

GIANT: Protein-Ligand Binding Affinity Prediction via Geometry-aware Interactive Graph Neural Network

Shuangli Li, Jingbo Zhou, Tong Xu, Liang Huang, Fan Wang,
Haoyi Xiong, Weili Huang, Dejing Dou, Hui Xiong

Abstract—

Drug discovery often relies on the successful prediction of protein-ligand binding affinity. Recent advances have shown great promise in applying graph neural networks (GNNs) for better affinity prediction by learning the representations of protein-ligand complexes. However, existing solutions usually treat protein-ligand complexes as topological graph data, thus the 3D geometry-based biomolecular structural information is not fully utilized. The essential intermolecular interactions with long-range dependencies, including type-wise interactions and molecule-wise interactions, are also neglected in GNN models. To this end, we propose a geometry-aware interactive graph neural network (GIANT) which consists of two components: 3D geometric graph learning network (3DG-NET) and pairwise interactive learning network (PI-NET). Specifically, 3DG-NET iteratively performs the node-edge interaction process to update embeddings of nodes and edges in a unified framework while preserving the 3D geometric factors among atoms, including spatial distance, polar angle and dihedral angle information in 3D space. Moreover, PI-NET is adopted to incorporate both element type-level and molecule-level interactions. Specially, interactive edges are gathered with a subsequent reconstruction loss to reflect the global type-level interactions. Meanwhile, a pairwise attentive pooling scheme is designed to identify the critical interactive atoms for complex representation learning from a semantic view. An exhaustive experimental study on two benchmarks verifies the superiority of GIANT.

Index Terms—Binding affinity prediction, graph neural network, geometry modeling, drug discovery, compound-protein interaction.



1 INTRODUCTION

The prediction of protein-ligand binding affinity has been widely considered as one of the most important tasks in computational drug discovery [1]. Here ligands are usually drug candidates including small molecules and biologics which can interact with proteins as agonists or inhibitors in the biological processes to cure diseases. Given a protein, we are interested in understanding how well a drug molecule (called a ligand) can interact with this protein. The strength of interaction between them can be quantified as a numerical score (called the binding affinity), which potentially

determines whether a ligand can have an effective influence on the protein (for example, to inactivate a protein to cure a disease). Therefore, the calculation of binding affinity is of great significance, and our target is to estimate this valuable interaction score. Although it can be measured by experimental methods, those biological tests are laborious and time-consuming. Thus, data-driven computational approaches have become increasingly necessary and achieved remarkable success in various drug discovery applications, including protein interaction mining [2], molecule generation [3], and drug reactions prediction [4], which highlight the efficacy of such methods in tackling complex problems for drug-based data mining and knowledge discovery. With similar data-driven learning models, binding affinities can be predicted in the early stage of drug discovery. Instead of applying costly biological methods directly to screen numerous candidate molecules, the prediction of binding affinity can help to rank drug candidates and prioritize the appropriate ones for subsequent testing to accelerate the process of drug screening [5].

With the development of structural biology and protein structure prediction [6], especially the recent Alphafold II model [7], there are growing three-dimensional (3D) structure protein data, which enables a new paradigm for structure-based drug discovery [8]–[10]. It has been demonstrated that 3D structural information can effectively contribute to the drug design [11]. Indeed, since there are already many accurate and robust algorithms to find poses of protein-ligand complexes (e.g., binding site prediction methods and docking methods), it is significant to focus on

- Shuangli Li is with Anhui Province Key Lab of Big Data Analysis and Application, School of Computer Science and Technology, University of Science and Technology of China and Business Intelligence Lab, Baidu Research. E-mail: lsl1997@mail.ustc.edu.cn. This work was done when he was an intern in Baidu Research under the supervision of Jingbo Zhou.
- Jingbo Zhou is with Business Intelligence Lab, Baidu Research. E-mail: zhoujingbo@baidu.com.
- Tong Xu is with Anhui Province Key Lab of Big Data Analysis and Application, School of Computer Science, University of Science and Technology of China. E-mail: tongxu@ustc.edu.cn.
- Liang Huang is with Oregon State University. E-mail: liang.huang.sh@gmail.com.
- Fan Wang and Haoyi Xiong are with Baidu Inc. E-mail: {wangfan04,xionghaoyi}@baidu.com.
- Weili Huang is with HWL Consulting LLC. E-mail: lwily99@gmail.com.
- Dejing Dou is with Business Intelligence Lab, Baidu Research. E-mail: dejingdou@gmail.com. His current affiliation is BCG X.
- Hui Xiong is with Artificial Intelligence Thrust, The Hong Kong University of Science and Technology (Guangzhou). E-mail: xionghui@ust.hk
- Jingbo Zhou, Dejing Dou and Hui Xiong are the corresponding authors.

Manuscript received April XX, 20XX; revised August XX, 20XX.

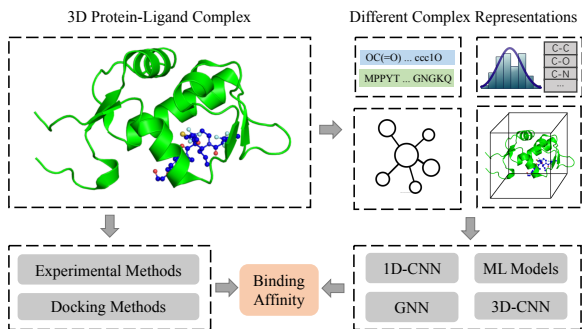


Fig. 1. A brief summary for protein-ligand binding affinity prediction. (1) Top left: An example of protein-ligand complex (Structure ID in Protein Data Bank (PDB): 5HMI). (2) Top right: Various complex representations. (3) Bottom left: Traditional Methods. (4) Bottom right: Machine learning and deep learning methods.

the much harder task of binding affinity prediction [12]. To learn useful 3D structure from a protein-ligand complex, as illustrated in Figure 1, many efforts have been devoted to estimating more accurate binding affinity for effective drug design. Docking methods [13]–[15] play an important role to predict how a specific ligand binds to the target protein with affordable computational costs. While the docking process can identify the binding pose of the protein-ligand complex with relatively high accuracy, its prediction of binding affinity is inaccurate and unreliable due to poor scoring functions [12], [16], which limits the applicability of docking methods in drug discovery. Compared to docking calculations, traditional machine learning methods [12], [17] have improved the performance by learning the extracted features from protein-ligand complexes. However, these approaches with limited generalizability require expert knowledge and heavily rely on feature engineering.

Recently, deep learning for binding affinity prediction has become an emerging research area, which represents the complex as sequence data [18], [19], 3D grid-like data [20] or graph data [21] to employ various neural networks. One of the key challenges of deep learning in structural biology is how to model the 3D spatial structure for better performance. To this end, most of the existing works [20], [22], [23] attempt to apply 3D convolutional neural networks (3D CNNs) by treating the complex as a 3D-grid representation. However, the cost of these models is huge, especially when considering long-range structural interactions. In addition, both the absence of topological information and the sensitivity to rotation in the complex have a negative effect on the prediction results.

Despite the powerful ability of graph neural networks (GNNs) to learn graph representations [24], there are only a few studies [21], [25] using GNNs to predict the protein-ligand binding affinity. By contrast, many researchers have developed GNN models in other fields of drug discovery [26], [27], such as predicting molecular property [28]–[31], biological network linking [32], and chemical reactions [33]. Nevertheless, these domain-specific models tend to lose their effectiveness when modeling the larger biomolecules, e.g., protein-ligand complexes. In general, most of the existing GNNs in drug design aim to learn the spatial structure by incorporating the distance information, which is insufficient to model the 3D geometric structure of complex. Moreover, the fundamental pairwise interactive information

between proteins and ligands, which is valuable for predicting the binding affinity [34], cannot be handled under the current GNN framework.

This paper is an extension of our preliminary work [35]. To overcome the above limitations, in this paper, we propose a novel Geometry-aware Interactive Graph Neural Network (GIANT) to learn the constructed complex interaction graph for protein-ligand binding affinity prediction. GIANT is equipped with two components to correspondingly address the challenges, namely the *3D geometric graph learning network* (3DG-NET) for modeling the geometric structure in 3D space and the *pairwise interaction learning network* (PI-NET) for leveraging both element type-level and molecule-level interactions with long-range intermolecular dependencies.

As the first part of GIANT, the key idea of 3DG-NET is illustrated in Figure 2 which aims to construct the spherical space for each central target and to apply the node-edge interactive scheme iteratively. 3DG-NET has the ability to preserve spatial distance, polar angle and dihedral angle information of neighbors when performing the aggregation process, thus it can effectively learn the 3D structure of protein-ligand complexes.

PI-NET is designed as the secondary part of GIANT to incorporate global intermolecular interactions. On the one hand, in view of the large size of the protein, it is redundant to contain the complete protein structure in the graph and we construct the spatial-based interaction graph from the central key structure of complex, but in this way the type-level long-range interactive information (including distant solvation effects [36] and electrostatic interactions [34]) between the protein and the ligand cannot be captured through such complicated graph without the complete structure. To deal with this issue, we employ an atomic type-aware pooling process on edges by introducing an auxiliary learning task to reconstruct the interactive matrix for type-level interaction injection. On the other hand, several important atoms of the complex can affect pairwise interactions and contribute to the binding affinity. Therefore, we finally utilize the molecule-level attentive pooling network to extract the informative biological semantics.

By means of 3DG-NET and PI-NET from two perspectives, our proposed GIANT can enhance the representation learning for complexes with involving both 3D geometric structures and global interactions. To summarize, the main contributions of this paper are as follows:

- To the best of our knowledge, we are among the first to develop graph neural networks from the perspective of comprehensive biochemical representation learning in 3D space for structure-based binding affinity prediction.
- We propose a novel geometry-aware interactive graph neural network (GIANT), which can capture not only 3D geometric information through distance-aware graph attention and angle-oriented graph convolution with a triangular fusion scheme in 3DG-NET, but also global long-range interactions through pairwise interaction learning network (PI-NET) in a semi-supervised manner.
- We conduct extensive experiments using two benchmark datasets to evaluate the performance of the proposed model, which demonstrates the superiority of GIANT compared with state-of-the-art baselines.

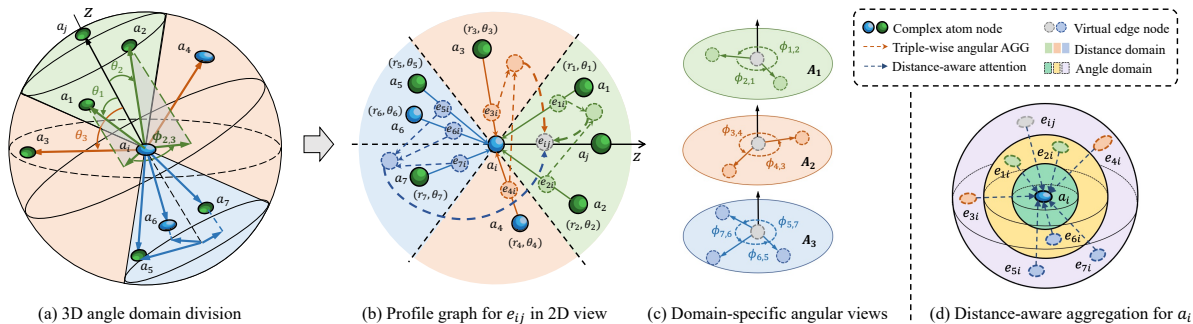


Fig. 2. An illustration of complex geometric division with three angle domains (in different colors) in 3D space, where θ_i represents polar angle and $\phi_{i,j}$ denotes dihedral angle between two adjacent planes.

Compared with our previous conference paper SIGN [35], the major improvements include: 1) For geometry modeling, we present a new sufficient paradigm to model the 3D view for protein-ligand complexes and replace the angle-oriented graph attention with a triple-wise dihedral graph aggregation process (TAGG) to enhance the structure learning. We fulfil the 3DG-NET for integrated complex modeling with complementary dihedral angle information. In this way, the proposed GIANT can learn the comprehensive 3D geometry instead of partial geometry in SIGN. 2) For interaction modeling, we devise a novel pairwise interaction learning network PI-NET to further facilitate the complex representation learning with adding the molecule-level interaction component, which can capture the interactive correlations of both biological element types and high-level molecules. 3) We significantly extend our experimental evaluation by comparing with our primary work [35] and showing additional quantitative results for model effectiveness and parameter analysis. 4) We also provide a case study to analyze the interpretability of our model in understanding the protein-ligand interaction patterns.

2 RELATED WORK

In this section, we first review the related literatures about predicting protein-ligand binding affinity and then detail recent advances in graph neural networks for drug discovery.

Protein-Ligand Binding Affinity Prediction. As a crucial stage in drug discovery, predicting protein-ligand binding affinity has been intensively studied for a long time [37], [38], which is of great importance for efficient and accurate drug screening. The earlier empirical-based methods [14], [39], [40] design docking and scoring functions specially to make predictions, while expert domain knowledge is required to encode internal biochemical interactions. Later on, statistical and machine learning-based methods [41] are developed to predict binding affinity based on data-driven learning, which attempt to extract protein-ligand features and use classic models for regression, such as random forest [12] and SVM [17]. These approaches are dependent on the quality of hand-crafted features and lack of generality on the larger dataset. Recently, several deep learning-based models [18], [19] utilize 1D convolutions and pooling to capture potential patterns from raw sequence information of both ligands and proteins. However, only using separate character representations fails to achieve desirable performance.

