERIAL

## A.1 INITIALIZATION FEATURES

Table 7: MD Encoder for Path Features

	Features Type	Description	Type	Size
1-Path (bond)	Bond Directionality	None, Beginwedge, Begindash, etc.	One-Hot	7
	Bond Type	Single, Double, Triple, or Aromatic.	One-Hot	4
	Bond Length	Numerical length of the bond.	Float	1
	In Ring	Indicates if the bond is part of a chemical ring.	One-Hot	2
1-Path (non-bond) cutoff=3	Atom charges	Atoms charges in Molecular ( $q_i, q_j, q_i \cdot q_j$ )	Float	3
	Distance between atoms	Distance between atoms ( $1/d_{ij}, 1/d_{ij}^6, 1/d_{ij}^{12}$ )	Float	3
2-Path	Centroid distance	Centroid position of the triangle formed by 2-path	Float	3
	Distance	Three bond lengths (two for covalent bond and one for non-covalent bond)	Float	3
	Area	Triangle area spanned by 2-path	Float	1
	Bond Angle	Bond angle for 2-path	Float	1
3-Path	Volume	Volume spanned by 3-path	Float	1
	Dihedral	Dihedral angle for 3-path	Float	1
	Total Area	Total Area of the corresponding four triangles	Float	1
	Bond Length	Non-covalent bond length ( $\{v_1v_3\}, \{v_2v_4\}, \{v_1v_4\}$ )	Float	3

As depicted in Figure 5, the  $C_2H_6O$  molecule is illustrated alongside its corresponding path simplices. Specifically, our 1-path simplex captures bond lengths, the 2-path simplex details bond angles, and the 3-path simplex reflects dihedral angles.

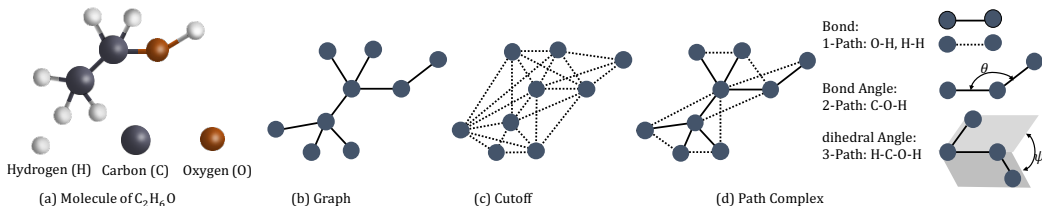


Figure 5: Different Representations of the  $C_2H_6O$  Molecule. (a) displays the molecular structure of  $C_2H_6O$ , including the oxygen (O), carbon (C), and hydrogen (H) atoms. (b) shows the graph representation based on chemical bonds. (c) illustrates the nearly fully connected graph generated based on a distance threshold (cutoff). (d) presents the representation using the path complex method and its physical implications. In the diagrams, solid lines represent chemical bonds, while dashed lines represent cutoff connections.

To fully incorporate MD force field information into molecular representation, we propose molecular path complex, which uses path simplices at different dimensions to explicitly characterize force field (covalent) bond terms. More specifically, our 1-path simplex represents bond length information, 2-path simplex describe bond angles, and 3-path simplex characterizes dihedral angle.



## A.2 PATH MESSAGE-PASSING MODULE

A central component of our PCMP model is path (simplex) Grigor'yan et al. (2024) message-passing module, where path features are updated based on path neighbors (same order paths), cofaces (higher-order paths), and faces (lower-order paths). Mathematically, each  $n$ -path will always have two unique  $(n - 1)$ -faces, but many  $n$ -path neighbors and  $(n + 1)$ -cofaces. In our PCMP framework, the simplex message-passing module contains two parts, i.e., message embedding and message updating. Two message embedding modules, i.e., upper embedding and lower embedding, are considered. In upper embedding module, path message will be generated using its neighbors and cofaces, while for lower embedding module, path messages will be generated using its neighbors and faces. The path feature will be updated using path messages from both upper and lower embedding through a message updating module. An illustration of our PCMP module is shown in Figure 6.

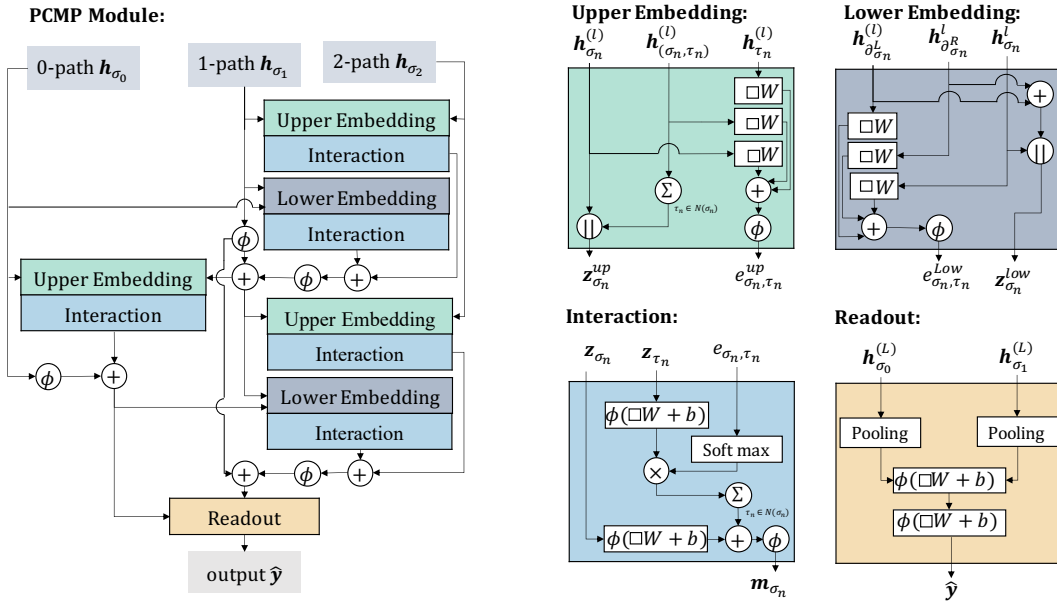


Figure 6: The PCMP Module.  $\square$  denotes the layer’s input,  $\parallel$  concatenation, and  $\phi$  a non-linearity. Upper embedding and Upper interaction refer to utilizing high-order path features to update low-order path features, while Lower embedding and Lower interaction refer to using low-order path features to update high-order path features.

The upper embedding module generates path message based on path neighbors and cofaces. For an  $n$ -path  $\sigma_n$  and its neighbors  $\tau_n$ , we denote their path feature vectors as  $\mathbf{h}_{\sigma_n}$  and  $\mathbf{h}_{\tau_n}$  respectively. The common coface of  $\sigma_n$  and  $\tau_n$  is denoted as  $(\tau_n, \sigma_n)$  and  $\mathbf{h}_{(\tau_n, \sigma_n)}$  the associate path feature. The upper attention score  $e_{\sigma_n, \tau_n}^{up}$  and upper concatenated feature  $\mathbf{z}_{\sigma_n}^{up}$  can be expressed as,

$$\begin{aligned} \mathbf{f}_{\sigma_n} &= \mathbf{h}_{\sigma_n} \mathbf{W}_n^{up}, \quad \mathbf{f}_{\tau_n} = \mathbf{h}_{\tau_n} \mathbf{W}_n^{up}, \quad \mathbf{f}_{(\sigma_n, \tau_n)} = \mathbf{h}_{(\sigma_n, \tau_n)} \mathbf{W}_{n+1}, \\ e_{\sigma_n, \tau_n}^{up} &= \text{ReLU}(\mathbf{f}_{\sigma_n} + \mathbf{f}_{\tau_n} + \mathbf{f}_{(\sigma_n, \tau_n)}), \quad \mathbf{z}_{\sigma_n}^{up} = [\mathbf{h}_{\sigma_n} \parallel \sum_{\tau_n \in \mathcal{N}(\sigma_n)} \frac{1}{|\mathcal{N}(\sigma_n)|} \mathbf{h}_{(\sigma_n, \tau_n)}], \end{aligned}$$

where ReLU is a non-linear activation function and  $\mathbf{W}_n^{up}$  and  $\mathbf{W}_{n+1}$  are weight matrices. Note that  $\parallel$  is the concatenation operator,  $\mathcal{N}(\sigma_n)$  denotes the neighbors of path  $\sigma_n$ , and  $|\mathcal{N}(\sigma_n)|$  is the total number of neighbors of path  $\sigma_n$ .

