ChiENN: Embracing Molecular Chirality with Graph Neural Networks

Piotr Gaiński[1(✉)], Michał Koziarski[2,3], Jacek Tabor[1], and Marek Śmieja[1]

[1] Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland
`piotr.gainski@doctoral.uj.edu.pl`
[2] Mila - Quebec AI Institute, Montreal, QC, Canada
[3] Université de Montréal, Montreal, QC, Canada

Abstract. Graph Neural Networks (GNNs) play a fundamental role in many deep learning problems, in particular in cheminformatics. However, typical GNNs cannot capture the concept of chirality, which means they do not distinguish between the 3D graph of a chemical compound and its mirror image (enantiomer). The ability to distinguish between enantiomers is important especially in drug discovery because enantiomers can have very distinct biochemical properties. In this paper, we propose a theoretically justified message-passing scheme, which makes GNNs sensitive to the order of node neighbors. We apply that general concept in the context of molecular chirality to construct Chiral Edge Neural Network (ChiENN) layer which can be appended to any GNN model to enable chirality-awareness. Our experiments show that adding ChiENN layers to a GNN outperforms current state-of-the-art methods in chiral-sensitive molecular property prediction tasks.

Keywords: Graph Neural Networks · GNN · Message-passing · Chirality · Molecular Property Prediction

1 Introduction

Recent advances in Graph Neural Networks (GNNs) have revolutionized cheminformatics and enabled learning the molecular representation directly from chemical structures [9,30]. GNNs are widely used in molecular property prediction [20,31,34,36], synthesis prediction [3,17], molecule generation [2,21,22], or conformer generation [6,11,19,33]. Surprisingly, typical GNNs cannot capture the concept of chirality, roughly meaning they do not distinguish between a molecule and its mirror image, called enantiomer (see Fig. 1). Although enantiomers share many physical, chemical, and biological properties, they may behave remarkably differently when interacting with other chiral molecules, e.g. chiral proteins. For this reason, capturing chirality is critical in the context of drug design [10,12,16,23,25] and should not be ignored by the design of GNN architecture.

A chiral molecule is a molecule with at least one chiral center which is usually a carbon atom with four non-equivalent constituents. The mirror image
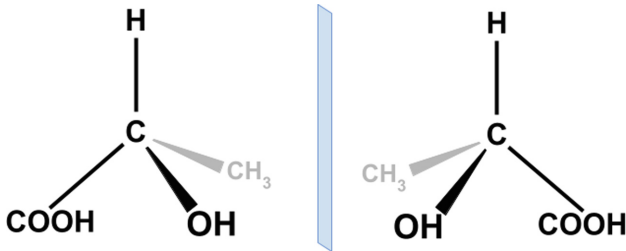
Fig. 1. An example of a chiral molecule (left) and its mirror image (right).

of a chiral molecule, called an enantiomer, cannot be superposed back to the original molecule by any combination of rotations, translations, and conformational changes (see Fig. 1). Therefore, enantiomers are molecules with different bond arrangements and the same graph connectivity. There are many examples of chiral drugs used in pharmacy whose enantiomers cause substantially different effects [23]. For instance (S)-penicillamine is an antiarthritic drug while its enantiomer (R)-penicillamine is extremely toxic [15].

Actually, chirality can be a characteristic of any class of graphs embedded in euclidean space (where we have an intuitive notion of reflection). For instance, Fig. 2 shows two 2D road maps that are mirror images of each other and possess different properties. For this reason, modeling chirality in GNNs is not restricted to the chemical domain.

In this paper, we propose and theoretically justify a novel order-sensitive message-passing scheme, which makes GNNs sensitive to chirality. In contrast to existing methods of embracing chirality, our framework is not domain specific and does not rely on arbitrary chiral tagging or torsion angles (see Sect. 2). The only inductive bias our method introduces to a GNN is the dependency on the orientation of the neighbors around a node, which lies at the core of chirality.
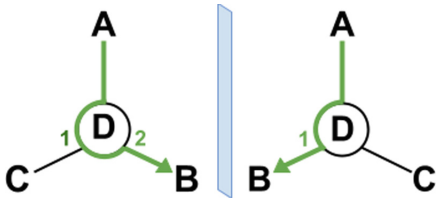
Fig. 2. An illustration of a road map (left) and its mirror image (right). We see that the maps share the same connectivity between cities, however, to get from city A to city B one has to take the second exit on a roundabout D for the left map, and the first exit for the right map.