Recently, AlphaFold II [7] makes a remarkable achievement in the field of protein structure prediction, which adopts the Transformer-based framework designed for predicting protein’s 3D structure given the amino acid sequence of the protein. As the increasing availability of 3D-structure protein-ligand data [42], there is another hot research area of studying structure-based approaches, which focus on learning from 3D-structure protein-ligand complexes to predict binding affinity. The problem of AlphaFold II and the binding affinity prediction problem are two complementary problems, both of which hold great importance for biological data mining and drug knowledge discovery. Some recent works [22], [23] represent the protein-ligand complex as 3D grid-like data and use 3D convolutions (3D-CNNs) to take advantage of spatially-local correlations. Though these approaches can learn spatial information, one limitation is that positions of proteins and ligands in different complexes are changeable, such as different angle rotations, which means the spatial structure of 3D grid-like modeling is inevitably incomplete. More recently, OnionNet [43] employs CNN models to learn the complex representation from the extracted element-specific interaction features between a protein and its ligand. However, all the above models neglect the critical topological structure information of complexes. In the work [25], a protein-ligand complex is represented as a weighted graph with distance information. Then graph attention networks are applied to predicting the interactions. Nevertheless, only distance information between atoms is not adequate to model 3D-structure interactions. In this paper, we also focus on the structure-based prediction of protein-ligand binding affinity with incorporating abundant spatial information.

Graph Neural Networks for Drug Discovery. Inspired by the great advantage of graph neural networks (GNNs) in modeling graph data, more attention has been devoted to applying them in computational drug discovery [26], such as the prediction of molecular property [44] and protein interface [45]. Treating the molecule as a graph, GNNs can learn the graph-level representation for drug or protein by aggregating structural information. GraphDTA [21] adopts GNN models [46]–[48] to learn drug presentation with combining the protein representation from 1D convolutions to predict binding affinity. In attributed molecular graphs, the edges between atoms contain valuable information, such as distance or bond order. To leverage rich attributes in the molecule, edge-oriented message passing neural networks [28], [49], [50] are proposed to update both node and edge

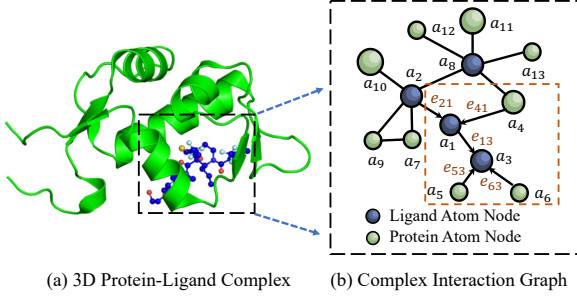


Fig. 3. An illustrative example of converting the protein-ligand complex into a complex interaction graph.

TABLE 1
Mathematical notations.

Notation	Description
$\mathcal{V}^P, \mathcal{V}^L$	The atom node sets of protein and ligand
M^P, M^L	The 3D position matrices of protein and ligand
$\mathcal{G}_I$	The complex interaction graph
a_i	The i -th atom node in $\mathcal{G}_I$
e_{ij}	The directed edge from atom a_i to atom a_j
$\mathcal{N}_e(a_i), \mathcal{N}_e(e_{ij})$	The neighboring edges of atom a_i and edge e_{ij}
$\mathbf{a}_i, \mathbf{e}_{ij}$	The embedding vectors of atom a_i and edge e_{ij}
$\mathbf{d}_{ij}$	The spatial embedding vector between a_i and a_j
Q	The number of angle domains
θ_{kij}	The polar angle between e_{ki} and e_{ij}
$\phi_{k_1 k_2}$	The dihedral angle between $e_{k_1 i}$ and $e_{k_2 i}$
$\mathbf{h}_P, \mathbf{h}_L$	The graph embeddings of protein and ligand

embeddings. Meanwhile, there are also some efforts to model the 3D-structure of molecule by improving GNNs with spatial information, such as distance [25], [29], angle [30], [51], and 3D coordinate [52]. However, these models fail to consider the spatial interactions between proteins and ligands. In addition, the function of learning angle information in [30] is designed for density functional theory, which is only beneficial for predicting molecular properties rather than protein-ligand binding affinity. Moreover, recently there are long-range interaction learning GNN models, while they are designed for specific applications (such as user-item interaction in social recommendation [53], [54]) and only focus on node-wise interactions [55]. To overcome these limitations, we propose an multi-level interaction-aware GNN framework with integrating both distance and angle factors harmoniously.

3 PRELIMINARIES

In this section, we introduce some definitions used in our model and formulate the structure-based prediction problem for protein-ligand binding affinity. The frequently used key notations in this paper are summarized in Table 1.

Definition 1. Complex Interaction Graph. Given a protein-ligand complex as shown in Figure 3(a), we define the atom node sets of protein and ligand as $\mathcal{V}^P = \{a_1^P, \dots, a_m^P\}$ and $\mathcal{V}^L = \{a_1^L, \dots, a_n^L\}$ with the position matrix $M^P \in \mathbb{R}^{m \times 3}$ and $M^L \in \mathbb{R}^{n \times 3}$ for 3D atomic coordinates, respectively. Then we define the complex interaction graph as a directional graph $\mathcal{G}_I = \langle \mathcal{V}, \mathcal{E} \rangle$, where the vertex set $\mathcal{V}$ is a subset of atom node sets of protein and ligand, i.e. $\mathcal{V} \subseteq \mathcal{V}^P \cup \mathcal{V}^L$ and the unweighted edge set $\mathcal{E} = f_e(\mathcal{V}^P, \mathcal{V}^L, M^P, M^L)$ is constructed based on the spatial positions of atoms in the complex. Specifically, except the $\mathcal{V}^L$, the protein's atoms

close to the ligand from $\mathcal{V}^P$ are selected to add into $\mathcal{V}$. We then update the complex edge set $\mathcal{E}$ by adding into the edges of atom pairs whose distances are shorter than the cutoff threshold θ_d . The distance between atom nodes is calculated using the Euclidean distance, which is a widely employed distance metric that measures the straight-line spatial distance between two points in three-dimensional space. By applying the Euclidean distance calculation, denoted as $d_{ij} = \sqrt{(M_i^L - M_j^P)^2}$, we can precisely quantify the spatial separation between atom nodes. Formally, the edge set is represented as $\mathcal{E} = \{(a_i, a_j) | a_i, a_j \in \mathcal{V}, s.t. d_{ij} \leq r_\theta\}$. The detailed construction process is described in Algorithm 1.

Algorithm 1: Graph Construction Process.

Input : The position matrix M^P and node set $\mathcal{V}^P$
The position matrix M^L and node set $\mathcal{V}^L$
The cutoff distance r_θ

Output: The graph $\mathcal{G}_I = \langle \mathcal{V}, \mathcal{E} \rangle$

```

1 Initialize  $\mathcal{V} \leftarrow \mathcal{V}^L, \mathcal{E} \leftarrow \{\}$ ;
2 for atom node pair  $(a_i, a_j) \in \mathcal{V}^L \times \mathcal{V}^P$  do
3   Calculate distance  $d_{ij} \leftarrow |M^L(a_i) - M^P(a_j)|$ ;
4   if  $d_{ij} \leq r_\theta$  then
5     Update node set  $\mathcal{V} \leftarrow \mathcal{V} \cup \{a_j\}$ ;
6   end
7 end
8 Combined position matrix
   $M \leftarrow \text{CONCAT}(M^L, M^P)$ ;
9 for atom node pair  $(a_i, a_j) \in \mathcal{V} \times \mathcal{V}$  do
10  Calculate distance  $d_{ij} \leftarrow |M(a_i) - M(a_j)|$ ;
11  if  $d_{ij} \leq r_\theta$  then
12    Update edge set  $\mathcal{E} \leftarrow \mathcal{E} \cup \{e_{ij} = (a_i, a_j)\}$ ;
13  end
14 end
15 return  $\mathcal{V}, \mathcal{E}$ 
```

Definition 2. Edge-oriented Neighbors. Given an atom node a_i or a directed edge e_{ij} (i.e., $a_i \rightarrow a_j$) in the complex interaction graph $\mathcal{G}_I$, the edge-oriented neighbors $\mathcal{N}_e$ of a_i or e_{ij} are defined as the sets of directed edges $\{e_{ki}, \dots, e_{li}\}$ which point to the target atom a_i or the target edge e_{ij} .

Taking Figure 3(b) as an example, the edges e_{21} and e_{41} are connected to the edge e_{13} via the common node a_1 , the edge-oriented neighbors of e_{13} are denoted as $\mathcal{N}_e(e_{13}) = \{e_{21}, e_{41}\}$. Similarly, the edges e_{13} , e_{53} and e_{63} point to the atom node a_3 , resulting in the neighbors set $\mathcal{N}_e(a_3) = \{e_{13}, e_{53}, e_{63}\}$.

Problem 1.

Structure-based Protein-Ligand Binding Affinity Prediction. Given a protein-ligand complex with 3D structure, i.e., the complex interaction graph $\mathcal{G}_I$ and the 3D position matrix M , our goal is to learn a regression model $f(\mathcal{G}_I, M)$ to precisely predict the numerical binding affinity score, which represents the strength of interaction. Then we can discover drug molecules (ligands) with higher scores, which reflect stronger interactions between proteins and ligands.

4 MODEL FRAMEWORK

In this section, we present the proposed GIANT model for protein-ligand binding affinity prediction. We first introduce the overall framework and then describe the details of each component for geometric and interactive modeling.

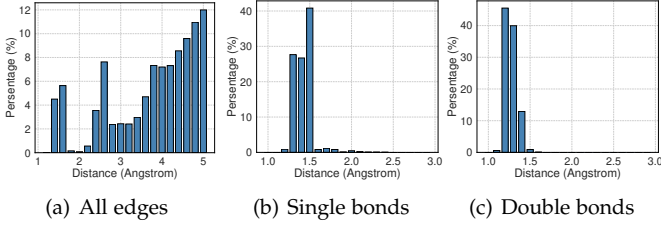


Fig. 4. The distribution of distances between atoms within 5 Å in the protein-ligand complex from PDBbind dataset.

4.1 Overview

Figure 5 exhibits the architecture which takes the complex interaction graph $\mathcal{G}_I$ as an input. We start with the 3D geometric graph learning network (3DG-NET), which is composed of *node*→*edge* and *edge*→*node* interaction layers. 3DG-NET can propagate the node’s and edge’s embeddings alternately with learning the geometric distance and angle information in the 3D space. The two parts of 3DG-NET play a synergistic effect on modeling the spatial structure of the complex. After that, we apply the pairwise interaction learning network (PI-NET) which explores the interactive correlations in both type-level and molecule-level views. The type-level interaction layer performs on the edges’ representations to obtain the atomic type-based interaction matrix of the complex, while the molecule-level interaction layer can learn the interactive pattern across the semantic structures of proteins and ligands to generate the ultimate representation. From a global view, PI-NET aims to approximate the overall interactions between proteins and ligands to improve the prediction performance. Finally, the model is trained through multi-task learning with augmented constraints for the interaction matrix and the spatial distribution, which serves as additional self-supervised tasks.

4.2 3D Geometric Graph Learning Network

Standard GNNs have shown great advantages in learning topological structure of the general graph, which cannot take atom’s spatial position into account in the 3-dimensional space. To model the 3D structure of a complex, an intuitive method is to provide atom’s 3-dimensional coordinate in the GNN architecture [52]. However, the position information under the Cartesian coordinate system is sensitive to both translations and rotations, causing poor generalization of models when learning the complex representation. Several models, such as GNN-DTI [30] and MAT [29], manage to combine the distance information in the aggregation process, while only pairwise distance is not adequate. Different from DimeNet [30], which specially design spherical Bessel functions in GNN for density functional theory (DFT) approximation with limited ability to model the larger biological complex, we propose to employ iterative geometry-aware *node*→*edge* and *edge*→*node* interaction layers to incorporate both distance and angle information from a spatial distribution perspective.

4.2.1 The Architecture of 3DG-NET

Inspired by the spherical coordinate which is composed of radial distance r , polar angle θ and dihedral angle ϕ , we develop an interaction-based graph neural network to leverage both the distance between nodes and the angles

between edges in a collaborative framework. As illustrated in Figure 2, when aggregating for edge e_{ij} , we treat it as the zenith direction z . To easily understand the geometric distribution in the original 3D space, we show the 2D view profile graph along the direction of z in Figure 2(b). To keep the original 3D geometric factors, in the profile graph of 2D view, we preserve the real polar angle and radial distance which are calculated in the 3D space for each edge. The edges in the 2D view graph can be seen as the rotated edges from the 3D space to the 2D plane with the same polar angles. Under such a definite polar coordinate system in 2D view graph, the edge-oriented neighbors are distributed around e_{ij} with unique identifying coordinates (r, θ) . Through the method of dividing angle domains in 3D space, the spatial distribution for the complex can be taken into account. In each angle domain, the dihedral angle ϕ is formed by two adjacent planes and we can obtain the domain-specific angular views in Figure 2(c). To integrate the dihedral angle and further learn the comprehensive 3D geometry, we develop the 3DG-NET by means of angle-oriented interaction scheme for edges for protein-ligand structure understanding in our framework. Specifically, the triple-wise angular aggregation (TAGG) module first locally incorporates projection angles in each domain, then all domain-level representations are combined via the global aggregation stage.