The lower embedding module generates path message based on path neighbors and faces. For an  $n$ -path  $\sigma_n$ , we use  $\partial_{\sigma_n}^R$  and  $\partial_{\sigma_n}^L$  to represent its right and left faces. The lower attention score  $e_{\sigma_n, \tau_n}^{low}$

and lower concatenated feature  $\mathbf{z}_{\sigma_n}^{low}$  can be expressed as,

$$\begin{aligned} \mathbf{f}_{\sigma_n} &= \mathbf{h}_{\sigma_n} \mathbf{W}_n^{low}, \quad \mathbf{f}_{\partial_{\sigma_n}^L} = \mathbf{h}_{\partial_{\sigma_n}^L} \mathbf{W}_{n-1}, \quad \mathbf{f}_{\partial_{\sigma_n}^R} = \mathbf{h}_{\partial_{\sigma_n}^R} \mathbf{W}_{n-1}, \\ e_{\sigma_n, \tau_n}^{low} &= \text{ReLU}(\mathbf{f}_{\partial_{\sigma_n}^L} + \mathbf{f}_{\partial_{\sigma_n}^R} + \mathbf{f}_{\sigma_n}), \quad \mathbf{z}_{\sigma_n}^{low} = [(\mathbf{h}_{\partial_{\sigma_n}^L} + \mathbf{h}_{\partial_{\sigma_n}^R}) \parallel \mathbf{h}_{\sigma_n}], \end{aligned}$$

where  $\mathbf{W}_n^{low}$  and  $\mathbf{W}_{n-1}$  are weight matrices.

The path feature is updated by using message from both upper embedding and low embedding. First, upper and lower path message is generated from the upper embedding and low embedding respectively as follows,

$$\begin{aligned} \mathbf{c}_{\sigma_n}^{up/low} &= \text{ReLU}(\mathbf{z}_{\sigma_n}^{up/low} \mathbf{W}_n^{up/low} + b^{up/low}), \quad \alpha_{\sigma_n, \tau_n}^{up/low} = \frac{e_{\sigma_n, \tau_n}^{up/low}}{\sum_{\kappa_n \in \mathcal{N}(\sigma_n)} e_{\sigma_n, \kappa_n}^{up/low}}, \\ \mathbf{m}_{\sigma_n}^{up/low} &= \text{LeakyReLU}(\mathbf{c}_{\sigma_n}^{up/low} + \sum_{\tau_n \in \mathcal{N}(\sigma_n)} \alpha_{\sigma_n, \tau_n}^{up/low} \mathbf{c}_{\tau_n}^{up/low}), \end{aligned}$$

then path feature is updated by using both messages as follows,

$$\mathbf{h}_{\sigma_n}^{(l+1)} = \text{LeakyReLU}((\mathbf{m}_{\sigma_n}^{low})^{(l)} + (\mathbf{m}_{\sigma_n}^{up})^{(l)}).$$

Note that  $\mathbf{h}_{\sigma_n}^{(l+1)}$  means the updated feature vector for  $n$ -path  $\sigma_n$  at the  $(l+1)$ -th layer. It depends on the upper and lower message information at the  $l$ -th layer.

### A.3 DATASET DETAILS, MIN-MAX SCALING, SPLITTING METHO AND MEAN ABSOLUTE ERR

In this study, we analyzed five datasets from MoleculeNet Wu et al. (2018) and MolBench Jiang et al. (2023): QM7 Blum & Reymond (2009), QM9 Ruddigkeit et al. (2012), Tox21, Hiv and Muv, all of which are publicly available on the MoleculeNet website: <https://moleculenet.org/datasets-1>. Details about these datasets are in Table 8. Note that the subindex indicates standard deviation val-

Table 8: The details of the datasets. Note that the subindex indicates standard deviation values.

Dataset	QM7	QM9	Tox21	HIV	MUV
No. molecules	6,830	133,885	7831	41127	93808
No. average atoms	16 <sub>(3)</sub>	18 <sub>(3)</sub>	36 <sub>(23)</sub>	46 <sub>(24)</sub>	43 <sub>(10)</sub>
No. tasks	1	3	12	1	17
Task type	Regression	Regression	Classification	Classification	Classification
Evaluation	MAE	MAE	ROC-AUC	ROC-AUC	ROC-AUC

ues. For instance, the element 16<sub>(13)</sub> means the number of average atoms in QM7 is 16, with 13 as its standard deviation. The QM7 dataset is a subset of the GDB-13 database Blum & Reymond (2009), which contains approximately 1 billion organic molecules with up to seven "heavy" atoms (C, N, O, S). The QM9 dataset, a subset of the GDB-17 database, provides twelve properties, encompassing geometric, energetic, electronic, and thermodynamic properties, following the baseline methods PCMP use the electronic spatial extent ( $\alpha$ ), and the energies of the highest occupied molecular orbital ( $\epsilon$ HOMO) and the lowest unoccupied molecular orbital ( $\epsilon$ LUMO) as the targets. Tox21 is qualitative toxicity measurements on 12 biological targets, including nuclear receptors and stress response pathways. HIV is experimentally measured abilities to inhibit HIV replication. MUV is subset of PubChem BioAssay by applying a refined nearest neighbor analysis, designed for validation of virtual screening techniques.

**Min-Max Scaling** Given that QM7 and QM9 involve regression, we applied min-max normalization to scale target values between 0 and 1. In multiple-target regression tasks, Min-Max Scaling is commonly used to normalize the targets. This technique linearly transforms the target values to a specified range between a minimum and maximum value. The transformation follows the formula:

$$\bar{y} = \frac{y - y_{\min}}{y_{\max} - y_{\min}}, \quad y_{\text{scal}} = y_{\max} - y_{\min} \quad (1)$$

Here,  $\bar{y}$  represents the normalized target value,  $y$  is the original target value,  $y_{\min}$  is the minimum value of the target, and  $y_{\max}$  is the maximum value of the target.

During prediction, the normalized predictions obtained from the model need to be transformed back to the original scale of the target values. The transformation is performed using the formula:

$$\tilde{y} = \hat{y} \cdot y_{\text{scal}} + y_{\min}, \quad y = \bar{y} \cdot y_{\text{scal}} + y_{\min} \quad (2)$$

where  $\hat{y}$  is the model output, and  $\tilde{y}$  and  $y$  are used for loss function computation and evaluation.

This normalization process ensures that all target values are scaled within a fixed range, typically between 0 and 1. It facilitates better convergence during model training and helps in handling targets with varying scales effectively. Furthermore, Min-Max Scaling maintains the relative relationships between target values while bringing them into the desired range, making it a suitable choice for multiple-target regression tasks.

**Splitting Method** Following the work of Bharath Ramsundar Ramsundar et al. (2019), we employed scaffold splitting to partition all datasets. This method segments molecules based on their scaffolds (molecular substructures). Scaffold splitting is a more challenging partitioning approach that can better evaluate a model’s generalization ability on out-of-distribution data samples. To ensure a fair comparison with other models, we adopted the same scaffold splitting method to divide the task datasets into training, validation, and test sets with a ratio of 8:1:1.

**MAE (Mean Absolute Error)** The Mean Absolute Error (MAE) is defined as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3)$$

where  $y_i$  and  $\hat{y}_i$  represent the true value and predicted value of the  $i^{th}$  sample respectively. MAE is a commonly used metric for evaluating regression performance. A lower MAE value indicates higher prediction accuracy, with a decrease in MAE typically suggesting improved model performance.

Table 9: Hyperparameters set up.

Dataset	QM7	QM9	Tox21	HIV	MUV
Learning rate	1e-4	1e-3	1.5e-4	1e-3	1e-4
Batch size	512	512	512	512	512
No.heads	1	6	6	2	1
No.layers	2	2	2	2	2
Train/Valid/Test	8:1:1	8:1:1	8:1:1	8:1:1	8:1:1
Loss function	L1	L1	BCE	BCE	BCE
Optimizer	ADAM	ADAM	ADAM	ADAM	ADAM
Epochs	500	500	1000	1000	1000
Seed	42	42	42	42	42

#### A.4 HYPERPARAMETERS SETUP

We have set up a set of hyperparameters for training the model are summarized in Table 9. In addition, the optimizer selected as ADAM, and the loss function chosen as L1. All models are trained using NVIDIA RTX A5000 32GB GPUs. The running times are show in the table

## B MATHEMATICAL ANALYSIS OF PATH COMPLEX

### B.1 PATH COMPLEX

**Definition B.1** (Elementary path Grigor’yan et al. (2012)). Given a set  $V$ , an elementary  $n$ -path of  $V$  is any sequence of  $n + 1$  elements  $v_0 v_1 \cdots v_n$  of  $V$ , denoted by  $\sigma_n = v_0 v_1 \cdots v_n$

**Definition B.2** (Path complex Grigor’yan et al. (2012)). A path complex  $P$  over the vertex set  $V$  is a collection of elementary paths of  $V$  such that  $\forall \sigma_n = v_0 v_1 \cdots v_n \in P, v_1 \cdots v_n \in P, v_0 v_1 \cdots v_{n-1} \in P$ .