The key component of the proposed framework is the message aggregation function. In a typical GNN, the messages incoming to a node from its neighbors are treated as a set and aggregated with a permutation-invariant function (sum, max, etc.). It makes the model unable to distinguish between chiral graphs with the same connectivity, but with different spatial arrangements. We re-invent this approach and introduce a message aggregation function that is sensitive to the spatial arrangement (order) of the neighbors. Our approach can be used in any

chiral-sensitive graph domain where chirality can be expressed by an order of the neighboring nodes.

We apply that general order-sensitive message-passing framework in the context of molecular chirality to construct Chiral Edge Neural Network (ChiENN) layer. The ChiENN layer can be appended to most molecular GNN models to enable chirality sensitiveness. Our experiments show that ChiENN can be successfully used within existing GNN models and as a standalone model consisting of stacked ChiENN layers. In both cases, ChiENN outperforms current state-of-the-art methods in chiral-sensitive molecular property prediction tasks by a large margin. We make our code publicly available[1].

Our contributions are as follows:

1. We propose and theoretically justify a general order-sensitive message-passing scheme. Our method can be adapted to any chiral-sensitive graph domain where chirality can be expressed by an order of the neighboring nodes (Sect. 3).
2. We use the proposed framework to construct a novel ChiENN layer that enables chirality awareness in any GNN model in the domain of molecular graphs (Sect. 4). The proposed ChiENN can be applied to any 3D graph task with the notion of chirality.
3. We evaluate and analyze the ChiENN layer and show that it outperforms current state-of-the-art methods in chiral-sensitive molecular property prediction tasks (Sect. 5).

2   Related Work

Explicit Tagging of Chiral Center. The most common approach for incorporating chirality into GNN is to use local or global chiral tags [13,18,27]. Both local and global tagging can be seen in the following way. Every carbon atom with four non-equivalent constituents, called a chiral center, is given a tag (CCW or CW) describing the orientation of its constituents. The orientation is defined using the enumeration of constituents computed by an arbitrary algorithm. The constituent with the highest number (4) is positioned so that it points away from the observer. The curve passing through the constituents with numbers 1, 2, and 3 respectively determines a clockwise (CW) or counterclockwise (CCW) orientation of the chiral center. Although enumeration algorithms for global and local tagging differ (the latter is not explicitly used in practice), the expressivity of both methods is limited, as we show in Sect. 5.

3D GNNs with Torsion Angles. Some recent GNN models enrich graphs with 3D information, like distances between atoms [4,20], angles between bonds [8,28], and torsion angles between two bonds joined by another bond [5,7]. As distances and angles are invariant to chirality, the torsion angles (that are negated

---

upon reflection) are required for 3D GNN to express the chirality. However, even access to a complete set of torsion angles does not guarantee expressivity in chiral-sensitive tasks as shown in [1]. Torsion angles are sensitive to bond rotations and can also be negated by the reflection of a non-chiral molecule. In [1], the authors proposed the ChIRo model that instead of embedding single torsion angles, embeds sets of torsion angles with a common bond. ChIRo is the current state-of-the-art method for chiral-sensitive tasks. In contrast to ChIRo, our proposed method does not incorporate distances, angles, or torsion angles. It only relies on the orientation of neighbors around a node, making it more general and easily adaptable to other chiral graph domains. Moreover, our experiments show that the ChiENN layer outperforms ChIRo by a large margin on chiral-sensitive molecular tasks (see Sect. 5).