Moreover, the distance factor is also helpful for structure modeling, which reveals spatial correlations. Figure 4 shows the statistical distribution of distance between atoms. It can be seen that the distances of covalent bonds mainly range from 1 to 2 Å, while noncovalent interactions, like hydrophobic and van der Waals interactions, and hydrogen bonds, are distributed over longer distances. Atomic interactions in the complex vary from different distances, which indicate different spatial relations for atom pairs. Given the radial distance r between atoms a_i and a_j , as shown in Figure 2(d), we first map r to a bucket (i.e., a distance domain corresponding to a type of relation) and obtain the one-hot vector $\mathbf{x}_{ij}$. Then we apply a dense layer transformation to get the spatial relation embedding:

$$\mathbf{d}_{ij} = \mathbf{W}_s \mathbf{x}_{ij}, \quad (1)$$

where $\mathbf{W}_s \in \mathbb{R}^{d_s \times b}$ is the transformation weight matrix and b is the number of buckets (i.e., spatial relations). To factor in these correlations, we design the distance-aware attention in the second interaction stage for nodes. As shown in Figure 5(a), the overall geometry-aware interaction process at l -th layer is defined as:

$$\begin{aligned} \mathbf{e}_{ij}^{(l)} &= f^{(l)}\left(\{\mathbf{a}_k^{(l-1)}, \mathbf{a}_i^{(l-1)}, \forall e_{ki} \in \mathcal{N}_e(e_{ij})\}\right), \\ \mathbf{a}_j^{(l)} &= g^{(l)}\left(\mathbf{a}_j^{(l-1)}, \{\mathbf{e}_{kj}^{(l)}, \forall e_{kj} \in \mathcal{N}_e(a_j)\}\right), \end{aligned} \quad (2)$$

where $\mathbf{e}_{ij}^{(l)}$ is the edge embedding, $\mathbf{a}_j^{(l)}$ is the node (atom) embedding, $\mathcal{N}_e(e_{ij})$ and $\mathcal{N}_e(a_j)$ are the edge-oriented neighbors of edge e_{ij} and node a_j respectively, and $f(\cdot)$ and $g(\cdot)$ are interaction functions of *node*→*edge* and *edge*→*node* layers which are introduced as follows.

4.2.2 Angle-oriented Node→Edge Interaction Layer

Failing to distinguish neighbor nodes from different directions in the aggregation process is a weakness of the existing

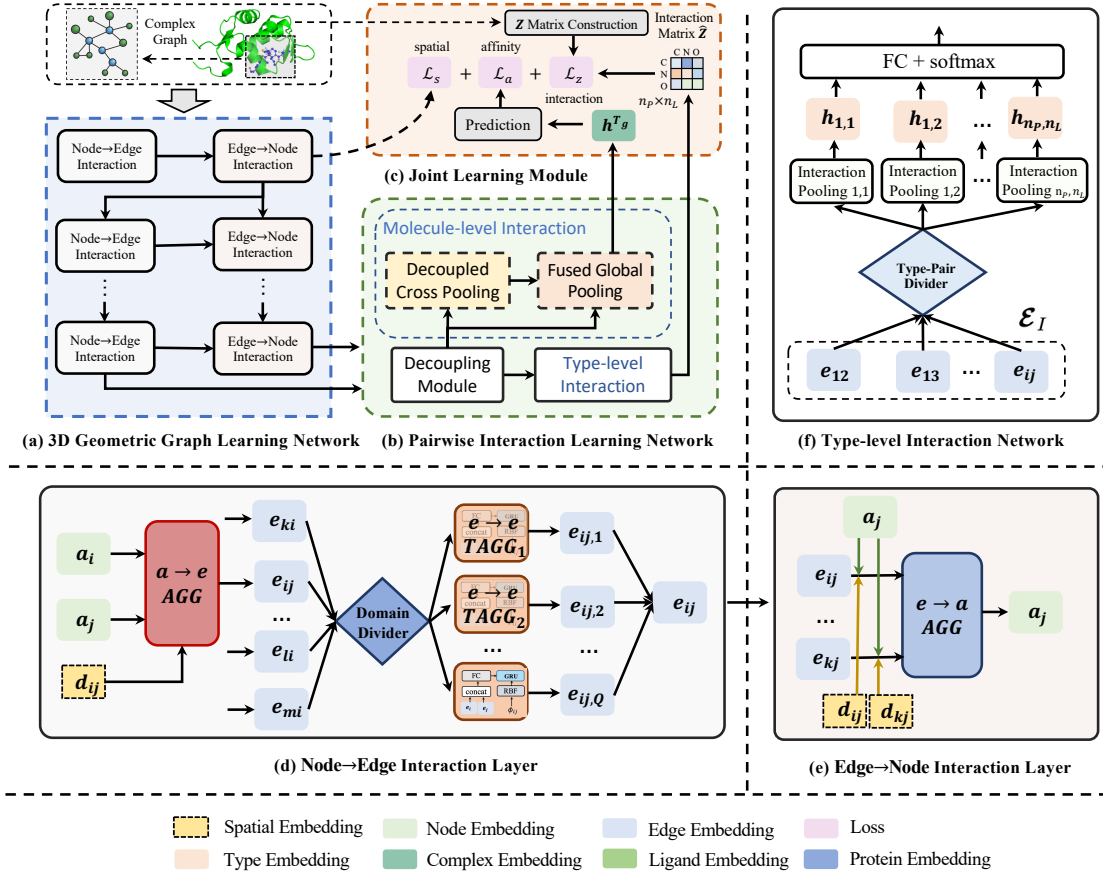


Fig. 5. Illustration of the proposed GIANT framework. (a), (b): The two key components 3DG-NET and PI-NET. (c): The joint multi-task learning module for model optimization. (d), (e): The two inner structures of component 3DG-NET. (f): The type-level interaction layer in the PI-NET.

GNN models. To overcome this inadequacy, we adopt an angle-oriented graph learning layer to update the edge representations with integrating abundant geometric angle information. Since the angle exists between the two edges, as shown in Figure 5(d), we first get the edge embedding through aggregating the node features:

$$e_{ij}^{(l)} = \sigma(\mathbf{W}_{a \rightarrow e}^{(l)} \cdot [a_i^{(l-1)} \parallel a_j^{(l-1)} \parallel d_{ij}]), \quad (3)$$

where $\mathbf{W}_{a \rightarrow e}^{(l)}$ is the transformation matrix for atomic combination, the operator $\parallel$ represents concatenation, and σ is the Relu function.

After obtaining the representations $\{e_{ij}^{(l)}, e_{ki}^{(l)}, \dots, e_{mi}^{(l)}\}$ of edge e_{ij} and its neighbors, we further separate the neighboring edges in 3-dimensional space by applying an angle-domain divider $\mathbf{D}_A$, which plays an intermediate role to assign each neighbor to the specific angle domain. For example, in Figure 2(b), there are seven edge-oriented neighbors $e_{1i}, e_{2i}, \dots$, and e_{7i} around the central target edge e_{ij} . These neighboring edges are located in three different local angle domains according to the polar angles between edge e_{ij} and its neighbors. Given the number of angle domains Q (e.g., $Q = 3$ in Figure 2) and the target edge e_{ij} for aggregation, $\mathbf{D}_A$ can map each neighbor e_{ki} to the located angle domain index:

$$Ind_{ki} = \mathbf{D}_A(e_{ki}, e_{ij}, Q) = \begin{cases} \lceil Q \cdot \frac{\theta_{kij}}{180^\circ} \rceil, & \theta_{kij} \leq 180^\circ \\ \lceil Q \cdot \frac{360^\circ - \theta_{kij}}{180^\circ} \rceil, & \theta_{kij} > 180^\circ \end{cases} \quad (4)$$

where $\lceil \cdot \rceil$ denotes rounding operation to get the integer index, $\theta_{kij} \in [0, 360^\circ]$ is the calculated polar angle between

edges e_{ki} and e_{ij} . Then the subset of edge-oriented neighbors located in the q -th angle domain can be defined as:

$$\mathcal{N}_e^q(e_{ij}) = \{e_{ki} \mid e_{ki} \in \mathcal{N}_e(e_{ij}) \wedge Ind_{ki} = q\}. \quad (5)$$

After reorganizing the neighbors of e_{ij} through divider $\mathbf{D}_A$ based on the projected profile in 2D view, we then feed all neighbor subsets from different angle domains into Q independent propagation layers to capture high-order dependencies in the complex interaction graph. As shown in Figure 2(a), the spatial area scope of each angle domain forms a spherical cone, where exists several local neighboring edges. The spatial distribution in 3D space is synergistically determined by polar angle θ and dihedral angle ϕ . While the separated angle domains can indicate the geometric information of polar angle, the other geometric factor, dihedral angle, should also be perceived by the edge-level interactive stage.

To incorporate the significant dihedral angle information into the GNN, we further devise a triple-wise local aggregation (TAGG) process in each angle domain. As illustrated in Figure 2(b) and Figure 2(c), all neighboring edges in one angle domain are projected into a two-dimensional plane, which is called azimuthal pattern, to incarnate the local distribution from the aspect of projection angle. To calculate the consecutive angles between edges, we specify the clockwise direction and obtain the dihedral angle $\phi_{k_1 i k_2}$ between two adjacent edges $e_{k_1 i}$ and $e_{k_2 i}$ in the q -th angle domain A_q . We define the update function for the q -th

azimuthal pattern as:

$$\begin{aligned} \mathbf{m}_{ij,q}^{(l)} &= \sum_{\substack{e_{k_1 i} \in \mathcal{N}_e^g(e_{ij}) \\ e_{k_2 i} \in \mathcal{N}_e^g(e_{ij})}} f_{\text{tri},q}(\mathbf{e}_{k_1 i}^{(l)}, \phi_{k_1 i k_2}, \mathbf{e}_{k_2 i}^{(l)}) \cdot f_{\text{adj}}(e_{k_1 i}, e_{k_2 i}), \\ \mathbf{e}_{ij,q}^{(l)} &= \mathbf{M}_q^{\text{update}}(\mathbf{M}_q^{\text{fe}}(\mathbf{e}_{ij}^{(l-1)}), \mathbf{m}_{ij,q}^{(l)}), \quad 1 \leq q \leq Q, \end{aligned} \quad (6)$$

where $\mathbf{m}_{ij,q}^{(l)}$ is the q -th local aggregated edge representation at l -th layer, $f_{\text{tri},q}$ represents the triple-wise angle learning function introduced later, the boolean function $f_{\text{adj}}(e_{k_1 i}, e_{k_2 i})$ returns 1 only if $e_{k_1 i}$ is adjacent to $e_{k_2 i}$ along the clockwise direction, $\mathbf{M}_q^{\text{update}}$ and $\mathbf{M}_q^{\text{fe}}$ refer to dense layers for edge transformation in each angle domain. Moreover, we first apply the RBF (Radial Basis Function) kernel which has been proved to be advantageous for geometry encoding [56] to convert the scalar dihedral angle $\phi_{k_1 i k_2}$ to the geometric feature:

$$\phi_{k_1 i k_2} = \text{RBF}(\phi_{k_1 i k_2}) = \frac{K}{k=1} \exp(-\beta(\phi_{k_1 i k_2} - \mu_k)^2), \quad (7)$$

where $\smile$ means the concatenation operator over the scalar angle value to form a K -dimensional geometric representation. The K central points $\{\mu_k\}$ are uniformly picked out between 0 and 2π , while β is $(\frac{2\pi}{K})^{-2}$. Under the assumption of equal distribution [56], RBF learning process can encode dihedral angles based on the fine-grained spatial splitting with learnable parameters and then map the geometry into comprehensive vectors. Since the two adjacent edges constitute the angle, we adopt the pairwise edge encoder to extract the angle context representation $\tilde{\mathbf{e}}_{k_1 i k_2}^{(l)}$:

$$\tilde{\mathbf{e}}_{k_1 i k_2,q}^{(l)} = \mathbf{M}_q^p(\mathbf{M}_q^{\text{fe}}(\mathbf{e}_{k_1 i}^{(l)}) \parallel \mathbf{M}_q^{\text{fe}}(\mathbf{e}_{k_2 i}^{(l)})), \quad (8)$$

The dense layers $\mathbf{M}_q^{\text{fe}}$ can project edge representations to the same space. $\mathbf{M}_q^p$ are utilized to combine the pairwise information. Then we propose a triangular fusion scheme to integrate informative dihedral angles continuously:

$$f_{\text{tri},q}(\mathbf{e}_{k_1 i}^{(l)}, \phi_{k_1 i k_2}, \mathbf{e}_{k_2 i}^{(l)}) = \text{GRU}(\tilde{\mathbf{e}}_{k_1 i k_2,q}^{(l)}, \mathbf{W}_{\phi,q}^{(l)} \phi_{k_1 i k_2}), \quad (9)$$

where $\mathbf{W}_{\phi,q}$ is the trainable parameter matrix for geometric transformation in q -th angle domain. GRU represents the Gated Recurrent Unit for learning spatial order-aware dependency, which assembles the pairwise edge context feature and the geometric information of dihedral angle to comprehend the spatial distribution in 3D space.

Finally we combine all aggregated edge embeddings obtained from Eq. (6). To completely preserve the geometric information in different local angle domains, we concatenate the representations as the global aggregation to update the angle-aware edge embedding:

$$\mathbf{e}_{ij}^{(l)} = [\mathbf{e}_{ij,1}^{(l)} \parallel \mathbf{e}_{ij,2}^{(l)} \parallel \cdots \parallel \mathbf{e}_{ij,Q}^{(l)}]. \quad (10)$$

4.2.3 Distance-aware Edge→Node Interaction Layer

After injecting the angle information into the edge embedding $\mathbf{e}_{ij}^{(l)}$, we make further efforts to develop an attention-based edge→node interaction layer to incorporate another geometric factor in the spherical coordinate system, that is distance. Specifically, since edges and nodes (atoms) have

different feature spaces, we first convert the edge embedding and node embedding into the hidden representation $\tilde{\mathbf{e}}_{ij}^{(l)}$ and $\tilde{\mathbf{a}}_j^{(l-1)}$ in the same vector space:

$$\begin{aligned} \tilde{\mathbf{e}}_{ij}^{(l)} &= \mathbf{W}_e^{(l)} \cdot \mathbf{e}_{ij}^{(l)}, \\ \tilde{\mathbf{a}}_j^{(l)} &= \mathbf{W}_a^{(l)} \cdot \mathbf{a}_j^{(l-1)}, \end{aligned} \quad (11)$$

where $\mathbf{W}_e^{(l)}$ and $\mathbf{W}_a^{(l)}$ are linear transformation matrices, $\mathbf{a}_j^{(l-1)}$ is the embedding of atom a_j from $(l-1)$ -th layer.