The element  $\sigma_n$  of  $P$  that has  $n + 1$  vertices is called an  $n$ -path of  $P$ . The path  $\sigma$  is called a face of the path  $\tau$  if  $\sigma$  is derived from  $\tau$  by removing the first or last vertex. The  $n$ -path  $\tau_n$  is called a coface of  $(n - 1)$ -path  $\sigma_{n-1}$  if  $\sigma_{n-1}$  is a face of  $\tau_n$ . Two  $n$ -paths are upper adjacent if they are faces of a common  $(n + 1)$ -path, lower adjacent if they have a common  $(n - 1)$ -path as face. For an  $n$ -path  $\sigma_n$ , let  $\mathcal{B}(\sigma_n)$  be the set of faces of  $\sigma_n$ ,  $\mathcal{C}(\sigma_n)$  be the set of cofaces of  $\sigma_n$ ,  $\mathcal{N}_\uparrow(\sigma_n)$  be the set of  $n$ -paths that are upper adjacent with  $\sigma_n$ ,  $\mathcal{N}_\downarrow(\sigma_n)$  be the set of  $n$ -paths that are lower adjacent with  $\sigma_n$ . Note that we can use the above four relations, including face-relation, coface-relation, upper adjacency and lower adjacency, to define the neighbors of an  $n$ -path  $\sigma_n$ . We give construction of path complex from graphs, simplicial complexes and hypergraphs. We give construction of path complex from graphs, simplicial complexes and hypergraphs.

### B.1.1 PATH COMPLEX FROM GRAPHS

**Definition B.3** (Path). Given an undirected graph  $G = (V, E)$  over the vertex set  $V$ , we define the  $n$ -path  $\sigma_n$  of  $G$  as any sequence of  $n + 1$  vertices  $v_0 v_1 \cdots v_n (v_i \in V)$  satisfying the following conditions:

1.  $\forall i (0 \leq i < n), (v_i, v_{i+1}) \in E$  or  $(v_{i+1}, v_i) \in E$ .
2.  $\forall i \neq j, v_i \neq v_j$ .

Note that for each  $n$ -path  $\sigma_n = v_0 v_1 \cdots v_n, \sigma'_n = v_n \cdots v_1 v_0$  is also an  $n$ -path, we identify these two paths as the same one.

**Definition B.4** (Path complex from graphs). Given an undirected graph  $G = (V, E)$ , let  $P_n$  be the set of all  $n$ -paths of  $G$ , then  $P_G = \bigcup_n P_n$  form a path complex. We call  $P_G$  the path complex derived from  $G$ .

It can be seen that the path complex  $P_G$  derived from  $G$  is determined by  $P_0$  and  $P_1$ .

**Theorem B.5** (Path Complex Invariance). *The graph neural network is invariant to the permutation of the simplexes in the path complex  $P_G$ , meaning the output of the network remains unchanged under any permutation  $\pi$  of the vertices.*

### B.1.2 PATH COMPLEX FROM SIMPLICIAL COMPLEX

**Definition B.6** (Simplicial complex). A simplicial complex  $\mathcal{K}$  over the vertex set  $V$  is a collection of vertex subsets of  $V$  satisfying that if  $\sigma \in \mathcal{K}, \tau \subset \sigma, \tau \in \mathcal{K}$ .

The element  $\sigma_k$  of  $\mathcal{K}$  that has  $k + 1$  vertices is called a  $k$ -simplex. A simplex  $\sigma$  is called a face of the simplex  $\tau$  and  $\tau$  is called a coface of  $\sigma$  if  $\sigma \subset \tau$ . For any two  $k$ -simplices  $\sigma_k, \tau_k \in \mathcal{K}$ ,  $\sigma_k$  and  $\tau_k$  are called upper adjacent if they are both faces of an  $(k + 1)$ -simplex  $\alpha_{k+1} \in \mathcal{K}$ . Two  $k$ -simplices  $\sigma_k, \tau_k$  are called lower adjacent if they share a common  $(k - 1)$ -simplex as faces.

**Definition B.7** (Path). Given a simplicial complex  $\mathcal{K}$ , we define an  $(k, n)$ -path  $e_n^k$  of  $\mathcal{K}$  as a sequence of  $n + 1$   $k$ -simplices  $\sigma_k^0 \sigma_k^1 \cdots \sigma_k^n$  satisfying the following conditions:

1.  $\forall i (0 \leq i < n), \sigma_k^i$  and  $\sigma_k^{i+1}$  are upper adjacent.
2.  $\forall i \neq j, \sigma_k^i \neq \sigma_k^j$

We can also use lower adjacent, face and the coface relation to define paths. Note that for each  $(k, n)$ -path  $\sigma_k^0 \sigma_k^1 \cdots \sigma_k^n$ , there is an  $(k, n)$ -path  $\sigma_k^n \cdots \sigma_k^0$ . We identify these two paths as the same one.

**Definition B.8** (Path complex from simplicial complex). Given a simplicial complex  $\mathcal{K}$ , let  $P_n^k$  be the set of all  $(k, n)$ -paths of  $\mathcal{K}$ , then  $P_{\mathcal{K}}^k = \bigcup_n P_n^k$  form a path complex.

For the simplicial complex  $\mathcal{K}$ , its one-skeleton forms a graph  $\mathcal{K}_1$ , we have  $P_{\mathcal{K}_1} = P_{\mathcal{K}}^0$ .

### B.1.3 PATH COMPLEX FROM HYPERGRAPHS

**Definition B.9** (Hypergraph). A hypergraph  $\mathcal{H}$  over the vertex set  $V$  is a collection of vertex subsets of  $V$ .

The element  $\sigma_k$  of  $\mathcal{H}$  that has  $k + 1$  vertices is called an  $k$ -hyperedge. Two hyperedges are called lower adjacent if their intersection is not empty.

**Definition B.10** (Path). Given a hypergraph  $\mathcal{H}$  over the vertex set  $V$ , we define an  $n$ -path of  $\mathcal{H}$  as a sequence of  $n + 1$  hyperedges  $\sigma^0 \sigma^1 \cdots \sigma^n$  such that any two adjacent hyperedges are lower adjacent and any two hyperedges are not same.

**Definition B.11** (Path complex from hypergraphs). Given a hypergraph  $\mathcal{H}$ , let  $P_n$  be the set of all  $n$ -paths of  $\mathcal{H}$ , then  $P_{\mathcal{H}} = \bigcup_n P_n$  form a path complex.

### B.2 HOMOLOGY OF PATH COMPLEX

The homology of path complex is a new homology theory that breaks the landscape of classical homology theory in algebraic topology, introducing a new framework for exploring the topology of more general mathematical structures Grigor'yan et al. (2012); Grigor'yan et al. (2014; 2020). This homology theory was initially called path homology and renamed GLMY homology in 2022, which advances the study of topological foundations for complex networks Chowdhury & Mémoli (2018); Chowdhury et al. (2022) and has been successfully applied in complex disease Wu et al. (2023), biology and material sciences Chen et al. (2023). Next, we give the construction of homology of path complexes.