Changing Aggregation Scheme. The method most related to our approach is the Tetra-DMPNN model from [24] which replaces a classic message-passing scheme with a chiral-sensitive one. The proposed aggregation scheme is guided by local chiral tags, meaning that it relies on some arbitrary rules for enumerating neighbors and cannot be applied to nodes other than chiral centers. Moreover, the Tetra-DMPNN method is computationally expensive and does not scale with the number of possible neighbors of a chiral center, making the model useful only in the context of chemistry. Our approach provides a general, efficient, and scalable chiral-sensitive message passing and outperforms the Tetra-DMPNN on chiral-sensitive molecular tasks by a large margin (see Sect. 5).

3   Order-Sensitive Message-Passing Scheme

Setting. Let us consider a directed graph $G = (X, E)$ in which every node $x_i \in X$ is represented by a N-dimensional encoding ($x_i \in \mathbb{R}^N$). Edge e_{ij} connects nodes x_i and x_j and is represented by M-dimensional encoding ($e_{ij} \in \mathbb{R}^M$).

In addition, we assume that for every node $x \in X$, we are given an order o of all its neighbors $o = (x_0, x_1, \ldots, x_{d-1})$. The order of neighbors forms a sequence, which stands in contrast to typical graphs, where neighbors are treated as an unordered set. Given a permutation π on $\{0, 1, \ldots, d-1\}$, we assume that two orders $o_1 = (x_0, \ldots, x_{d-1})$ and $o_2 = (x_{\pi(0)}, \ldots, x_{\pi(d-1)})$ are equivalent if and only if π is a shift i.e. $\pi(i) = (i + k) \bmod d$, for a fixed $k \in \mathbb{Z}$. In other words, the neighbors form the sequence on a ring.

One of the most common mechanisms in GNN is message-passing, which updates the representation of a node x by the information coming from its neighbors $(x_0, \ldots, x_{d-1})$, which can be written as:

$$x' = f(x; x_0, \ldots, x_{d-1}).$$

In this paper, we are going to describe the general message-passing scheme, which is aware of the neighbors' order. Before that, we discuss possible choices of the aggregation function f.

Vanilla Message-Passing as a Permutation-Invariant Transformation.
Let us first discuss a basic case, where f is a permutation-invariant function, i.e.

$$f(x; x_0, \ldots, x_{d-1}) = f(x; x_{\pi(0)}, \ldots, x_{\pi(d-1)}),$$

for every permutation π of $\{0, 1, \ldots, d-1\}$. This aggregation ignores the order
of neighbors and lies in a heart of typical GNNs.

Let us recall that f is permutation-invariant with respect to $\{x_0, x_1, \ldots, x_{d-1}\}$
if and only if it can be decomposed in the form [35]:

$$f(x_0, x_1, \ldots, x_{d-1}) = \rho(\sum_{i=0}^{d-1} \phi(x_i)),$$

for suitable transformations ϕ and ρ. In the context of graphs, a general form of
a permutation-invariant aggregation of neighbors $\{x_0, x_1, \ldots, x_{d-1}\}$ of x is:

$$x' = f(x; x_0, \ldots, x_{d-1}) = \rho(x; \sum_{i=0}^{d-1} \phi(x; x_i)), \tag{1}$$

for suitable transformations ϕ and ρ. By specifying ρ, ϕ as neural networks, we
get the basic formula of vanilla message-passing.

Shift-Invariant Aggregation. Vanilla message-passing relies on permutation-
invariant aggregation and it does not take into account the neighbor's order.
Thus we are going to discuss the weaker case of aggregation function f and
assume that f is shift-invariant, i.e.

$$f(x; x_0, \ldots, x_{d-1}) = f(x; x_{0+p}, \ldots, x_{d-1+p}),$$

for any shift by a number $p \in 0, \ldots, d-1$, where the additions on indices are
performed modulo d. This assumption is consistent with our initial requirement
that shifted orders are equivalent.

The following theorem gives a general formula for shift-invariant mappings.

Theorem 1. *The function f is shift-invariant if and only if f can be written
as:*

$$f(x_0, \ldots, x_{d-1}) = \sum_{p=0}^{d-1} g(x_{0+p}, \ldots, x_{d-1+p})$$

for an arbitrary function g.