As a result of the variant distances and atomic attributes, the neighboring edges have different impacts on the target node. However, the existing GNN models cannot effectively capture the influence of the distance factor. Hence, as shown in Figure 2(d) and 5(e), we propose to extend the original GAT [47] with the distance-aware attention to fuse the distance information with the capability of discriminating multiple spatial relations among atoms:

$$\begin{aligned} w_{ij}^{(l)} &= \text{LeakyRelu}(\mathbf{v}_l^T \cdot [\tilde{\mathbf{e}}_{ij}^{(l)} \parallel \tilde{\mathbf{a}}_j^{(l)} \parallel \mathbf{W}_d^{(l)} \mathbf{d}_{ij}]), \\ \beta_{ij}^{(l)} &= \frac{\exp(w_{ij}^{(l)})}{\sum_{e_{tj} \in \mathcal{N}_e(a_j)} \exp(w_{tj}^{(l)})}, \end{aligned} \quad (12)$$

where $\mathbf{v}_l$ is the trainable parameter of edge→node attention at l -th layer, $\mathbf{W}_d^{(l)}$ is the trainable parameter matrix for distance transformation, the final calculated attention weight $\beta_{ij}^{(l)}$ reflects how important the edge e_{ij} is for the node a_j . Then we develop the distance-aware attention to multi-head attention version as GAT for better stability and apply the aggregation process from edge to node:

$$\mathbf{a}_j^{(l)} = \frac{1}{C} \sum_{c=1}^C \sum_{e_{ij} \in \mathcal{N}_e(a_j)} \beta_{ij,c}^{(l)} \cdot \tilde{\mathbf{e}}_{ij,c}^{(l)}, \quad (13)$$

where C is the number of independent attention heads. Due to the angle injection for edge embedding $\tilde{\mathbf{e}}_{ij,c}^{(l)}$ and the distance injection for attention weight $\beta_{ij,c}^{(l)}$, our proposed model can comprehensively incorporate multiple geometric information in the complex.

After performing L geometric interaction layers, finally we obtain the node embedding $\mathbf{a}_j = \mathbf{a}_j^{(L)}$ for atom a_j and the edge embedding $\mathbf{e}_{ij} = \mathbf{e}_{ij}^{(L)}$ between atoms a_i and a_j .

4.3 Pairwise Interaction Learning Network

The binding affinity between the protein and the ligand is heavily dependent on multiple cross-molecular interactions, which include atom-level, element type-level and molecule-level correlations. A pair of influential protein-ligand atomic nodes can be disconnected and distant from each other (e.g., over 10 angstroms). Since interactive atomic pairs may not reach to each other through a multi-hop message passing process in a topological view, stacking many GNN layers may still fail to capture such distant correlations even if it possibly suffers from the over-smooth problem [57]. As a result, only node-wise interaction is not adequate without high-level correlations. As presented in Figure 6, our model aims to capture multi-level correlations with following a hierarchical interaction learning scheme: node-wise interaction $\rightarrow$ type-wise interaction $\rightarrow$ molecule-wise interaction.

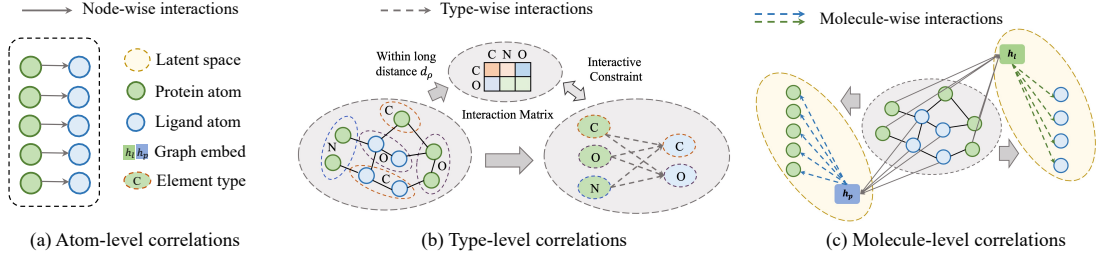


Fig. 6. The illustration of three different correlations in the protein-ligand complex.

Despite the design of 3DG-NET for atom-level interactions, the global inter-relationship between two molecules is still not taken into consideration. Specifically, while 3DG-NET treats the protein-ligand complex as a geometric-enhanced graph and focuses on aggregating local spatial information for node-wise interaction, this component fails to capture the type-wise and molecule-wise interaction. As illustrated in Figure 6(b) and 6(c), type embedding and molecular embedding are essential inputs to capture high-level correlations. However, the 3DG-NET can not utilize these inputs, which motivates us to design PI-NET to handle global protein-ligand correlations at the level of element types and molecules. To tackle this problem, we manage to develop the PI-NET to thoroughly learn the representation for the interaction graph, which can enhance the global protein-ligand interactions from two perspectives: long-range dependence modeling with type-wise interaction and important atom identification with molecule-wise interaction.

From the view of type-wise interaction, the correlations exist between element type-based aggregations for the protein and the ligand. We obtain the type embeddings and construct the interactive matrix with type-wise interactions, which can be optimized by the first module of PI-NET (Type-level Pairwise Interactive Constraint). From the view of molecule-wise interaction, the second module (Molecule-level Pairwise Interaction Modeling) is adopted to take advantage of molecular embeddings and identify the interactive atoms for the binding affinity.

4.3.1 Type-level Pairwise Interactive Constraint

The complete long-range intermolecular interactions at the type level between protein and ligand have effects on the binding affinity [12], [34], while $\mathcal{G}_I$ cannot provide such type-level interactive information. To capture the element type-wise long-range interactions in the complex (e.g., the Carbon-Carbon co-occurrence interaction), we further input the obtained intermolecular edge representation set $\mathcal{E}_I$ as well as the interaction matrix $\mathbf{Z}$ constructed from the additional complete protein structure and introduce an additional type-level pairwise interactive constraint via self-supervised training.

Specifically, we first construct the pairwise interaction matrix $\mathbf{Z} \in \mathbb{R}^{|S_P| \times |S_L|}$ from the complete protein and its ligand, where S_P and S_L are atomic type sets of the protein and its ligand. Each element T_k in S_P or T_l in S_L represents the atomic number (e.g., 6) of a certain atom (e.g., carbon atom C). Following the previous work [12], we calculate the number of occurrences for a specific atomic type pair (T_k, T_l) (e.g., (6, 7) for <C,N> pair) within a certain distance

and normalize the result to get the matrix $\mathbf{Z}$:

$$n(T_k, T_l) = \sum_{a_i \in \mathcal{V}^P} \sum_{a_j \in \mathcal{V}^L} \delta(\tau(a_i), T_k) \delta(\tau(a_j), T_l) \Theta(d_p - d_{ij}),$$

$$\mathbf{Z}_{kl} = \frac{n(T_k, T_l)}{\sum_{(a_i, a_j) \in \mathcal{V}^P \times \mathcal{V}^L} \Theta(d_p - d_{ij})}, \quad (14)$$

where the function $\tau(a_i)$ returns the atomic type of a_i , $\delta(\cdot, \cdot)$ is a Kronecker delta function which outputs 1 only if the type of atom is T_k (or T_l) and 0 otherwise, d_p is referred to as the interaction cutoff distance and a Heaviside step function Θ is adopted to count protein-ligand atomic type pairs within the distance d_p .

Since the complex interaction graph is composed of a protein molecule and a ligand molecule, we aim to leverage the global intermolecular interactions which focus on the high-level type pairwise relationships. We use the atomic type-specific aggregation for node representations which can integrate the chemical attribute knowledge and structural context at an appropriate granularity. Specifically, we take the edge embeddings obtained from 3DG-NET as input to the atomic type-aware pooling layer, which is shown in Figure 5(b). There are $|S_P| \times |S_L|$ pooling blocks for type pairs. One block to gather edge representations belonging to atomic type pair (T_k, T_l) can be formulated as:

$$\mathbf{h}_{k,l} = \sum_{e_{ij} \in \mathcal{E}_I} \underbrace{\delta(\tau(a_i), T_k) \delta(\tau(a_j), T_l)}_{\text{Divider}} \mathbf{W}_h \mathbf{e}_{ij}^{(L)}, \quad (15)$$

where $\mathbf{W}_h$ is the shared parameter matrix for edge pooling, $\mathcal{E}_I$ contains all representations of the intermolecular edges in the complex $\mathcal{G}_I$, a_i and a_j are atom nodes connected by e_{ij} , the two $\delta(\cdot, \cdot)$ functions act as a divider to pick up the corresponding edges. Then we calculate each value of the approximate interaction matrix:

$$\tilde{\mathbf{Z}}_{kl} = \frac{\exp(\mathbf{q}^T \mathbf{h}_{k,l})}{\sum_{i,j} \exp(\mathbf{q}^T \mathbf{h}_{i,j})}, \quad (16)$$

where $\mathbf{q}$ is the trainable parameter. In the training stage, we use an additional proximity loss to draw the interaction matrix $\tilde{\mathbf{Z}}$ and $\mathbf{Z}$ closer:

$$\mathcal{L}_z = \sum_{\mathcal{G}_I \in \mathcal{D}} \|F(\tilde{\mathbf{Z}}) - F(\mathbf{Z})\|, \quad (17)$$

where $F(\cdot)$ is the flatten operation for matrix, $\mathcal{D}$ is the training set.

4.3.2 Molecule-level Pairwise Interaction Modeling

After obtaining the representations of node (atom) and edge (bond) in the complex interaction graph $\mathcal{G}_I$, we select out

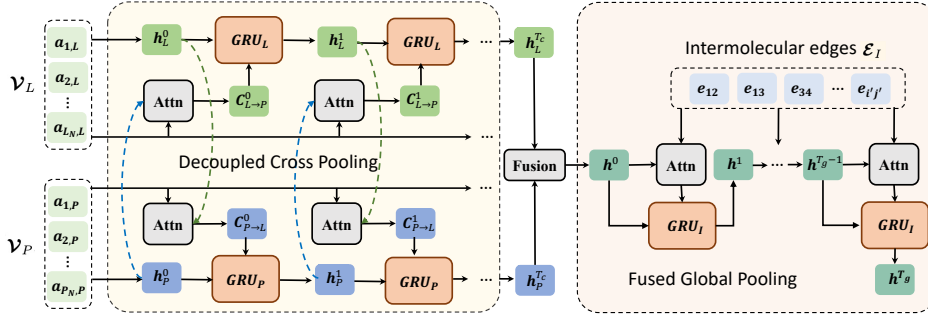


Fig. 7. The architecture of the molecule-level interaction network.

three representation sets (protein node set $\mathcal{V}_P$, ligand node set $\mathcal{V}_L$ and intermolecular edge set $\mathcal{E}_I$) with re-indexing through the decoupling module:

$$\begin{aligned}\mathcal{V}_P &= \{a_{i',P} \mid a_i \in \mathcal{V}^P\}, \\ \mathcal{V}_L &= \{a_{i',L} \mid a_i \in \mathcal{V}^L\}, \\ \mathcal{E}_I &= \{e_{i'j'} \mid a_i \in \mathcal{V}^P, a_j \in \mathcal{V}^L\}.\end{aligned}\quad (18)$$

Then the decoupled representations are fed into the interaction-aware attentive pooling scheme. We leverage the GRU as the basic block for molecule-level interaction learning, which can integrate pairwise molecular embeddings with the semantic dependence. Different from AttentiveFP [33] which learns on the small molecular graph and fails to handle the associated double molecular structures, the novel PI-NET aims to aggregate the interactive information from the other molecule (protein or ligand) into each graph embedding for pairwise characteristics learning.