Given a path complex  $P$  over  $V$ , We fix a field coefficient  $\mathbb{F}$ , let  $\Lambda_n(P)$  be the vector space spanned by all the elementary  $n$ -paths of  $P$ . Considering the standard boundary operator  $\partial_n : \Lambda_n(P) \rightarrow \Lambda_{n-1}(P)$

$$\forall \sigma_n = v_0 v_1 \cdots v_n \in P, \quad \partial_n(\sigma_n) = \sum_{i=0}^n (-1)^i v_0 \cdots v_{i-1} v_{i+1} \cdots v_n$$

We have  $\partial_n \partial_{n+1} = 0$ . Let  $\mathcal{A}_n(P)$  be the vector space spanned by all the  $n$ -paths of  $P$ , usually  $\partial(\mathcal{A}_n) \not\subset \mathcal{A}_{n-1}(P)$ . We consider the following subspace  $\Omega_n(P)$  of  $\mathcal{A}_n(P)$

$$\Omega_n(P) = \{u \in \mathcal{A}_n(P) \mid \partial(u) \in \mathcal{A}_{n-1}(P)\}$$

Then we have  $\partial_n(\Omega_n(P)) \subset \Omega_{n-1}(P)$ . Consequently, we get a chain complex  $(\Omega_*(P), \partial_*)$

$$\cdots \rightarrow \Omega_{n+1}(P) \xrightarrow{\partial_{n+1}} \Omega_n(P) \xrightarrow{\partial_n} \Omega_{n-1}(P) \rightarrow \cdots$$

**Definition B.12** (Homology of path complex). Given a path complex  $P$ , its  $k$ -homology is defined as the  $k$ -th homology of the chain complex  $(\Omega_*(P), \partial_*)$

$$H_k(P) = H_k((\Omega_*, \partial_*))$$

This definition can be directly applied to the path complexes derived from graphs, simplicial complexes and hypergraphs.

### B.3 WEAK ISOMORPHISM INVARIANCE OF THE PATH COMPLEX HOMOLOGY

For an undirected graph  $G = (V, E)$ , the degree of a vertex  $v \in V$  is the number of edges that contain  $v$  and we denoted it by  $deg(v)$ .

**Definition B.13** (Graph collapse and expansion). Given a graph  $G = (V, E)$ , take an edge  $(v_1, v_2) \in E$  such that  $deg(v_1) = 1$ . Let  $V' = V \setminus \{v_1\}$ ,  $E' = E \setminus \{(v_1, v_2)\}$ , then  $G' = (V', E')$  is a new graph. We say that  $G'$  is derived from  $G$  by a graph collapse and  $G$  is derived from  $G'$  by a graph expansion.

**Definition B.14** (Weak isomorphic). Given two graphs  $G_1, G_2$ ,  $G_1$  and  $G_2$  are called weak isomorphic if  $G_1$  can be derive from  $G_2$  by a sequence of graph collapse and expansion operations.

It is obvious that two graphs are weak isomorphic if they are isomorphic.

1134 **Theorem B.15.** *If two graphs  $G_1$  and  $G_2$  are weak isomorphic, then*

- 1135  
1136 1. *The number of connected components of  $G_1$  and  $G_2$  are same.*  
1137  
1138 2. *The number of cycles of  $G_1$  and  $G_2$  are same.*

1139  
1140 *Proof.* Let  $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ , without loss of generality, we can assume that  $G_2$  is  
1141 derived by collapsing an edge  $(v_1, v_2) \in E_1$  from  $G_1$  and  $\deg(v_1) = 1$ .

- 1142  
1143 1. This is obvious.  
1144  
1145 2. Let  $C(G_i)$  be the set of cycles of  $G_i$ , then we have  $C(G_2) \subset C(G_1)$  because  $G_2$  is a  
1146 subgraph of  $G_1$ .  $\forall c \in C(G_1)$ ,  $c$  is a sequence of vertices such that the degree of each  
1147 vertex is 2. So  $v_1$  is not contained in  $c$ ,  $c \in C(G_2)$ . Consequently,

$$C(G_1) = C(G_2)$$

1148  
1149  
1150 □

1151 **Theorem B.16.** *Given two graphs  $G_1, G_2$ , let  $P_{G_1}, P_{G_2}$  be the path complexes derived from  $G_1$  and  
1152  $G_2$  respectively. If  $G_1$  and  $G_2$  are weak isomorphic, then*

$$1153 \quad H_k(P_{G_1}) \cong H_k(P_{G_2}) \quad (k \geq 0)$$

1154  
1155  
1156  
1157  
1158 *Proof.* Let  $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ , without loss of generality, we can assume that  $G_2$  is  
1159 derived by collapsing an edge  $(v_1, v_2) \in E_1$  from  $G_1$  and  $\deg(v_1) = 1$ .

- 1160  
1161 1.  $k = 0$

$$1162 \quad \Omega_0(P_{G_1}) = \langle v_1 \rangle \oplus \Omega_0(P_{G_2}), \quad \Omega_0(P_{G_2}) = \langle v \mid v \in V_2 \rangle$$

$$1163 \quad \Omega_1(P_{G_1}) = \langle v_1 v_2 \rangle \oplus \Omega_1(P_{G_2}), \quad \Omega_1(P_{G_2}) = \langle e \mid e \in E_2 \rangle$$

1164 We have  $\text{Ker} \partial|_{\Omega_0(P_{G_1})} = \langle v_1 \rangle \oplus \text{Ker} \partial|_{\Omega_0(P_{G_2})}$ ,  $\text{Im} \partial|_{\Omega_1(P_{G_1})} = \langle v_2 - v_1 \rangle$   
1165  $\oplus \text{Im} \partial|_{\Omega_1(P_{G_2})} = \langle v_1 \rangle \oplus \text{Im} \partial|_{\Omega_1(P_{G_2})}$ . Consequently,

$$1166 \quad H_0(P_{G_1}) = \frac{\text{Ker} \partial|_{\Omega_0(P_{G_1})}}{\text{Im} \partial|_{\Omega_1(P_{G_1})}}$$

$$1167 \quad = \frac{\langle v_1 \rangle \oplus \text{Ker} \partial|_{\Omega_0(P_{G_2})}}{\langle v_1 \rangle \oplus \text{Im} \partial|_{\Omega_1(P_{G_2})}}$$

$$1168 \quad = \frac{\text{Ker} \partial|_{\Omega_0(P_{G_2})}}{\text{Im} \partial|_{\Omega_1(P_{G_2})}}$$

$$1169 \quad = H_0(P_{G_2})$$

- 1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178 2.  $k = 1$

$$1179 \quad \Omega_1(P_{G_1}) = \langle v_1 v_2 \rangle \oplus \Omega_1(P_{G_2}), \quad \Omega_1(P_{G_2}) = \langle e \mid e \in E_2 \rangle$$

1180 The degree of vertex  $v_1$  is one means  $v_1$  only appears in the 1-path  $(v_1, v_2)$ , so  $(v_1, v_2)$   
1181 cannot be contained in the kernel of  $\partial$  on  $\Omega_1(P_{G_1})$ , which means that

$$1182 \quad \text{Ker} \partial|_{\Omega_1(P_{G_1})} = \text{Ker} \partial|_{\Omega_1(P_{G_2})}$$

$$1183 \quad \mathcal{A}_2(P_{G_1}) = \langle v_1 v_2 v \mid v_2 \neq v \in V_2 \rangle \oplus \mathcal{A}_2(P_{G_2})$$

1184 Note that  $(v_1, v)$  is not a 1-path for any  $v \in V_2 (v \neq v_2)$ , so

$$1185 \quad \Omega_2(P_{G_1}) = \Omega_2(P_{G_2})$$

Consequently,

$$\begin{aligned} H_1(P_{G_1}) &= \frac{\text{Ker}\partial|_{\Omega_1(P_{G_1})}}{\text{Im}\partial|_{\Omega_2(P_{G_1})}} \\ &= \frac{\text{Ker}\partial|_{\Omega_1(P_{G_2})}}{\text{Im}\partial|_{\Omega_2(P_{G_2})}} \\ &= H_1(P_{G_2}) \end{aligned}$$

3.  $k \geq 2$ . It suffices to prove that

$$\Omega_k(P_{G_1}) = \Omega_k(P_{G_2}) \quad (k \geq 2)$$

It is obvious that  $\Omega_k(P_{G_2}) \subset \Omega_k(P_{G_1})$ . So we only need to prove that  $\Omega_k(P_{G_1}) \subset \Omega_k(P_{G_2})$ .