Proof. If f is shift invariant, then $f(x_0, \ldots, x_{d-1}) = f(x_{0+p}, \ldots, x_{d-1+p})$ for
every p, and consequently

$$f(x_0, \ldots, x_{d-1}) = \sum_{p=0}^{d-1} \frac{1}{d} f(x_{0+p}, \ldots, x_{d-1+p}).$$

On the other hand, if the function f can be written as $\sum_{p=0}^{d-1} g(x_{0+p}, \ldots, x_{d-1+p})$,
then it is shift-invariant for arbitrary function g.

Following the above theorem, we get a general formula for shift-invariant aggregation applicable to graphs:

$$x' = \rho(x; \sum_{p=0}^{d-1} \psi(x; x_{0+p}, \ldots, x_{d-1+p})), \tag{2}$$

for suitable ρ and ψ, where all additions are performed modulo d.

Now, we want to ensure that our function f is not only shift-invariant but also order-sensitive

Order-Sensitive Message-Passing. Let us assume that we are in the class of shift-invariant transformations. We are going to specify the formula (2) to obtain ab aggregation, which is sensitive to any permutation other than shift. More precisely, we say that f is order-sensitive if and only if for every permutation π, we have:

$$f(x; x_0, \ldots, x_{d-1}) = f(x; x_{\pi(0)}, \ldots, x_{\pi(d-1)}) \iff \pi(i) = (i+k) \bmod d.$$

Let us investigate typical functions ψ in formula (2), which can be implemented using neural networks. We start with the simplest case, where ψ is linear. Then

$$\sum_{p=0}^{d-1} \psi(x_{0+p}, \ldots, x_{d-1+p}) = \sum_{p=0}^{d-1} \sum_{i=0}^{d-1} w_i x_{i+p} = \sum_{i=0}^{d-1} w_i \sum_{p=0}^{d-1} x_{i+p} = \sum_{i=0}^{d-1} w_i \sum_{j=0}^{d-1} x_j,$$

does not depend on the order of the neighbors $(x_0, ..., x_{d-1})$. To construct more complex functions, we use an arbitrary Multi-Layer Perceptron (MLP) as ψ. Since MLPs are universal approximators (for a sufficiently large number of hidden units), we can find such parameters θ that $\sum_{p=0}^{d-1} \psi_\theta(x_{\pi(0)+p}, \ldots, x_{\pi(d-1)+p})$ returns a different value for every permutation π that is not a shift. Therefore our aggregation scheme with ψ given by MLP can learn order-sensitive mapping.
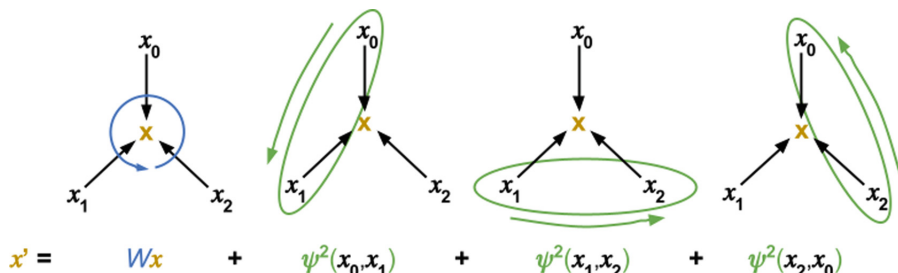
Following the above observations, we implement our order-sensitive message-passing using MLP as ψ. To match our construction to various numbers of neighbors in a graph, we restrict ψ to be k-ary (denoted as ψ^k) for some fixed $k > 1$ and overload it so that:

$$\psi^k(x_0, \ldots, x_{d-1}) = \psi^k(x_0, \ldots, x_{d-1}, \underbrace{0, \ldots, 0}_{k-d}) \text{ for } d < k.$$

Given that, we implement the Eq. (2) with the following neural network layer:

$$x' = Wx + \sum_{p=0}^{d-1} \psi^k(x_{0+p}, ..., x_{k-1+p}),$$
$$\psi^k(x_{0+p}, ..., x_{k-1+p}) = W_1 \sigma(W_2(x_{0+p}|...|x_{k-1+p})). \tag{3}$$

$$x' = \quad Wx \quad + \quad \psi^2(x_0,x_1) \quad + \quad \psi^2(x_1,x_2) \quad + \quad \psi^2(x_2,x_0)$$

Fig. 3. An illustration of our update rule for node x with 3 ordered neighbors (x_0, x_1, x_2) and for $k = 2$. We see that ψ^k is used to embed pairs of consecutive nodes.