Decoupled Cross Pooling. As shown in Figure 7, we first sum the atomic vectors of each molecule and get the initial molecule-level representations h_P^0 for protein and h_L^0 for ligand. Then the crossed graph attentive pooling layers with communicative channels for protein and ligand are employed to learn the interactive context with fusing the updated molecular embedding. For the protein channel, the aggregation step with cross attention mechanism at the t -th layer can be formalized as follows:

$$\begin{aligned}\alpha_{i,P}^{t-1} &= \text{LeakyRelu}((q_P^{t-1})^T \cdot [h_L^{t-1} \parallel a_{i,P}]), \\ C_{L \rightarrow P}^{t-1} &= \sum_{a_{i,P} \in \mathcal{V}_P} \text{softmax}(\alpha_{i,P}^{t-1}) \mathbf{W}_P^{t-1} a_{i,P},\end{aligned}\quad (19)$$

where q_P^{t-1} is the trainable parameter to calculate the cross attention weight $\alpha_{i,P}^{t-1}$ with combining the ligand embedding in an interactive manner. $C_{L \rightarrow P}^{t-1}$ represents the aggregated context vector, which can figure out the influential protein atoms for the ligand molecule to guide the interaction. Since the decoupled pooling process is symmetric for the protein and the ligand, the context vector $C_{P \rightarrow L}^{t-1}$ from protein to ligand can be generated through the similar cross attention layer. Then we employ the GRU block to update the decoupled embeddings for the protein and the ligand:

$$\begin{aligned}h_P^t &= \text{GRU}(h_P^{t-1}, C_{L \rightarrow P}^{t-1}), \\ h_L^t &= \text{GRU}(h_L^{t-1}, C_{P \rightarrow L}^{t-1}),\end{aligned}\quad (20)$$

Fused Global Pooling. After the derivation of protein embedding $h_P^{T_c}$ and ligand embedding $h_L^{T_c}$ with T_c stacked cross pooling layers, we further introduce the fused global

pooling layer to better capture the protein-ligand interactions. We first get the integrated complex representation h^0 through the weighting fusion:

$$h^0 = \alpha h_L^{T_c} + (1 - \alpha) h_P^{T_c} \quad (21)$$

where $\alpha \in [0, 1]$ measures the factor between two molecules when contributing to the fused comprehensive representation. Considering the situation that the binding affinity derives from the complicated effects of the interactive edges between the protein and the ligand, we propose to apply the attentive global pooling to aggregate the significant edge embeddings, some of which carry the crucial reactive information. The pooling mechanism over intermolecular edges is given as:

$$\begin{aligned}\gamma_{ij}^{t-1} &= \text{LeakyRelu}((q_I^{t-1})^T \cdot [h^{t-1} \parallel e_{ij}]), \\ C_{inter}^{t-1} &= \sum_{e_{ij} \in \mathcal{E}_I} \text{softmax}(\gamma_{ij}^{t-1}) \mathbf{W}_I^{t-1} e_{ij}, \\ h^t &= \text{GRU}(h^{t-1}, C_{inter}^{t-1}), \quad t = 0, 1, \dots, T_g\end{aligned}\quad (22)$$

where q_I^{t-1} and $\mathbf{W}_I^{t-1}$ are learnable parameters at the t -th global pooling layer. The attention coefficient γ_{ij}^{t-1} reveals how important the intermolecular edge e_{ij} is to the affinity. The GRU block is adopted to determine how much information aggregated from abundant edge embeddings to be leveraged and how much information of the fused molecular embedding to be reserved. Benefiting from the global interactive attention mechanism between protein and ligand, the influential atomic pairs which greatly contribute to the binding interactions can be picked out to provide valuable explanations for the prediction result.

4.4 Optimization Objective

In the last part, we get the final complex representation h^{T_g} after the interaction network PI-NET and use MLP layers as the regressor to predict the protein-ligand binding affinity:

$$\hat{y} = \text{MLP}(h^{T_g}). \quad (23)$$

Then the absolute error between the predicted binding affinity $\hat{y}$ and the measured ground truth y is used to calculate the loss. Thus, we adopt the L1 loss function to optimize the model:

$$\mathcal{L}_a = \sum_{g_I \in \mathcal{D}} |\hat{y} - y|, \quad (24)$$

where $\mathcal{D}$ contains all the protein-ligand complexes with binding affinities. Moreover, we introduce the graph regularization for spatial angle domains into the loss function,

which is in view of the assumption that adjacent angle domains in 3D space should have similar azimuthal transformation matrices in the aggregation process of 3DG-NET:

$$\mathcal{L}_s = \sum_{l=1}^L \sum_{q=1}^{Q-1} \|W_{\phi,q+1}^l - W_{\phi,q}^l\|^2, \quad (25)$$

To integrate the effectiveness of both interaction and geometry for better complex representation learning, we further combine with the complex interaction constraint in Eq. (17) as well as the spatial constraint from a geometric view, and we reach the following overall objective function:

$$\mathcal{L} = \mathcal{L}_a + \lambda_z \mathcal{L}_z + \lambda_s \mathcal{L}_s, \quad (26)$$

where λ_z and λ_s are balancing hyper-parameters to control the strength of interaction loss and geometric graph regularization respectively.

5 EXPERIMENTS

In this section, we conduct experiments on two standard datasets to investigate the following research questions:

- **RQ1.** How does the proposed GIANT model perform compared against the state-of-the-art methods?
- **RQ2.** How does the generalizability of GIANT when trained on the larger but lower-quality dataset?
- **RQ3.** Will the geometric and interactive factors be effective and benefit the prediction?
- **RQ4.** How do the parameter settings (e.g., cutoff distance and angle domain divisions) affect the prediction result?
- **RQ5.** How efficient is GIANT compared with the state-of-the-art models?
- **RQ6.** Can GIANT provides reasonable interpretations about the binding affinity from an interactive view?

5.1 Experiment Settings

5.1.1 Datasets

We evaluate all models on the following public standard datasets for protein-ligand binding affinity prediction.

PDBbind¹ is a well-known public dataset [42] in development which provides 3D binding structures of protein-ligand complexes with experimentally determined binding affinities. In our experiment, we mainly use the PDBbind v2016 dataset, which is most frequently used in recent works [23], [43]. Specifically, it includes three overlapping subsets, i.e., *general*, *refined* and *core set*. The *general set* contains all 13,283 protein-ligand complexes, while the 4,057 complexes in *refined set* are selected out of the *general set* with better quality. Moreover, the *core set* with 290 complexes serves as the highest quality benchmark for testing through a careful selection process [58]. Conveniently, we call the difference between the *refined* and *core* subsets, that is 3,767 complexes, as *refined set* of PDBbind in the following.

CSAR-HiQ² is an additional benchmark dataset [59], containing two subsets with 176 and 167 protein-ligand complexes. We use this external dataset from an independent source to further evaluate the generalization ability.

5.1.2 Setup

Following [12], we choose the *refined set* of PDBbind as our primary training data because there is considerable overlap between the full *general set* and CSAR-HiQ dataset. We randomly split the protein-ligand complexes in *refined set* with a ratio of 9:1 for training and validation. For testing sets, we use the *core set* and CSAR-HiQ set with removing the complexes present in *refined set*.

Since the lower-quality data of the *general set* can still improve the performance of models [60], we conduct the supplemental experiment on the full *general set* which is larger but of worse quality to analyze the generalizability of our model and baseline methods. As stated above, we can only evaluate the performance on the *core set* due to the overlapping problem of CSAR-HiQ dataset. Following [23], [43], We randomly select 1,000 complexes from the *refined set* as the validating set. The remaining 11,993 complexes in the *general set* are used for training.

5.1.3 Evaluation Metrics

To comprehensively evaluate the model performance, following [23], [43], we use Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Pearson’s correlation coefficient (R) and the standard deviation (SD) in regression to measure the prediction error. As introduced in [23], SD is defined as: $SD = \sqrt{\frac{1}{|\mathcal{D}|-1} \sum_{i=1}^{|\mathcal{D}|} [y_i - (a + b\hat{y}_i)]^2}$, where $\hat{y}_i$ and y_i respectively represent the predicted and experimental value of the i -th complex in dataset $\mathcal{D}$, and a and b are the intercept and the slope of the regression line, respectively.

5.1.4 Baselines

We compare our proposed model with comparative methods including machine learning-based methods (**LR**, **SVR**, and **RF-Score** [12]), CNN-based methods (**Pafnucy** [23] and **OnionNet** [43]), GNN models **GraphDTA** [21] for protein-ligand binding affinity prediction, and GNN-based global structure learning models (**GraphTrans** [61] and **NL-GCN** [55]). Moreover, various state-of-the-art GNN-based models which also consider the geometric information (e.g., distance or angle) for molecular modeling (**SGCN** [52], **DMPNN** [28], **MAT** [29], **DimeNet** [30], and **CMPNN** [49]) and protein-ligand structure learning (**GNN-DTI** [25] and **SIGN** [35]) are compared to evaluate the performance of GIANT. The details of baseline descriptions for **GraphTrans**, **NL-GCN** and **SIGN** are as follows, while other baselines are introduced in the original paper [35].

- **GraphTrans** [61] combines the Transformer-based self-attention with GNN model to learn long-range pairwise relationship, which can obtain the global graph embedding with topological structural information.
- **NL-GCN** [55] enables the non-local aggregation scheme with an attention-guided sorting mechanism for GNNs. A single calibration vector is utilized to redefine non-local neighborhoods for global structure learning.
- **SIGN** [35] is our preliminary state-of-the-art graph network model for the prediction of binding affinity. It is a basic variant of GIANT, which captures the distance and angle information as well as structural interactions.

1. <http://www.pdbbind-cn.org>

2. <http://www.csardock.org>

TABLE 2
Performance comparison on *PDBbind core set* and *CSAR-HiQ set* based on the *refined* training set.

Method		PDBbind core set				CSAR-HiQ set			
		RMSE ↓	MAE ↓	SD ↓	R ↑	RMSE ↓	MAE ↓	SD ↓	R ↑
ML-based Methods	LR	1.675 (0.000)	1.358 (0.000)	1.612 (0.000)	0.671 (0.000)	2.071 (0.000)	1.622 (0.000)	1.973 (0.000)	0.652 (0.000)
	SVR	1.555 (0.000)	1.264 (0.000)	1.493 (0.000)	0.727 (0.000)	1.995 (0.000)	1.553 (0.000)	1.911 (0.000)	0.679 (0.000)
	RF-Score	1.446 (0.008)	1.161 (0.007)	1.335 (0.010)	0.789 (0.003)	1.947 (0.012)	1.466 (0.009)	1.796 (0.020)	0.723 (0.007)
CNN-based Methods	Pafnucy	1.585 (0.013)	1.284 (0.021)	1.563 (0.022)	0.695 (0.011)	1.939 (0.103)	1.562 (0.094)	1.885 (0.071)	0.686 (0.027)
	OnionNet	1.407 (0.034)	1.078 (0.028)	1.391 (0.038)	0.768 (0.014)	1.927 (0.071)	1.471 (0.031)	1.877 (0.097)	0.690 (0.040)
GraphDTA Methods	GCN	1.735 (0.034)	1.343 (0.037)	1.719 (0.027)	0.613 (0.016)	2.324 (0.079)	1.732 (0.065)	2.302 (0.061)	0.464 (0.047)
	GAT	1.765 (0.026)	1.354 (0.033)	1.740 (0.027)	0.601 (0.016)	2.213 (0.053)	1.651 (0.061)	2.215 (0.050)	0.524 (0.032)
	GIN	1.640 (0.044)	1.261 (0.044)	1.621 (0.036)	0.667 (0.018)	2.158 (0.074)	1.624 (0.058)	2.156 (0.088)	0.558 (0.047)
	GAT-GCN	1.562 (0.022)	1.191 (0.016)	1.558 (0.018)	0.697 (0.008)	1.980 (0.055)	1.493 (0.046)	1.969 (0.057)	0.653 (0.026)
GNN-based Methods	SGCN	1.583 (0.033)	1.250 (0.036)	1.582 (0.320)	0.686 (0.015)	1.902 (0.063)	1.472 (0.067)	1.891 (0.077)	0.686 (0.030)
	GraphTrans	1.539 (0.044)	1.182 (0.046)	1.521 (0.042)	0.714 (0.019)	1.950 (0.072)	1.508 (0.069)	1.886 (0.083)	0.687 (0.033)
	NL-GCN	1.516 (0.019)	1.198 (0.013)	1.511 (0.024)	0.720 (0.010)	1.840 (0.024)	1.393 (0.016)	1.817 (0.028)	0.716 (0.011)
	GNN-DTI	1.492 (0.025)	1.192 (0.032)	1.471 (0.051)	0.736 (0.021)	1.972 (0.061)	1.547 (0.058)	1.834 (0.090)	0.709 (0.035)
	DMPNN	1.493 (0.016)	1.188 (0.009)	1.489 (0.014)	0.729 (0.006)	1.886 (0.026)	1.488 (0.054)	1.865 (0.035)	0.697 (0.013)
	MAT	1.457 (0.037)	1.154 (0.037)	1.445 (0.033)	0.747 (0.013)	1.879 (0.065)	1.435 (0.058)	1.816 (0.083)	0.715 (0.030)
	DimeNet	1.453 (0.027)	1.138 (0.026)	1.434 (0.023)	0.752 (0.010)	1.805 (0.036)	1.338 (0.026)	1.798 (0.027)	0.723 (0.010)
	CMPNN	1.408 (0.028)	1.117 (0.031)	1.399 (0.025)	0.765 (0.009)	1.839 (0.096)	1.411 (0.064)	1.767 (0.103)	0.730 (0.052)
	SIGN	1.316 (0.031)	1.027 (0.025)	1.312 (0.035)	0.797 (0.012)	1.735 (0.031)	1.327 (0.040)	1.709 (0.044)	0.754 (0.014)
Ours	GIANT	1.269 (0.020)	0.999 (0.018)	1.265 (0.024)	0.814 (0.008)	1.666 (0.024)	1.242 (0.030)	1.633 (0.034)	0.779 (0.011)

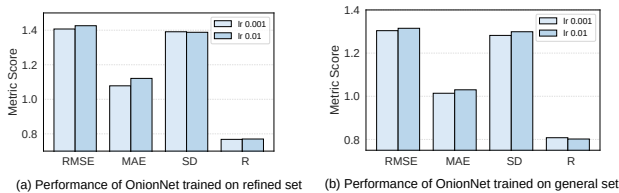


Fig. 8. Performance of OnionNet models with different learning rates.

5.1.5 Parameter Settings

For the proposed GIANT, we use Adam optimizer for model training with a learning rate of 0.001 and set the batch size as 32. The balancing hyper-parameters λ_z and λ_s are set to 1.75 and $1e-4$ respectively according to the performance on validation set. We construct the complex interaction graph and interaction matrix with cutoff-threshold $\theta_d = 5\text{\AA}$ and $\theta_p = 12\text{\AA}$ as suggested in [36], respectively. The basic dimensions of node and edge embeddings are both set to 128. The number of buckets for spatial relation b is set to 4 with the splitting granularity of $1\text{\AA}$. For 3DG-NET with two layers, we set the dimension K of the RBF kernel to 64, the number of attention heads C to 4, the dropout rate to 0.2, and the number of angle domains N to 6. For semantic interaction modeling in PI-NET, we set the number of decoupled cross pooling layers and fused global pooling layers as 2 and 1 respectively. The fusion parameter α is set to 0.7. For structure-level interaction layer in PI-NET, there are 36 pooling blocks in total, where the two atomic type sets S_P and S_L are defined as stated in [12]. We implement GIANT based on PaddlePaddle and the code is available³.