(a) We prove that

$$\Omega_k(P_{G_1}) \subset \mathcal{A}_k(P_{G_2})$$

$\forall \omega_k \in \Omega_k(P_{G_1}), \omega_k \in \mathcal{A}_k(P_{G_1}), \partial(\omega_k) \in \mathcal{A}_{k-1}(P_{G_1})$ . Note that every  $k$ -path in  $P_{G_1}$  is either an  $k$ -path in  $P_{G_2}$  or starts with  $v_1v_2$ , so  $\omega_k$  can be represented as

$$\omega_k = v_1v_2e_{k-2} + e_k$$

where  $e_{k-2} \in \mathcal{A}_{k-2}(P_{G_2})$  is a linear combination of  $(k-2)$ -paths  $\sum v_{i_0}v_{i_1} \cdots v_{i_{k-2}}$  ( $v_{i_0} \neq v_1$ ) of  $P_{G_2}$  and  $e_k$  is an  $k$ -path of  $P_{G_2}$ . We have

$$\partial(\omega_k) = (v_2 - v_1)e_{k-2} + v_1v_2\partial(e_{k-2}) + \partial(e_k)$$

Note that  $v_1e_{k-2}$  is a linear combination of  $(k-1)$ -paths  $v_1v_{i_0} \cdots v_{i_{k-2}}$  ( $v_1 \neq v_{i_0}$ ). Since  $v_1v_{i_0}$  is not an edge, these paths are not contained in  $P_{G_1}$ , but  $\partial(\omega_k) \in \mathcal{A}_{k-1}(P_{G_1})$ , so  $v_1e_{k-2}$  must add some item in the right part to become zero. There is not any item in the right part of the equation has  $v_1v_{i_0} \cdots v_{i_{k-2}}$ , so  $v_1e_{k-2}$  must be zero, which means that  $e_{k-2}$  is zero. Consequently,

$$\omega_k = e_k \in \mathcal{A}_k(P_{G_2})$$

(b) We prove that

$$\Omega_k(P_{G_2}) = \mathcal{A}_k(P_{G_2}) \cap \Omega_k(P_{G_1})$$

It is obvious that  $\Omega_k(P_{G_2}) \subset \mathcal{A}_k(P_{G_2}) \cap \Omega_k(P_{G_1})$ , so we only need to prove that  $\mathcal{A}_k(P_{G_2}) \cap \Omega_k(P_{G_1}) \subset \Omega_k(P_{G_2})$ .

$\forall e \in \mathcal{A}_k(P_{G_2}) \cap \Omega_k(P_{G_1})$ , Since  $e \in \mathcal{A}_k(P_{G_2})$ ,  $v_1$  will not appear in  $e$ , which means  $e$  is a path of  $P_{G_2}$ . Note that  $e \in \Omega_k(P_{G_1})$ , so  $\partial(e) \in \mathcal{A}_{k-1}(P_{G_1})$ , with the property that  $e \in P_{G_2}$ , we have  $\partial(e) \in \mathcal{A}_{k-1}(P_{G_2})$ , which means

$$e \in \Omega_k(P_{G_2})$$

Combining the results of (a) and (b), we have

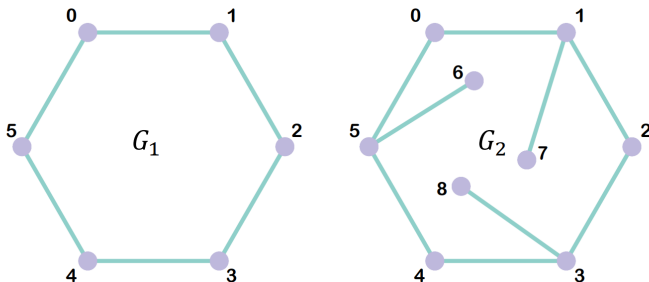
$$\Omega_k(P_{G_1}) \subset \mathcal{A}_k(P_{G_2}) \cap \Omega_k(P_{G_1}) = \Omega_k(P_{G_2})$$

□

Theorem B.16 means the path complex homology is a graph weak isomorphism invariant. Consequently, for two graphs  $G_1$  and  $G_2$ , if there exists  $k$  such that  $H_k(P_{G_1}) \not\cong H_k(P_{G_2})$ , then  $G_1$  and  $G_2$  are not weak isomorphic and not isomorphic.

Figure 7 illustrates an example of the graph weak isomorphism. As shown in Figure 1,  $G_1 = (V_1, E_1)$  where  $V_1 = \{0, 1, 2, 3, 4, 5\}$ ,  $E_1 = \{(0, 1), (1, 2), (2, 3), (3, 4), (4, 5), (0, 5)\}$ .  $G_2 = (V_2, E_2)$  where  $V_2 = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$  and  $E_2 = \{(0, 1), (1, 2), (2, 3), (3, 4), (4, 5), (0, 5), (5, 6), (1, 7), (3, 8)\}$ .  $G_1$  can be derived by doing the graph collapse operation on  $G_2$  through  $\{7, (1, 7)\}$ ,  $\{6, (5, 6)\}$  and  $\{8, (3, 8)\}$  one by one, so  $G_1$  and  $G_2$  are weak isomorphic. It can be seen that  $G_1$  and  $G_2$  both have one cycle and one connected component.

1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251



1252  
1253  
1254  
1255

Figure 7: Illustration of the graph weak isomorphism.  $G_1$  can be derived by doing the graph collapse operations on  $G_2$  through  $\{7, (1, 7)\}$ ,  $\{6, (5, 6)\}$  and  $\{8, (3, 8)\}$  one by one, so  $G_1$  and  $G_2$  are weak isomorphic.

1256

### C PATH COMPLEX INVARIANCE

1257

1258

1259

1260

1261

1262

1263

1264

1265

1266

1267

1268

1269

1270

1271

1272

1273

1274

1275

1276

1277

1278

1279

1280

1281

1282

1283

1284

1285

1286

1287

1288

1289

1290

1291

1292

1293

1294

1295

Molecules can be effectively represented as graphs  $G = (V, E)$ , where vertices  $V$  correspond to atoms, and edges  $E$  represent chemical bonds. Building on this, a path complex  $P_G$  is constructed from the graph, extending the connectivity into higher-dimensional simplices based on paths in  $G$ . These simplices encapsulate relationships such as chemical bond types and angular constraints between bonds, further enriching the molecular representation.

**Definition C.1** (Permutation Operation). A permutation  $\pi$  is a bijection on the set of vertices  $V$ , representing a relabeling of the atoms in the molecule. This permutation induces a corresponding bijection on the path complex  $P_G$ , noted as  $\pi(P_G)$ , which preserves the connectivity structure of the original path complex.

**Definition C.2** (Graph Neural Network Layer for Path Complexes). Consider a generic layer of a graph neural network adapted for path complexes, updating the features of each simplex based on a message passing scheme:

$$h_{\sigma_n}^{(k+1)} = \text{RELU} \left( W^{(k)} \cdot h_{\sigma_n}^{(k)} + \sum_{\tau_n \in \mathcal{N}(\sigma_n)} U^{(k)} \cdot h_{\tau_n}^{(k)} \right)$$

$h_{\sigma_n}^{(k)}$  denotes the feature vector of the  $n$ -order path  $\sigma_n$  at layer  $k$ .  $W^{(k)}$  and  $U^{(k)}$  are learnable parameters of the network at layer  $k$ .  $\mathcal{N}(\sigma_n)$  represents the neighborhood set of simplex  $\sigma_n$  in  $P_G$ , including faces and cofaces.

*Proof of Theorem B.5.* 1. Initially, let  $h_{\sigma_n}^{(0)} = h_{\pi(\sigma_n)}^{(0)}$  for all  $\sigma_n \in P_G$ , assuming that the initial features are assigned consistently regardless of the labeling of the paths.

2. Assume that  $h_{\sigma_n}^{(k)} = h_{\pi(\sigma_n)}^{(k)}$  holds true for some layer  $k$ .

3. To prove for layer  $k + 1$ :

$$h_{\sigma_n}^{(k+1)} = \text{RELU} \left( W^{(k)} \cdot h_{\sigma_n}^{(k)} + \sum_{\tau_n \in \mathcal{N}(\sigma_n)} U^{(k)} \cdot h_{\tau_n}^{(k)} \right)$$

$$h_{\pi(\sigma_n)}^{(k+1)} = \text{RELU} \left( W^{(k)} \cdot h_{\pi(\sigma_n)}^{(k)} + \sum_{\pi(\tau_n) \in \mathcal{N}(\pi(\sigma_n))} U^{(k)} \cdot h_{\pi(\tau_n)}^{(k)} \right)$$

Given the inductive hypothesis that  $h_{\sigma_n}^{(k)} = h_{\pi(\sigma_n)}^{(k)}$  and  $h_{\tau_n}^{(k)} = h_{\pi(\tau_n)}^{(k)}$  for all  $\tau_n \in \mathcal{N}(\sigma_n)$ , and recognizing that  $\mathcal{N}(\pi(\sigma_n)) = \pi(\mathcal{N}(\sigma_n))$ , it follows that:



1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349

$$h_{\sigma_n}^{(k+1)} = h_{\pi(\sigma_n)}^{(k+1)}$$

□

## D PATH WEISFEILER LEHMAN (PWL) TEST

### D.1 PATH COMPLEX

**Definition D.1** (Path Complex Isomorphism). Given two path complexes  $P_1, P_2$  over the vertices  $V_1, V_2$ .  $P_1$  and  $P_2$  are called isomorphic if there is a map  $f : V_1 \rightarrow V_2$  such that  $\sigma_n = v_0 v_1 \cdots v_n \in P_1 \iff f(\sigma) = f(v_0) f(v_1) \cdots f(v_n) \in P_2$ .