Our k-ary message function ψ^k is composed of concatenation operator | and two-layer MLP with ELU as σ. Intuitively, the output of $\psi^k(x_{0+p}, ..., x_{k-1+p})$ can be seen as a message obtained jointly from k consecutive neighbors starting from a neighbor p in order $(x_0, ..., x_{d-1})$ which is illustrated in Fig. 3.

## 4   ChiENN: Chiral-Aware Neural Network

In this section, we apply the order-sensitive message-passing framework to molecular graphs. We show that order-sensitive aggregation is a key factor for embracing molecular chirality. Roughly speaking, in contrast to vanilla message-passing, the proposed ChiENN (Chiral-aware Edge Neural Network) is able to distinguish enantiomers, where one molecule is a mirror image of the second. Although we evaluate the ChiENN model in the context of molecular property prediction, the proposed model can be applied to any 3D graph task with the notion of chirality.

To construct ChiENN based on our order-sensitive message-passing scheme from Eq. (3), we need to define a notion of neighbors' order in molecular graphs that grasps the concept of chirality (see Fig. 4). We introduce this notion of order for edge (dual) molecular graphs and provide a simple transformation from standard molecular graphs to edge molecular graphs. Therefore the rest of the section is organized into three subsections:

1. **Edge Graph** describing the transformation from a molecular graph to its edge (dual) form used in our ChiENN model,
2. **Neighbors Order** defining the order of the neighbors in an edge graph,
3. **Chiral-Aware Update** constructing order-sensitive update rule using our order-sensitive framework from the Sect. 3.

4.1   Edge Graph

Let us suppose, we have a directed graph $G = (X, C, E)$ that represents a concrete conformation (3D embedding) of a molecule. The node encoding $x_i \in X$
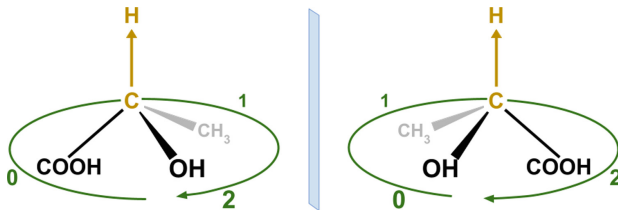
Fig. 4. An intuitive illustration of the neighbor ordering in a molecule. First, we pick a directed bond from atom C to H and then order the rest of the neighbors around that bond. We observe that for a chiral molecule (left) and its mirror image (right), we obtain different orders of the $COOH$, CH_3, and OH constituents.

corresponds to an i-th atom from a molecule, $c_i \in C \subseteq \mathbb{R}^3$ are its coordinates in 3D space, and the edge encoding $e_{ij} \in E$ represents a bond between i-th and j-th atoms.

To make the definition of neighbor order straightforward, our ChiENN model operates on an edge (dual) graph $G' = (X', C', E')$ which swaps nodes with edges from the original graph G. It means that the node $x_{ij} \in X'$ represents the edge $e_{ij} \in E$, while the edge $e_{ij,jk} \in E'$ represents the node x_j that connects edge $e_{ij} \in E$ with $e_{jk} \in E$. Similarly, $c'_{ij} \in \mathbb{R}^3 \times \mathbb{R}^3$ is now a 3D coordinate vector that links positions c_i and c_j. Formally, we have:

$$X' = \{x_{ij} = e_{ij} : e_{ij} \in E\},$$
$$C' = \{c_{ij} = c_i | c_j : c_i, c_j \in C, e_{ij} \in E\},$$
$$E' = \{e_{ij,jk} = e_{ij} | x_j | e_{jk} : e_{ij}, e_{jk} \in E, x_j \in X\},$$