For baseline models, we tune the parameters of each method based on recommended settings in the paper to ensure the best performance. For our primary model SIGN [35], the hyper-parameters are the same as described in the original paper. For ML-based baselines, the number of decision trees in RF-score is set to 100, the max-depth of trees is set to 5, the maximum number of features is set to 3 and the minimum number of samples required to split is set to

10. For the CNN-based models, we set the channels of three-layer 3D convolutions for Pafnucy with the learning rate of $1e-5$ as 64, 128 and 256. For OnionNet, we experimented with different learning rates and found that a learning rate of 0.001 achieved the best performance in our experimental setup. As shown in Figure 8, instead of using the default learning rate (0.01) in the original paper [43], our tuned experiments show that using a learning rate of 0.001 can further improve OnionNet’s performance in our running environment with 1.3% and 3.8% improvements of RMSE and MAE when trained on refined set. Thus, we report the best performance achieved by a learning rate of 0.001 in our experimental setup. The number of input features is 3840 and there are 32, 64, and 128 filters in the three convolutional layers with the kernel size as 4. The maximum length of protein sequences is set to 1000 and the learning rate is set to $5e-4$ in GraphDTA. The number of graph learning layers for GCN, GAT, GCN-GAT and GIN are set to 3, 2, 2 and 5 respectively. For GNN-based models, the number of filters in three-layer SGCN is set to 32 with the dimension as 36 and the learning rate as $5e-4$. We also apply the data augmentation process to ensure optimal performance. For fair comparison, the embedding dimension of other baselines is set to 128 (same as GIANT). For GraphTrans with 5 GNN layers and 4 transformer layers, the learning rate is set to $1e-3$ with a cosine annealing schedule for learning rate decay. The kernel size of convolution in NL-GCN is 5, while the learning rate is fixed to $1e-3$. Note that we stack two layers for non-local graph convolution and apply the global summation readout with a two-layer MLP regressor for the binding affinity prediction task. GNN-DTI with the learning rate of $1e-4$ adopts the four-layer graph attention module, and the initial μ and δ for distance learning in GAT layers are set to 4.0 and 1.0, respectively. The weighting coefficients for self-attention, distance, and adjacency matrices in MAT with the learning rate of $5e-4$ are set to 0.3, 0.3, and 0.4, respectively. For DimeNet with the learning rate of $1e-3$, the number of spherical harmonics and radial basis functions are set to 4 and 3, respectively. We use two-layer interaction blocks and three-layer bilinear layers

3. https://github.com/PaddlePaddle/PaddleHelix/tree/dev/apps/drug_target_interaction/giant

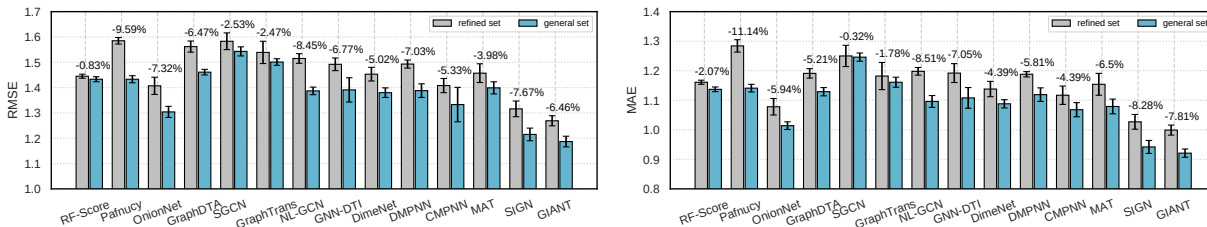


Fig. 9. Performance improvements on PDBbind benchmark when training on the *general* training set.

to make DimeNet work in our experiment. For DMPNN and CMPNN with the learning rate of $1e-3$, the number of edge-oriented message passing layers is set to 3 and we use MLP as the communication module in CMPNN.

5.2 Performance Comparison

5.2.1 Overall Comparison (RQ1)

We first compare our proposed GIAN with baseline approaches on two benchmark datasets. As shown in Table 2, the average and the standard deviation of four indicators for testing performance are reported across five random runs. In general, we can observe that our proposed SIGN and GIAN achieves the best performance on two datasets, especially GIAN is with 9.8% and 7.7% improvement of RMSE over the best baseline models on PDBbind and CSAR-HiQ datasets, respectively. We also observe the significant improvements by comparing GIAN with its variant SIGN owing to the designed 3D geometric learning and the additional semantic interaction modeling. We further have the following observations.

Among all baselines, GraphDTA methods show relatively poor performance due to the failure of considering the spatial structure and interactions between proteins and ligands. It indicates that simply modeling the molecular graph with protein sequence information is not capable of predicting structure-based protein-ligand binding affinity. By contrast, from the perspective of interaction modeling, the machine learning-based methods and OnionNet model take advantage of long-range interaction features and achieve better results. Please note that the original paper of OnionNet only conducted experiments on the larger general training set. As shown in Figure 8, we conducted comprehensive experiments under both refined and general training settings to provide a more thorough evaluation. In this experiment, we focus on comparing the performance of OnionNet trained on the refined set, while we will further compare and discuss the experimental results of OnionNet trained on the general set in the next Section 5.2.2. However, these data-driven approaches relying on feature engineering ignore the informative spatial structures of complexes and have limited generalization capability on the additional CSAR-HiQ dataset. For the global graph structure learning methods, GraphTrans and NL-GCN perform significantly better than classic GCN and GAT models due to the capability of capturing non-local graph topological information. Nevertheless, the lack of global type-wise and molecule-wise interaction leads to unsatisfying performance, and these methods perform much worse than GIAN. From the perspective of spatial structural modeling, we find that SGCN and GNN-DTI which incorporate position and distance information exhibit considerable improvement over the vanilla GCN

and GAT. Since SGCN takes atomic position coordinates as input directly, it will be easily affected by the rotation and translation of atoms, and the 3D CNN model Pafnucy suffers from a similar issue. Thus, the prediction results are not ideal. Despite leveraging a transformer-like attention mechanism to handle the spatial structure, MAT is not better than RF-Score and OnionNet, suggesting the importance of combining spatial and interactive information. The edge-oriented model CMPNN outperforms the above methods because it enhances DMPNN with communication while propagating the distance information, which shows the significance of node-edge message passing process. Although DimeNet can learn from angle information and perform slightly better, the performance is still not ideal due to its limited ability of modeling larger biomolecules. Our proposed GIAN can not only capture more comprehensive 3D angle-enhanced geometric information instead of just distance, but also handle both type-level and molecule-level interactions in the complex through attentive pooling and multi-task learning framework. Therefore, GIAN is much effective for modeling the protein-ligand complex and can accurately predict the binding affinity.

5.2.2 Generalizability Comparison (RQ2)

There is increasing 3D structure-based protein-ligand data with binding affinity, whereas the amount of high-quality data in *refined set* is relatively small. Thus, the ability of utilizing more lower-quality data to improve performance shows the generalizability of model, which is another necessary measurement of performance evaluation. As introduced in Section 5.1.2, we conduct the extra experiment of generalizability on the *general set* of PDBbind dataset. As illustrated in Figure 9, we compare the proposed GIAN with major competitive baselines on two training sets. The results show that GIAN gets the lowest prediction error remarkably under both training settings. More importantly, our model improves the performance by around 6.5% on RMSE and 7.8% on MAE when trained on the *general set* and it further expands the prediction advantage compared to baselines. Therefore, GIAN is proved to be more generalizable to more data in large quantity but poor quality.

5.3 Effectiveness of Geometric-Interactive Learning

To verify the effectiveness of factors that influence the final performance, we compare GIAN with its variants on the two benchmarks. Due to the page limit, we only report the performance on RMSE metric since the results on other metrics are also consistent with our analysis.

5.3.1 Impact of 3D Geometry Modeling (RQ3)

We first investigate the impact of geometric learning component. The geometry-reduced variants are as follows:

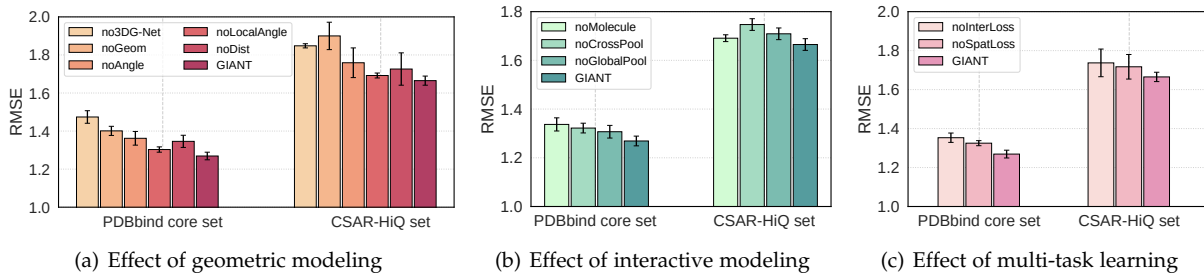


Fig. 10. Ablation study for geometry, interaction, and multi-task learning.

- **noDist** excludes the distance-aware graph attention.
- **noLocalAngle** excludes the triple-wise aggregation module without azimuthal angle information.
- **noAngle** excludes angle-oriented interaction component.
- **noGeom** excludes 3D geometry factors in 3DG-NET.
- **no3DG-Net** replaces 3DG-NET with a distance weighted GCN for protein-ligand structure learning.

Figure 10(a) presents the comparison results of geometry learning on RMSE. Specifically, noDist, noLocalAngle, and noAngle perform worse than GIANT since they can only capture the one-sided geometric structural information, i.e., distance or angle information in the complex. We can also observe that noLocalAngle achieves better performance than noAngle, demonstrating the importance of azimuthal angle factor for structure learning. The GCN-based variant no3DG-Net performs much worse than GIANT, which shows the general GNN model with distance information is not effective for complicated protein-ligand geometry learning. Furthermore, the prediction errors of noGeom and no3DG-Net are especially high among all variants. It indicates that modeling the complete geometric structure has a significant impact on performance improvement. Overall, our proposed 3DG-NET component of GIANT is effective for the geometry-based graph and greatly benefits the protein-ligand representation learning.

5.3.2 Impact of Interactive Modeling (RQ3)

In addition to geometric factors, we also conduct experiments to analyze the impact of pairwise interactive modeling. We study the performance of GIANT with the following settings on the interaction learning network PI-NET:

- **noCrossPool** excludes decoupled cross pooling process.
- **noGlobalPool** is lack of fused global pooling process.
- **noMolecule** removes molecule-level interaction module.

As we can see in Figure 10(b), GIANT outperforms all the variants removing either cross pooling or global pooling, verifying each pooling component has its contribution. The results also show that the lack of molecule-level interactions leads to performance reduction, which confirms that only utilizing the geometric factors is insufficient and will lose the important molecule-wise interactive information.

5.3.3 Impact of Multi-task Learning (RQ3)

To further validate the effectiveness of the multi-task learning architecture about the incorporation of geometry and interaction, we remove the additional loss for experiments:

- **noInterLoss** removes the interaction constraint $\mathcal{L}_z$.
- **noSpatLoss** removes the spatial regularization loss $\mathcal{L}_s$.

From 10(c), we can note that GIANT achieves better performance than noInterLoss and noSpatLoss on two datasets.

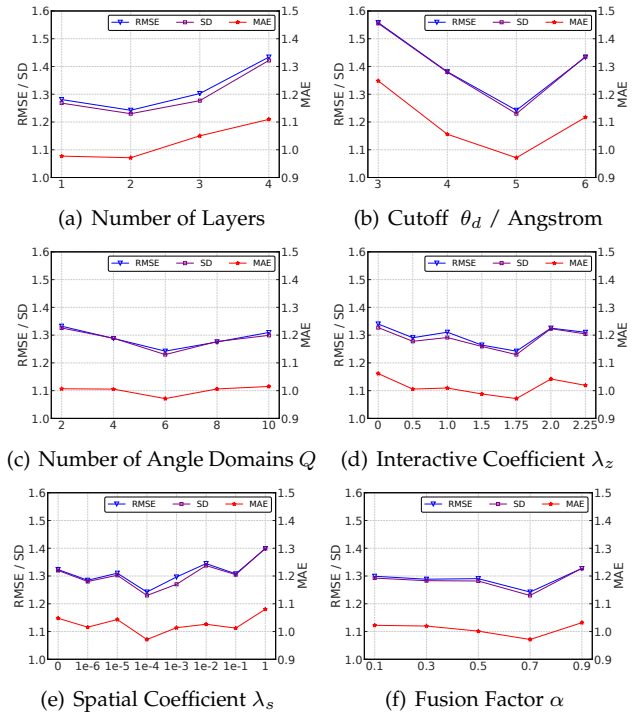


Fig. 11. Parameter sensitivity on PDBbind dataset.

This indicates the complementary supervised information from both 3D geometry learning and type-level interactions can boost the model training and contributes to the performance improvements.

As a whole, Figure 10 shows that the proposed GIANT significantly outperforms all other variants, proving the necessity of handling the geometric and interactive information synergistically which is essential for protein-ligand binding affinity prediction.