**Theorem D.2.** Given two graphs  $G_1, G_2$ , let  $P_{G_1}, P_{G_2}$  be the path complexes derived from  $G_1, G_2$  respectively. We have

$$G_1 \cong G_2 \iff P_{G_1} \cong P_{G_2}$$

### D.2 PATH COMPLEX COLORING

**Definition D.3** (Path Coloring). A path coloring is a map  $c$  such that for each path complex  $P$  and any path  $\sigma$  of  $P$ ,  $c(\sigma)$  is a color from a fixed color table. We denote this color by  $c_\sigma^P$ .

We will often omit  $P$  in the subscript when the underlying path complex is arbitrary.

**Definition D.4.** Given two path complexes  $P_1, P_2$  and a path coloring  $c$ .  $P_1$  and  $P_2$  are called  $c$ -similar, denoted by  $c^{P_1} = c^{P_2}$ , if for any dimension  $n$ , we have the color multi-sets equality

$$\{\{c_\sigma^{P_1} \mid \dim(\sigma) = n, \sigma \in P_1\}\} = \{\{c_\tau^{P_2} \mid \dim(\tau) = n, \tau \in P_2\}\}$$

**Definition D.5** (PWL). We give a path complex version of the WL test to derive a message passing procedure that can retain the expressive power of the test. We call this the Path WL (PWL), the steps of general PWL are as follows:

1. Given a path complex  $P$ , all the paths of  $P$  are initialized with the same color.
2. For the color  $c_\sigma^t$  of path  $\sigma$  at iteration  $t$ , the color  $c_\sigma^{t+1}$  of  $\sigma$  at the next iteration is computed by perfectly hashing the color multi-set of the neighbors of  $\sigma$ .
3. The algorithm stops once a stable coloring is reached. Two path complexes are considered non-isomorphic if their color histograms are different at some dimensions.

**Neighbor Color Multi-set** Based on the four neighbor definitions, we have four types of neighbor color multi-sets. Let  $c^t$  be the coloring of PWL for path complex  $P$  at iteration  $t$ , four types of color multi-sets are as follows

1.  $c_{\mathcal{B}}^t(\sigma) = \{\{c_\tau^t \mid \tau \in \mathcal{B}(\sigma)\}\}$
2.  $c_{\mathcal{C}}^t(\sigma) = \{\{c_\tau^t \mid \tau \in \mathcal{C}(\sigma)\}\}$
3.  $c_{\uparrow}^t(\sigma) = \{\{(c_\tau^t, c_{\sigma \cup \tau}^t) \mid \tau \in \mathcal{N}_{\uparrow}(\sigma)\}\}$
4.  $c_{\downarrow}^t(\sigma) = \{\{(c_\tau^t, c_{\sigma \cap \tau}^t) \mid \tau \in \mathcal{N}_{\downarrow}(\sigma)\}\}$

Having the neighbor color multi-sets, we obtain the following update rule that contains all four types of neighbors:

$$c_\sigma^{t+1} = \text{HASH}\{c_\sigma^t, c_{\mathcal{B}}^t(\sigma), c_{\mathcal{C}}^t(\sigma), c_{\uparrow}^t(\sigma), c_{\downarrow}^t(\sigma)\}$$

Actually, certain neighbors can be removed without affecting the expressive power of PWL test in terms of path complex that can be differentiated.

**Theorem D.6.** PWL with  $\text{HASH}\{c_\sigma^t, c_{\mathcal{B}}^t(\sigma), c_{\uparrow}^t(\sigma)\}$  is as powerful as PWL with the four-neighbor-updating strategy  $\text{HASH}\{c_\sigma^t, c_{\mathcal{B}}^t(\sigma), c_{\mathcal{C}}^t(\sigma), c_{\uparrow}^t(\sigma), c_{\downarrow}^t(\sigma)\}$ .

**Theorem D.7.** PWL is strictly more powerful than WL.

**Theorem D.8.** PWL is no less powerful than SWL Bodnar et al. (2021b) with the clique complex lifting.

### D.3 PATH COMPLEX MESSAGE PASSING

We propose a general Path Complex Message Passing (PCMP) using the following messages passing operations. For a path  $\sigma$  in  $P$ , we have

$$m_{\mathcal{B}}^{t+1}(\sigma) = AGG_{\tau \in \mathcal{B}(\sigma)}(M_{\mathcal{B}}(h_{\sigma}^t, h_{\tau}^t)) \quad (4)$$

$$m_{\uparrow}^{t+1}(\sigma) = AGG_{\tau \in \mathcal{N}_{\uparrow}(\sigma)}(M_{\uparrow}(h_{\sigma}^t, h_{\tau}^t, h_{\sigma \cup \tau}^t)) \quad (5)$$

Then, the updating function considers these two types of messages and the previous color of  $\sigma$ :

$$h^{t+1}(\sigma) = U(h_{\sigma}^t, m_{\mathcal{B}}^t(\sigma), m_{\uparrow}^t(\sigma)) \quad (6)$$

After  $L$  layers of the message passing process, the readout function takes the color multi-sets at all dimensions as input:

$$h_P = \text{READOUT}(\{\{h_{\sigma}^L\}_{\dim(\sigma)=0}, \dots, \{h_{\tau}^L\}_{\dim(\tau)=p}\}) \quad (7)$$

**Theorem D.9.** *PCMP with sufficient layers and injective neighborhood aggregators are as powerful as PWL.*

### D.4 PROOF OF MAIN RESULTS

In order to prove the main results, we give some notations.

**Definition D.10** (Path Coloring Refinement). A path coloring  $c$  refines a path coloring  $d$ , denoted by  $c \sqsubseteq d$ , if for any path complex  $P_1, P_2$  and  $\sigma \in P_1, \tau \in P_2$ ,  $c_{\sigma}^{P_1} = c_{\tau}^{P_2}$  implies  $d_{\sigma}^{P_1} = d_{\tau}^{P_2}$ . Additionally, if  $d \sqsubseteq c$ , we say that  $c$  and  $d$  are equivalent.

**Lemma D.11.** *Given two path complexes  $P_1, P_2$  with  $A \subset P_1, B \subset P_2$ . Assume  $c$  and  $d$  are two path coloring such that  $c \sqsubseteq d$ . If  $\{\{d_{\sigma}^{P_1} | \sigma \in A\}\} \neq \{\{d_{\tau}^{P_2} | \tau \in B\}\}$ , then  $\{\{c_{\sigma}^{P_1} | \sigma \in A\}\} \neq \{\{c_{\tau}^{P_2} | \tau \in B\}\}$ .*

*Proof.* Let  $C_1 = \{\{c_{\sigma}^{P_1} | \sigma \in A\}\}$ ,  $C_2 = \{\{c_{\tau}^{P_2} | \tau \in B\}\}$ . Assume  $C_1 = C_2$ , then there is a bijection  $f : A \rightarrow B$  such that  $\forall \sigma \in A, \tau = f(\sigma)$ , we have  $c_{\sigma}^{P_1} = c_{\tau}^{P_2}$ . From  $c \sqsubseteq d$  we know  $d_{\sigma}^{P_1} = d_{\tau}^{P_2}$ . Consequently,  $\{\{d_{\sigma}^{P_1} | \sigma \in A\}\} = \{\{d_{f(\sigma)}^{P_2} | \sigma \in A\}\} = \{\{d_{\tau}^{P_2} | \tau \in B\}\}$ , which contradicts with the condition that  $\{\{d_{\sigma}^{P_1} | \sigma \in A\}\} \neq \{\{d_{\tau}^{P_2} | \tau \in B\}\}$ . Hence the assumption is wrong.  $\square$

**Corollary D.12.** *Given two path colorings  $c$  and  $d$  such that  $c \sqsubseteq d$ . If  $d^{P_1} \neq d^{P_2}$ , then  $c^{P_1} \neq c^{P_2}$ .*

*Proof.* This follows by replacing the subsets  $A, B$  by the sets of  $n$ -paths of  $P_1$  and  $P_2$  respectively in the proof of Lemma D.11.  $\square$

The above corollary D.12 means that if  $c$  refines  $d$ , then  $c$  is able to distinguish all the path complex pairs that  $d$  can distinguish. In this sense, we can say that  $c$  is at least as powerful as  $d$ . If  $c$  and  $d$  are equivalent, we say they have the same expressive power.