5.4 Parameters Analysis (RQ4)

As depicted in Figure 11, we investigate the performance variation for GIANT w.r.t several necessary hyperparameters by varying each parameter while keeping others fixed as default settings.

Number of 3DG-NET layers L . We first study the influence of multi-hop propagation with stacking node-edge interaction layers from 1 to 4. We observe that increasing the number of layers would not always give rise to a better result in Figure 11(a). The model with one 3DG-NET layer has limited ability to model high-order information in the complex. As a result of over-fitting, the performance of the model using more than two layers starts to degenerate gradually. Therefore, applying two interaction layers in GIANT is enough to capture sufficient spatial information.

TABLE 3
The efficiency studies on the PDBbind dataset.

Model	Training (s/epoch)	Inference (ms/sample)
Pafnucy	1637.13	11.91
OnionNet	13.54	1.25
DimeNet	30.61	3.23
CMPNN	118.48	30.34
SIGN	34.47	5.46
GIANT	47.43	7.45

Cutoff distance θ_d . As illustrated in Figure 11(b), we then analyze the effect of cutoff distance for complex interaction graph construction when varying θ_d from 3 to 6. With the increase of θ_d , more geometric information in the complex is available to our model and beneficial for learning complex representation better, which leads to dramatic performance improvements when $\theta_d \leq 5$ Å. After that, too long cutoff distance will introduce additional redundancies and degrade the performance.

Angle domain divisions Q . To look deeper into the impact of angle information in our model, we divide the angle domains varying from 2 to 10. The results are reported in Figure 11(c). We can see that the model performs best when the number of angle domains is 6. Too fine-grained or coarse-grained divisions will result in performance degradation. One possible explanation is that too fine-grained divisions cannot provide distinguishable information in 3D space while the angle domain at a too big granularity contains quite sparse atomic neighbors, both of which have an adverse effect on learning 3D geometric structures.

Interactive coefficient λ_z . Moreover, we change the coefficient λ_z to control the trade-off between the prediction loss and interaction loss. Figure 11(d) shows that the performance first tends to get better with incorporating more type-level interactive information for long-range dependencies, and then begins to drop off slightly. In general, our model is stable with varying coefficients and always achieves better performance than all baseline methods.

Spatial coefficient λ_s . We also perform experiments to evaluate the sensitivity of another balancing coefficient for spatial regularization. From the results in Figure 11(e), we observe that applying the spatial regularization can remarkably improve the prediction performance. Similarly, the appropriate spatial coefficient can strengthen the ability of learning the geometric information, while too much effort on this regularization can also challenge the model training and make the performance start dropping.

Fusion Factor α . Finally, we test the effect of the fusion factor α in Eq. (21), and vary it from 0.1 to 0.9. As shown in Figure 11(f), the performances increase first and then start to decrease. GIANT achieves the best performance when α is 0.7, which demonstrates the proposed model should pay more attention to the central ligand molecule with fewer atoms for better fused representation learning.

5.5 Efficiency Analysis (RQ5)

To evaluate the computation cost for efficiency analysis, we compare the proposed GIANT with several competitive CNN-based and GNN-based models in Table 3, which shows the training time per epoch (3390 samples) and the average inference time for one sample. It can be observed that Pafnucy runs much slower due to learning from the

3D structure with the time-consuming 3D-CNN. What’s more, the 2D-CNN model OnionNet is the most efficient method thanks to the preliminary feature extraction and reduces the model running time. However, the complicated pre-processing stage requires expert knowledge and the performance is not satisfying. The GNN methods DimeNet and CMPNN are less efficient since they apply the more effective message passing scheme to complex structures and need iterative computation for node-edge interactions. Note that our GIANT runs much faster than CMPNN and has the moderate computation cost with significant performance improvements (9.8% better than OnionNet and 12.7% better than DimeNet). Moreover, the average prediction time per protein-ligand complex is around 7.45ms and GIANT is slightly slower than the primary SIGN, which demonstrates the computational efficiency of the additionally designed 3D geometric structure learning and molecule-level interactive modeling. In summary, the experimental results show that our model achieves the best prediction accuracy with low computational cost, which is a promising method for real-world drug discovery applications in practice.

5.6 Interactive Visualization and Interpretation (RQ6)

Since our proposed GIANT can achieve state-of-the-art performance, it is quite profitable to further analyze the predictive interpretability for understanding protein-ligand interaction patterns. We manage to extract intermolecular attention coefficients which can reveal significant atomic pairs in semantic modeling layer to explain the derivation of the reactive binding affinity. The higher coefficient γ_{ij} indicates the more important atomic pair (a_i, a_j) (i.e., e_{ij}) for interactive contribution. As presented in Figure 12, we exhibit the top four most influential pairs and top four least influential pairs between the protein and the ligand to demonstrate the effective interpretability of GIANT. From the visualization result for 3D structure, we can observe that the proposed model is capable of identifying the meaningful interaction patterns, which are almost consistent with the knowledge-based interaction analysis by expert. Specifically, the oxygen atom in the ligand molecule are highly compatible with the SER-963 and GLU-966 residues and can promote the final binding affinity. What’s more, the GLU-957 residue and the nearby nitrogen atom can form the hydrogen bond and give a high contribution to the protein-ligand interactions. On the contrary, the LEU-881 and VAL-889 residues are less correlated to the ligand and have little impact on the interactive process. In brief, GIANT shows good agreement with the expert-level domain knowledge and can provide valuable suggestions to understand the complex interactions.

6 CONCLUSION AND FUTURE WORK

In this paper, we demonstrated how to improve the prediction of binding affinity between proteins and ligands. Specifically, we proposed a GNN-based model, GIANT, to learn the representations of protein-ligand complexes for better binding affinity prediction by leveraging the fine-grained geometric structures and interaction information among atoms. Along this line, we designed the 3D geometric graph learning network (3DG-NET) to integrate both distance and angle information for 3D spatial structure modeling. Also, to

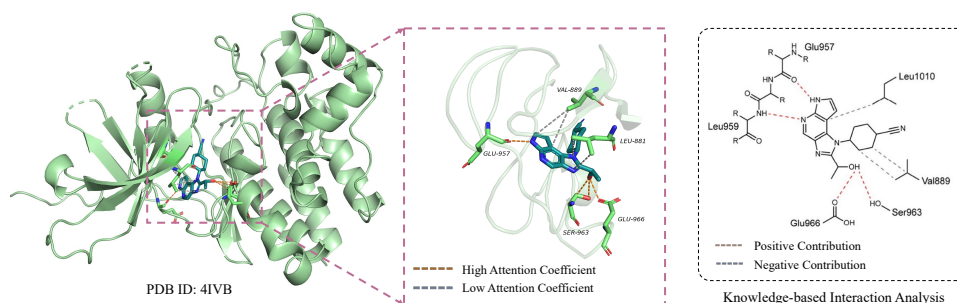


Fig. 12. Protein-ligand interaction analysis for the complex (4IVB in the core testing set) with the knowledge-based result adopted from [62].

further improve the prediction performance, we introduced a well-designed pairwise molecular interaction network PI-NET to capture semantic and structural interactions for comprehensive representation learning. Finally, the experimental results on two benchmarks showed the effectiveness and the generalizability of the proposed model.

The integration of heterogeneous graph neural networks (such as HetGNN [63] and HPN [64]) into the proposed GIANT framework can potentially improve the model's ability to handle diverse interactions among the protein and the ligand in drug discovery applications. In the future, we will consider various node types and multiple edge types for different interactions between proteins and ligands to construct the heterogeneous interaction graph. Furthermore, we will combine geometric structure modeling with heterogeneous graph learning through the meta-path feature fusion and an effective training strategy, which enables our model to learn from large-scale heterogeneous drug data.

REFERENCES

- [1] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: methods and applications," *Nature reviews Drug discovery*, vol. 3, no. 11, pp. 935–949, 2004.
- [2] J. Li and H. Li, "Using fixed point theorems to model the binding in protein-protein interactions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 8, pp. 1079–1087, 2005.
- [3] T. Fu, C. Xiao, L. M. Glass, and J. Sun, "Moler: incorporate molecule-level reward to enhance deep generative model for molecule optimization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 11, pp. 5459–5471, 2021.
- [4] X. Su, Z.-H. You, D.-s. Huang, L. Wang, L. Wong, B. Ji, and B. Zhao, "Biomedical knowledge graph embedding with capsule network for multi-label drug-drug interaction prediction," *IEEE Transactions on Knowledge and Data Engineering*, 2022.
- [5] Q. Li and S. Shah, "Structure-based virtual screening," *Methods in molecular biology (Clifton, N.J.)*, vol. 1558, p. 111–124, 2017.
- [6] A. Elnaggar, M. Heinzinger, C. Dallago, G. Rehawi, W. Yu, L. Jones, T. Gibbs, T. Feher, C. Angerer, M. Steinegger, D. Bhowmik, and B. Rost, "Prottrans: Towards cracking the language of life's code through self-supervised deep learning and high performance computing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [7] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.
- [8] H. Jhoti and A. R. Leach, *Structure-based drug discovery*. Springer, 2007, vol. 1.
- [9] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular docking: a powerful approach for structure-based drug discovery," *Current computer-aided drug design*, vol. 7, no. 2, pp. 146–157, 2011.
- [10] M. Batool, B. Ahmad, and S. Choi, "A structure-based drug discovery paradigm," *International journal of molecular sciences*, vol. 20, no. 11, p. 2783, 2019.
- [11] A. R. Leach, B. K. Shoichet, and C. E. Peishoff, "Prediction of protein-ligand interactions. docking and scoring: successes and gaps," *Journal of medicinal chemistry*, vol. 49, no. 20, pp. 5851–5855, 2006.
- [12] P. J. Ballester and J. B. Mitchell, "A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking," *Bioinformatics*, vol. 26, no. 9, pp. 1169–1175, 2010.
- [13] A. N. Jain, "Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine," *Journal of medicinal chemistry*, vol. 46, no. 4, pp. 499–511, 2003.
- [14] O. Trott and A. J. Olson, "Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of computational chemistry*, vol. 31, no. 2, pp. 455–461, 2010.
- [15] W. J. Allen, T. E. Balius, S. Mukherjee, S. R. Brozell, D. T. Moustakas, P. T. Lang, D. A. Case, I. D. Kuntz, and R. C. Rizzo, "Dock 6: Impact of new features and current docking performance," *Journal of computational chemistry*, vol. 36, no. 15, pp. 1132–1156, 2015.
- [16] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, and C. Corbeil, "Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go," *British journal of pharmacology*, vol. 153, no. S1, pp. S7–S26, 2008.
- [17] S. L. Kinnings, N. Liu, P. J. Tonge, R. M. Jackson, L. Xie, and P. E. Bourne, "A machine learning-based method to improve docking scoring functions and its application to drug repurposing," *Journal of chemical information and modeling*, vol. 51, no. 2, pp. 408–419, 2011.
- [18] H. Öztürk, A. Özgür, and E. Ozkirimli, "Deepdta: deep drug-target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, 2018.
- [19] T. Li, X. Zhao, and L. Li, "Co-vae: Drug-target binding affinity prediction by co-regularized variational autoencoders," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [20] I. Wallach, M. Dzamba, and A. Heifets, "Atomnet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery," *arXiv preprint arXiv:1510.02855*, 2015.
- [21] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh, "GraphDTA: Predicting drug-target binding affinity with graph neural networks," *Bioinformatics*, 10 2020, btaa921.
- [22] M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri, and D. R. Koes, "Protein-ligand scoring with convolutional neural networks," *Journal of chemical information and modeling*, vol. 57, no. 4, pp. 942–957, 2017.
- [23] M. M. Stepniewska-Dziubinska, P. Zielenkiewicz, and P. Siedlecki, "Development and evaluation of a deep learning model for protein-ligand binding affinity prediction," *Bioinformatics*, vol. 34, no. 21, pp. 3666–3674, 2018.
- [24] S. Li, J. Zhou, T. Xu, H. Liu, X. Lu, and H. Xiong, "Competitive analysis for points of interest," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 1265–1274.
- [25] J. Lim, S. Ryu, K. Park, Y. J. Choe, J. Ham, and W. Y. Kim, "Predicting drug-target interaction using a novel graph neural network with 3d structure-embedded graph representation," *Journal of chemical information and modeling*, vol. 59, no. 9, pp. 3981–3988, 2019.
- [26] M. Sun, S. Zhao, C. Gilvary, O. Elemento, J. Zhou, and F. Wang, "Graph convolutional networks for computational drug development and discovery," *Briefings in bioinformatics*, vol. 21, no. 3, pp. 919–935, 2020.
- [27] C. Zang and F. Wang, "Moflow: an invertible flow model for generating molecular graphs," in *Proceedings of the 26th ACM*

- SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 617–626.
- [28] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea *et al.*, “Analyzing learned molecular representations for property prediction,” *Journal of chemical information and modeling*, vol. 59, no. 8, pp. 3370–3388, 2019.
- [29] Ł. Maziarka, T. Danel, S. Mucha, K. Rataj, J. Tabor, and S. Jastrzebski, “Molecule attention transformer,” *arXiv preprint arXiv:2002.08264*, 2020.
- [30] J. Klicpera, J. Groß, and S. Günnemann, “Directional message passing for molecular graphs,” in *International Conference on Learning Representations*, 2020.
- [31] S. Li, J. Zhou, T. Xu, D. Dou, and H. Xiong, “Geomgcl: Geometric graph contrastive learning for molecular property prediction,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 4, 2022, pp. 4541–4549.
- [32] L. Cai, J. Li, J. Wang, and S. Ji, “Line graph neural networks for link prediction,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [33] K. Do, T. Tran, and S. Venkatesh, “Graph transformation policy network for chemical reaction prediction,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.
- [34] D. Leckband, J. Israelachvili, F. Schmitt, and W. Knoll, “Long-range attraction and molecular rearrangements in receptor-ligand interactions,” *Science*, vol. 255, no. 5050, pp. 1419–1421, 1992.
- [35] S. Li, J. Zhou, T. Xu, L. Huang, F. Wang, H. Xiong, W. Huang, D. Dou, and H. Xiong, “Structure-aware interactive graph neural networks for the prediction of protein-ligand binding affinity,” in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 975–985.
- [36] I. Muegge and Y. C. Martin, “A general and fast scoring function for protein-ligand interactions: a simplified potential approach,” *Journal of medicinal chemistry*, vol. 42, no. 5, pp. 791–804, 1999.
- [37] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, “Protein–ligand docking: current status and future challenges,” *Proteins: Structure, Function, and Bioinformatics*, vol. 65, no. 1, pp. 15–26, 2006.
- [38] L. Jacob and J.-P. Vert, “Protein-ligand interaction prediction: an improved chemogenomics approach,” *Bioinformatics*, vol. 24, no. 19, pp. 2149–2156, 2008.
- [39] H. Gohlke, M. Hendlich, and G. Klebe, “Knowledge-based scoring function to predict protein-ligand interactions,” *Journal of molecular biology*, vol. 295, no. 2, pp. 337–356, 2000.
- [40] R. Wang, L. Lai, and S. Wang, “Further development and validation of empirical scoring functions for structure-based binding affinity prediction,” *Journal of computer-aided molecular design*, vol. 16, no. 1, pp. 11–26, 2002.
- [41] L. J. Colwell, “Statistical and machine learning approaches to predicting protein–ligand interactions,” *Current opinion in structural biology*, vol. 49, pp. 123–128, 2018.
- [42] R. Wang, X. Fang, Y. Lu, C.-Y. Yang, and S. Wang, “The pdbname database: methodologies and updates,” *Journal of medicinal chemistry*, vol. 48, no. 12, pp. 4111–4119, 2005.
- [43] L. Zheng, J. Fan, and Y. Mu, “Onionnet: a multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction,” *ACS omega*, vol. 4, no. 14, pp. 15956–15965, 2019.
- [44] Z. Hao, C. Lu, Z. Huang, H. Wang, Z. Hu, Q. Liu, E. Chen, and C. Lee, “Asgn: An active semi-supervised graph neural network for molecular property prediction,” in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 731–752.
- [45] Y. Liu, H. Yuan, L. Cai, and S. Ji, “Deep learning of high-order interactions for protein interface prediction,” in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 679–687.
- [46] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *International Conference on Learning Representations*, 2017.
- [47] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” in *International Conference on Learning Representations*, 2018.
- [48] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, “How powerful are graph neural networks?” in *International Conference on Learning Representations*, 2019.
- [49] Y. Song, S. Zheng, Z. Niu, Z.-H. Fu, Y. Lu, and Y. Yang, “Communicative representation learning on attributed molecular graphs,” in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020, pp. 2831–2838.
- [50] J. Zhou, S. Li, L. Huang, H. Xiong, F. Wang, T. Xu, H. Xiong, and D. Dou, “Distance-aware molecule graph attention network for drug-target binding affinity prediction,” *arXiv preprint arXiv:2012.09624*, 2020.
- [51] Y. Liu, L. Wang, M. Liu, X. Zhang, B. Oztekin, and S. Ji, “Spherical message passing for 3d graph networks,” *arXiv preprint arXiv:2102.05013*, 2021.
- [52] T. Danel, P. Spurek, J. Tabor, M. Śmieja, Ł. Struski, A. Slowik, and Ł. Maziarka, “Spatial graph convolutional networks,” in *International Conference on Neural Information Processing*. Springer, 2020, pp. 668–675.
- [53] H. Xu, C. Huang, Y. Xu, L. Xia, H. Xing, and D. Yin, “Global context enhanced social recommendation with hierarchical graph neural networks,” in *2020 IEEE International Conference on Data Mining*. IEEE, 2020, pp. 701–710.
- [54] Y. Pang, L. Wu, Q. Shen, Y. Zhang, Z. Wei, F. Xu, E. Chang, B. Long, and J. Pei, “Heterogeneous global graph neural networks for personalized session-based recommendation,” in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 2022.
- [55] M. Liu, Z. Wang, and S. Ji, “Non-local graph neural networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [56] O. T. Unke and M. Meuwly, “Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges,” *Journal of chemical theory and computation*, vol. 15, no. 6, pp. 3678–3693, 2019.
- [57] Q. Li, Z. Han, and X.-M. Wu, “Deeper insights into graph convolutional networks for semi-supervised learning,” in *Thirty-Second AAAI conference on artificial intelligence*, 2018.
- [58] M. Su, Q. Yang, Y. Du, G. Feng, Z. Liu, Y. Li, and R. Wang, “Comparative assessment of scoring functions: the casf-2016 update,” *Journal of chemical information and modeling*, vol. 59, no. 2, pp. 895–913, 2018.
- [59] J. B. Dunbar Jr, R. D. Smith, C.-Y. Yang, P. M.-U. Ung, K. W. Lexa, N. A. Khazanov, J. A. Stuckey, S. Wang, and H. A. Carlson, “Cesar benchmark exercise of 2010: selection of the protein–ligand complexes,” *Journal of chemical information and modeling*, vol. 51, no. 9, pp. 2036–2046, 2011.
- [60] H. Li, K.-S. Leung, M.-H. Wong, and P. J. Ballester, “Low-quality structural and interaction data improves binding affinity prediction via random forest,” *Molecules*, vol. 20, no. 6, pp. 10947–10962, 2015.
- [61] Z. Wu, P. Jain, M. Wright, A. Mirhoseini, J. E. Gonzalez, and I. Stoica, “Representing long-range context for graph neural networks with global attention,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 13266–13279, 2021.
- [62] H. Cho, E. K. Lee, and I. S. Choi, “Layer-wise relevance propagation of interactionnet explains protein–ligand interactions at the atom level,” *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.
- [63] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, “Heterogeneous graph neural network,” in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 793–803.
- [64] H. Ji, X. Wang, C. Shi, B. Wang, and S. Y. Philip, “Heterogeneous graph propagation network,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 1, pp. 521–532, 2021.



Shuangli Li is working toward his PhD degree in the School of Computer Science and Technology at the University of Science and Technology of China. His research interests include spatial data mining, geometric deep learning and graph neural networks for drug discovery. He has published several research papers in conferences and journals, including KDD, AAAI and TKDE.

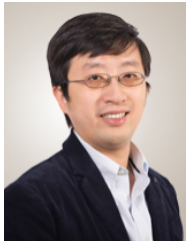


Jingbo Zhou is a staff research scientist at the Business Intelligence Lab of Baidu Research. He holds a Ph.D. in Computer Science from the National University of Singapore (2014) and a B.E. degree from Shandong University (2009). His research focuses on spatio-temporal big data, deep geometric learning, knowledge graph, and their applications in areas such as urban computing, drug discovery, and intelligent healthcare. He has published over 30 papers in top conferences and journals, including KDD,

SIGMOD, ICDE, AAAI, Nature Machine Intelligence, Nature Sustainability, Bioinformatics, and TKDE. Jingbo regularly serves as a committee member to prestigious conferences such as KDD, AAAI, IJCAI, ACL, and NeurIPS. Some of his research and development accomplishments have been applied across multiple Baidu products, including Baidu Maps, Baidu Search Ads, Baidu City Brain, and Baidu Computational Biology Platform.



Tong Xu is currently working as an Associate Professor at the University of Science and Technology of China (USTC), Hefei, China. He received the Ph.D. degree in USTC, 2016. He has authored more than 90 top-tier journal and conference papers in the fields of social network and social media analysis, including TIP, TKDE, TMC, TMM, TOMM, TOIS, KDD, SIGIR, WWW, etc. He was the recipient of the Best Paper Award of KSEM 2020 and the Best Student Paper Finalist of CICA 2021.



Liang Huang is a Full Professor of Computer Science at Oregon State University, and until recently, a Distinguished Scientist at Baidu Research USA. His research interests are natural language processing and computational biology. His honors include NAACL 2022 Best Demo Paper Award, ACL 2019 Keynote, ACL 2008 Best Paper Award, and EMNLP 2016 Best Paper Honorable Mentions. His recent work applies NLP algorithms to RNA folding and RNA design, with papers in Nature (2023) and PNAS (2021).



Fan Wang holds the position of Distinguished Architect within the Department of Natural Language Processing at Baidu. He earned his M.S. degree from the University of Colorado at Boulder in 2010 and his B.S. degree from the University of Science and Technology of China in 2008. His research encompasses a wide range of interests, spanning Natural Language Processing, Bio-Computing, Robotics, and Artificial General Intelligence.

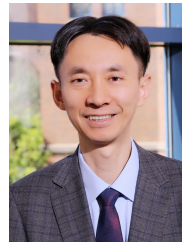


Haoyi Xiong (Senior Member, IEEE) received Ph.D in computer science from Télécom Sud-Paris and Université Pierre et Marie Curie, France in 2015. From 2016 to 2018, he was a Tenure-Track Assistant Professor with the Department of Computer Science, Missouri University of Science and Technology, Rolla, MO and a Postdoc at University of Virginia, Charlottesville, VA from 2015 to 2016. He is currently a principal architect at Big Data Lab, Baidu Inc., Beijing, China, and also a Graduate Faculty Scholar affiliated to University of Central Florida, Orlando FL. His current research

interests include AutoDL and ubiquitous computing. He has published more than 70 papers in top computer science conferences and journals, including UbiComp, ICML, ICLR, RTSS, KDD, AAAI/IJCAI, TMC, TC, TKDE, TNNLS, IOTJ, and TKDD. He was a co-recipient of the 2020 IEEE TCSC Award for Excellence in Scalable Computing (Early Career Researcher) and the prestigious Science & Technology Advancement Award (First Prize) from Chinese Institute of Electronics in 2019.



Dr. Weili Huang is a consultant/advisor in pharmaceutical and AI industries. In 2010, Dr. Huang founded Miracle Query, Inc in Eugene, Oregon. The primary goal of this company is to provide an intelligent system for public to access to health care information and enhance the knowledge sharing between the public and the Agency (FDA). Before that, Dr. Huang was a FDA pharmacologist reviewer in Center for Drug Evaluation and Research (CDER). As a reviewer, Dr. Huang prepared and published over 40 drug bioequivalence guidances on the FDA website. She was honored for her work in the Bioequivalence Database Group with an FDA Commissioner's Special Citation (2008-2009). Before joining the FDA, Dr. Huang was a research fellow at the University of Washington, Department of Pharmaceutics and she was awarded the Elmer M. and Joy B. Plein Fellowship (2006-2007) and Elmer M. Plein Endowed Research Fund (2006-2007). In 2006, Dr. Huang received her Ph.D. degree in Pharmaceutics from the University of Washington. Before coming to the U.S, Dr. Huang studied in the master's program of pharmacology at the Peking Union Medical College (China) from 1997 to 1999. Dr. Huang received B.S. degree of Pharmacy from the Peking University Health Science Center (China) in 1997.



Dejing Dou is the Chief Data Scientist, Partner and Vice President of BCG X. Previously, he was the Head of Big Data Lab (BDL) and Business Intelligence Lab (BIL) at Baidu Research. He was also a full Professor from the Computer and Information Science Department at the University of Oregon and has led the Advanced Integration and Mining (AIM) Lab since 2005. He has been the Director of the NSF IUCRC Center for Big Learning (CBL) since 2018. He was a visiting associate Professor at Stanford Center

for Biomedical Informatics Research during 2012-2013. Prof. Dou received his bachelor degree from Tsinghua University, China in 1996 and his Ph.D. degree from Yale University in 2004. His research areas include artificial intelligence, data mining, data integration, NLP, and health informatics. Dejing Dou has published more than 160 research papers, some of which appear in prestigious conferences and journals like AAAI, IJCAI, ICML, NeurIPS, ICLR, KDD, ICDM, ACL, EMNLP, CVPR, ICCV, CIKM, ISWC, TKDD, JIIS, and JoDS, with more than 4900 Google Scholar citations. His DEXA'15 paper received the best paper award. His KDD'07 paper was nominated for the best research paper award. His COLING'18 paper was Area Chair Favorites (excellent). He is on the Editorial Boards of Journal on Data Semantics, Journal of Intelligent Information Systems, and PLOS ONE. He is an Editor-in-Chief of AIMS Electronic Research Archive. He has been serving as program committee members for major international conferences and as program co-chairs for five of them. He has received over \$5 million PI research grants from the NSF and the NIH. Dejing Dou is a senior member of AAAI, ACM, and IEEE.



Hui Xiong (F'20) is a Chair Professor, Associate Vice President (Knowledge Transfer), and Head of the AI Thrust at Hong Kong University of Science and Technology (Guangzhou). His research interests span Artificial Intelligence, data mining, and mobile computing. He obtained his PhD in Computer Science from the University of Minnesota, USA. Dr. Xiong has served on numerous organization and program committees for conferences, including as Program Co-Chair for the Industrial and Government Track for the

18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Program Co-Chair for the IEEE 2013 International Conference on Data Mining (ICDM), General Co-Chair for the 2015 IEEE International Conference on Data Mining (ICDM), and Program Co-Chair of the Research Track for the 2018 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. He received several awards, such as the 2021 AAAI Best Paper Award and the 2011 IEEE ICDM Best Research Paper award. For his significant contributions to data mining and mobile computing, he was elected as a Fellow of both AAAS and IEEE in 2020.