*Proof of Theorem D.2.* It is easy to see that if  $G_1 \cong G_2$ , then  $P_{G_1} \cong P_{G_2}$ . The inverse statement follows from the fact that any graph is a subcomplex of its derived path complex by considering the 0-paths and 1-paths.  $\square$

*Proof of Theorem D.6.* Let  $a^t$  be the coloring at iteration  $t$  of the updating strategy

$$\text{HASH}\{a_{\sigma}^t, a_{\mathcal{B}}^t(\sigma), a_{\uparrow}^t(\sigma), a_{\downarrow}^t(\sigma)\}$$

$b^t$  be the coloring at iteration  $t$  of the updating strategy

$$\text{HASH}\{b_{\sigma}^t, b_{\mathcal{B}}^t(\sigma), b_{\uparrow}^t(\sigma), b_{\downarrow}^t(\sigma)\}$$

$c^t$  be the coloring at iteration  $t$  of the updating strategy

$$\text{HASH}\{c_{\sigma}^t, c_{\mathcal{B}}^t(\sigma), c_{\uparrow}^t(\sigma)\}$$

We firstly prove that  $a^t$  and  $b^t$  are equivalent, then prove that  $b^t$  and  $c^t$  are equivalent.

- 1404 1.  $a^t$  and  $b^t$  are equivalent. We have  $a^t \sqsubseteq b^t$  because  $a^t$  contains additional colors of its  
 1405 coface neighbors in the color updating rule. It suffices to prove that  $b^t \sqsubseteq a^t$ . We do this  
 1406 by induction. The base case holds since all the paths are initialized with the same color.  
 1407 Assume the result holds for  $t = k$ , we prove that  $b^{k+1} \sqsubseteq a^{k+1}$ . Let  $\sigma \in P_1$  and  $\tau \in P_2$   
 1408 be two  $n$ -paths from two arbitrary path complexes, suppose  $b_\sigma^{k+1} = b_\tau^{k+1}$ , we prove that  
 1409  $a_\sigma^{k+1} = a_\tau^{k+1}$ .

1410 The equation  $b_\sigma^{k+1} = b_\tau^{k+1}$  means that the hash function at iteration  $t+1$  have the same  
 1411 arguments. Consequently,  $b_\sigma^k = b_\tau^k$ ,  $b_{\mathcal{B}}^k(\sigma) = b_{\mathcal{B}}^k(\tau)$ ,  $b_{\uparrow}^k(\sigma) = b_{\uparrow}^k(\tau)$ ,  $b_{\downarrow}^k(\sigma) = b_{\downarrow}^k(\tau)$ . We  
 1412 prove that  $b_{\mathcal{C}}^k(\sigma) = b_{\mathcal{C}}^k(\tau)$ .

1413 We have  $b_{\uparrow}^k(\sigma) = b_{\uparrow}^k(\tau)$  and

$$1414 \quad b_{\uparrow}^k(\sigma) = \{\{(b_e^k, b_{\sigma \cup e}^k) | e \in \mathcal{N}_{\uparrow}(\sigma)\}\}, b_{\uparrow}^k(\tau) = \{\{(b_e^k, b_{\tau \cup e}^k) | e \in \mathcal{N}_{\uparrow}(\tau)\}\} \quad (8)$$

1415 Replacing the first component of the tuple by the same color, we have

$$1416 \quad \{\{(-, b_{\sigma \cup e}^k) | e \in \mathcal{N}_{\uparrow}(\sigma)\}\} = \{\{(-, b_{\tau \cup e}^k) | e \in \mathcal{N}_{\uparrow}(\tau)\}\} \quad (9)$$

1417 By the definition of upper adjacency and coface we have

$$1418 \quad b_{\mathcal{C}}^k(\sigma) = \{\{b_w^k | w \in \mathcal{C}(\sigma)\}\} = \{\{b_{\sigma \cup e}^k | e \in \mathcal{N}_{\uparrow}(\sigma)\}\} \quad (10)$$

$$1419 \quad b_{\mathcal{C}}^k(\tau) = \{\{b_w^k | w \in \mathcal{C}(\tau)\}\} = \{\{b_{\tau \cup e}^k | e \in \mathcal{N}_{\uparrow}(\tau)\}\} \quad (11)$$

1420 Combining Equation (8), (9), (10), (11), we have  $b_{\mathcal{C}}^k(\sigma) = b_{\mathcal{C}}^k(\tau)$ .

1421 From the induction hypothesis, we have  $a_\sigma^k = a_\tau^k$ ,  $a_{\mathcal{B}}^k(\sigma) = a_{\mathcal{B}}^k(\tau)$ ,  $a_{\mathcal{C}}^k(\sigma) = a_{\mathcal{C}}^k(\tau)$ ,  
 1422  $a_{\uparrow}^k(\sigma) = a_{\uparrow}^k(\tau)$ ,  $a_{\downarrow}^k(\sigma) = a_{\downarrow}^k(\tau)$ , so  $a_\sigma^{k+1} = a_\tau^{k+1}$ .

- 1423 2.  $b^t$  and  $c^t$  are equivalent. Similarly we have  $b^t \sqsubseteq c^t$ , we further prove that  $c^{2t} \sqsubseteq b^t$ . We  
 1424 do this by induction. The base case is obvious because all the paths are initialized with the  
 1425 same color. Assume the results holds for  $t = k$ , we prove that  $c^{2k+2} \sqsubseteq b^{k+1}$ . Let  $\sigma \in P_1$   
 1426 and  $\tau \in P_2$  be two  $n$ -paths from two arbitrary path complexes, suppose  $c_\sigma^{2k+2} = c_\tau^{2k+2}$ ,  
 1427 we prove that  $b_\sigma^{k+1} = b_\tau^{k+1}$ .

1428 For  $c_\sigma^{2k+2} = c_\tau^{2k+2}$ , by going back two steps of the hash function, we have  $c_\sigma^{2k} =$   
 1429  $c_\tau^{2k}$ ,  $c_{\mathcal{B}}^{2k}(\sigma) = c_{\mathcal{B}}^{2k}(\tau)$ ,  $c_{\uparrow}^{2k}(\sigma) = c_{\uparrow}^{2k}(\tau)$ . We want to prove that  $c_{\downarrow}^{2k}(\sigma) = c_{\downarrow}^{2k}(\tau)$ .

1430 Assume  $c_{\downarrow}^{2k}(\sigma) \neq c_{\downarrow}^{2k}(\tau)$ , then there is a color pair  $(c_0, c_1)$  such that  $(c_0, c_1)$  appears more  
 1431 times in  $c_{\downarrow}^{2k}(\sigma)$  (without loss of generality) than in  $c_{\downarrow}^{2k}(\tau)$ . For any path  $\delta$  and  $\lambda$ , define

$$1432 \quad A(\delta) = \{\{(c_\phi^{2k} = c_0, c_\delta^{2k} = c_1) | \phi \in \mathcal{C}(\delta)\}\} \quad (12)$$

$$1433 \quad C_\lambda = \{\{|A(\delta)| | \delta \in \mathcal{B}(\lambda)\}\} \quad (13)$$

1434 Then we have

$$1435 \quad C_\sigma = \{\{|A(\delta)| | \delta \in \mathcal{B}(\sigma)\}\} = \{\{|(c_\phi^{2k} = c_0, c_\delta^{2k} = c_1) | \delta \in \phi \cap \sigma\}\} \quad (14)$$

$$1436 \quad C_\tau = \{\{|A(\delta)| | \delta \in \mathcal{B}(\tau)\}\} = \{\{|(c_\phi^{2k} = c_0, c_\delta^{2k} = c_1) | \delta \in \phi \cap \tau\}\} \quad (15)$$

1437 So  $C_\sigma \neq C_\tau$ .

1438 Considering the path coloring  $d(\delta) = |A(\delta)|$ . For two  $n$ -paths  $\delta_1, \delta_2$ , if  $d(\delta_1) \neq d(\delta_2)$ ,  
 1439 we can assume that  $|A(\delta_1)| > |A(\delta_2)|$  without loss of generality, then the number of up-  
 1440 per adjacent neighbors of  $\delta_1$  and  $\delta_2$  up to color pair  $(c_0, c_1)$  are different, which means  
 1441  $c_{\uparrow}^{2k}(\delta_1) \neq c_{\uparrow}^{2k}(\delta_2)$ . So  $c_{\delta_1}^{2k+1} \neq c_{\delta_2}^{2k+1}$ , which means  $c^{2k+1} \not\sqsubseteq d$ .

1442 Applying Lemma D.11 to  $\mathcal{B}(\sigma)$  and  $\mathcal{B}(\tau)$ , we have

$$1443 \quad \{\{c_{\delta_1}^{2k+1} | \delta_1 \in \mathcal{B}(\sigma)\}\} \neq \{\{c_{\delta_2}^{2k+1} | \delta_2 \in \mathcal{B}(\tau)\}\} \quad (16)$$

1444 The above multi-sets are exactly the color multi-sets of the faces of  $\sigma$  and  $\tau$ , which means  
 1445  $c_{\mathcal{B}}^{2k+1}(\sigma) \neq c_{\mathcal{B}}^{2k+1}(\tau)$ . Consequently,  $c_\sigma^{2k+2} \neq c_\tau^{2k+2}$ , which contradicts with the induction  
 1446 hypothesis, so  $c_{\downarrow}^{2k}(\sigma) = c_{\downarrow}^{2k}(\tau)$ .

1447 From the induction hypothesis, we have  $b_\sigma^k = b_\tau^k$ ,  $b_{\mathcal{B}}^k(\sigma) = b_{\mathcal{B}}^k(\tau)$ ,  $b_{\uparrow}^k(\sigma) = b_{\uparrow}^k(\tau)$ ,  $b_{\downarrow}^k(\sigma) =$   
 1448  $b_{\downarrow}^k(\tau)$ , so  $b_\sigma^{k+1} = b_\tau^{k+1}$ .

1458

□

1459

1460

1461

*Proof of Theorem D.7.* Given a path complex  $P$ , let  $a^t$  be the coloring of the vertices of  $P$  at iteration  $t$  of WL and  $b^t$  be the coloring of the same vertices at iteration  $t$  of PWL. We firstly prove that  $b^t \sqsubseteq a^t$ , then give a pair of graphs to show that they cannot be differentiated by WL but can be differentiated by PWL.

1462

1463

1464

1465

1466

1467

1468

1469

1470

1471

1472

1473

1474

1475

1476

1477

1478

1479

1480

1481

1482

1483

1484

1485

1486

1487

1488

1489

1490

1491

1492

1493

1494

1495

1496

1497

1498

1499

1500

1501

1502

1503

1504

1505

1506

1507

1508

1509

1510

1511

1.  $b^t \sqsubseteq a^t$ . We do this by induction. The base case holds because all vertices are initialized with the same color. Suppose the result holds for  $t = k$ , we prove that  $b^{k+1} \sqsubseteq a^{k+1}$ . Let  $v$  and  $w$  be two vertices of two arbitrary path complexes  $P_1, P_2$ , suppose  $b_v^{k+1} = b_w^{k+1}$ , we prove that  $a_v^{k+1} = a_w^{k+1}$ .

Note that vertices only has upper adjacent neighbors, so we have  $b_v^k = b_w^k, b_{\uparrow}^k(v) = b_{\uparrow}^k(w)$ . The second equation means

$$\{\{b_x^k | (b_x^k, -) \in b_{\uparrow}^k(v)\}\} = \{\{b_y^k | (b_y^k, -) \in b_{\uparrow}^k(w)\}\}$$

This can be equivalently written as

$$\{\{b_x^k | x \in \mathcal{N}_{\uparrow}(v)\}\} = \{\{b_y^k | y \in \mathcal{N}_{\uparrow}(w)\}\}$$

From the induction hypothesis, we have  $a_v^k = a_w^k$  and

$$\{\{a_x^k | x \in \mathcal{N}_{\uparrow}(v)\}\} = \{\{a_y^k | y \in \mathcal{N}_{\uparrow}(w)\}\}$$

These are the arguments of the hash function for WL to compute the colors of  $v$  and  $w$  in the next iteration, so  $a_v^{k+1} = a_w^{k+1}$ .

2. Considering the graphs in Figure 8, they cannot be differentiated by WL test. In PWL test, the path complex derived from the right graph has not any 3-path while the derived path complex from the left graph has 3-paths.

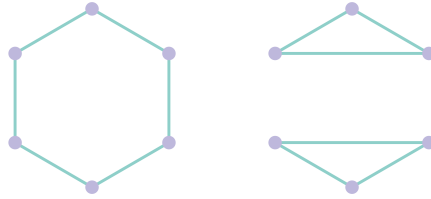


Figure 8: Two graphs that cannot be distinguished by WL but can be differentiated by PWL.

1497

1498

1499

1500

1501

1502

1503

1504

1505

1506

1507

1508

1509

1510

1511

*Proof of Theorem D.8.* Considering the graphs in Figure 9, they cannot be differentiated by SWL test. In PWL test, the path complex derived from the right graph has not any 4-path while the derived path complex from the left graph has 4-paths. □

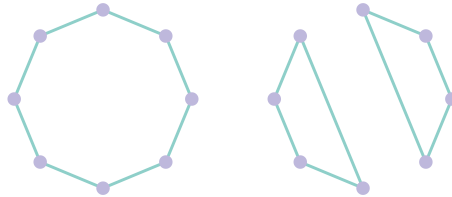


Figure 9: Two graphs that cannot be distinguished by SWL but can be differentiated by PWL.

1512 *Proof of Theorem D.9.* Let  $b^t$  and  $d^t$  be the coloring at iteration  $t$  of PWL and the  $t$ -th layer of an  
 1513 PCMP respectively. Assume the PCMP has  $L$  layers and assume  $d^t = d^L$  ( $t > L$ ). We use induction  
 1514 to prove that  $d^t \sqsubseteq b^t$ . The base case holds by definition. Suppose the result holds for  $t = k$ , when  
 1515  $t = k + 1$ , we prove that  $d^{k+1} \sqsubseteq b^{k+1}$ . For any two  $n$ -paths  $\sigma, \tau$  of any two path complexes  $P_1, P_2$   
 1516 such that  $d_\sigma^{k+1} = d_\tau^{k+1}$ , we prove that  $b_\sigma^{k+1} = b_\tau^{k+1}$ .

1517 The condition means all the update, aggregate and message functions are injective, so their compo-  
 1518 sition is also injective. Hence  $d_\sigma^k = d_\tau^k, d_{\mathcal{B}}^k(\sigma) = d_{\mathcal{B}}^k(\tau), d_\uparrow^k(\sigma) = d_\uparrow^k(\tau)$ .

1520  $d_{\mathcal{B}}^k(\sigma) = d_{\mathcal{B}}^k(\tau)$  means

$$\{\{d_s^k | s \in \mathcal{B}(\sigma)\}\} = \{\{d_t^k | t \in \mathcal{B}(\tau)\}\}$$

1522  $d_\uparrow^k(\sigma) = d_\uparrow^k(\tau)$  means

$$\{\{(d_s^k, d_{s \cup \sigma}^k) | s \in \mathcal{N}_\uparrow(\sigma)\}\} = \{\{(d_t^k, d_{t \cup \tau}^k) | t \in \mathcal{N}_\uparrow(\tau)\}\}$$

1526 By the induction hypothesis, we have  $b_\sigma^k = b_\tau^k$ .

$$\{\{b_s^k | s \in \mathcal{B}(\sigma)\}\} = \{\{b_t^k | t \in \mathcal{B}(\tau)\}\}$$

$$\{\{(b_s^k, b_{s \cup \sigma}^k) | s \in \mathcal{N}_\uparrow(\sigma)\}\} = \{\{(b_t^k, b_{t \cup \tau}^k) | t \in \mathcal{N}_\uparrow(\tau)\}\}$$

1531 So  $b_\sigma^k = b_\tau^k, b_{\mathcal{B}}^k(\sigma) = b_{\mathcal{B}}^k(\tau), b_\uparrow^k(\sigma) = b_\uparrow^k(\tau)$ , these are the arguments of the hash function in PWL,  
 1532 so  $b_\sigma^{k+1} = b_\tau^{k+1}$ .

1535  $\square$

1536

1537

1538

1539

1540

1541

1542

1543

1544

1545

1546

1547

1548

1549

1550

1551

1552

1553

1554

1555

1556

1557

1558

1559

1560

1561

1562

1563

1564

1565