where $|$ stands for a concatenation operator. Clearly, the constructed edge graph $G' = (X', C', E')$ can be fed to any GNN that can take as an input the original graph $G = (X, C, E)$.
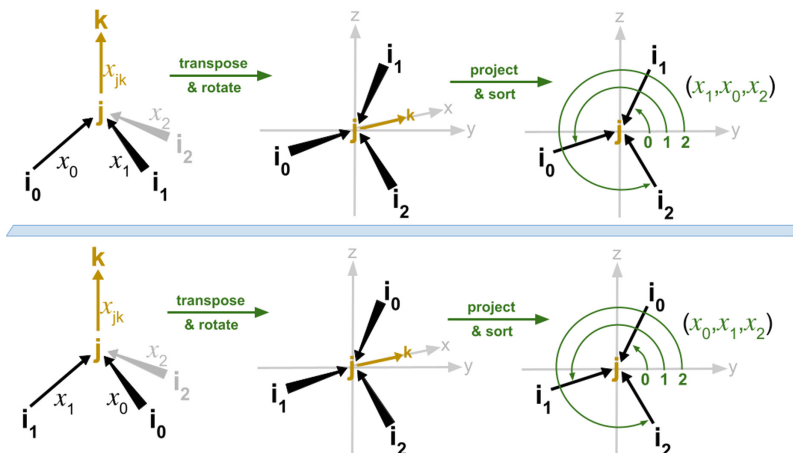
4.2   Neighbors Order

In an edge molecular graph $G = (X, C, E)$, a node $x_{jk} \in E$ represents a directed bond from atom j to atom k in the original molecule. It is assigned with a 3D vector $c_{jk} \in C \subseteq \mathbb{R}^3 \times \mathbb{R}^3$ spanned from atom j to atom k. Therefore, we will sometimes refer to nodes as if they were 3D vectors.

Let us consider the node x_{jk} and the set of its incoming neighbors: $N(x_{jk}) = \{x_{i_1j}, x_{i_2j}, ..., x_{i_dj}\}$. By construction of G, every node x_{jk} has a corresponding parallel node x_{kj}. For simplicity, we will treat this parallel node separately and exclude it from the set of neighbors, i.e. $x_{kj} \notin N(x_{jk})$.

The construction of the neighbors $N(x_{jk})$ order is illustrated in Fig. 5 and consists of two steps:

1. Transformation: first, we perform a sequence of 3D transformations on x_{jk} and $N(x_{jk})$ to make x_{jk} anchored to coordinate origin, perpendicular to yz plane and pointed away from the observer (see Fig. 5b)).

Fig. 5. An illustration of ordering the neighbors $\{x_0, x_1, x_2\}$ of x_{jk} for a chiral molecule (top) and its mirror image (bottom) around the chiral center j. First, we perform a sequence of 3D transformations on x_{jk} and its neighbors to make x_{jk} anchored to coordinate origin, perpendicular to yz plane and pointed away from the observer. Next, we project the transformed neighbors to the yz plane and sort the projections by the angle to the y axis. We see that for the chiral molecule (top) and its mirror image (bottom), we obtained non-equivalent orders (x_1, x_0, x_2) and (x_0, x_1, x_2).

2. Sorting: second, we project the transformed neighbors $N(x_{jk})$ to the yz plane and sort the projections by the angle to the y axis.

Details of the above construction are presented in the supplementary materials.

Two observations can be made regarding the above construction:

Observation 1. *The above construction returns non-equivalent orders for a chiral center and its mirror image.*

Observation 2. *Any $SE(3)$ transformation of a molecule coordinates C and any internal rotation of its bonds (conformation) can only change the shift of the order o, resulting in equivalent order o'.*

Therefore, the above construction grasps the notion of chirality in a molecule and is additionally $SE(3)$- and conformation-invariant.

We artificially excluded x_{kj} from a set of x_{jk} neighbors, because its parallel to x_{jk} and therefore its angle to y axis after the sequence of transformations is undefined. In theory, another neighbor x_{ij} can also be parallel to x_{jk} and should also be excluded from the neighbor set, but we have not observed such a case in our experiments and decided not to take it into account.

4.3   Chiral-Aware Update

Once we transformed a molecular graph to edge (dual) molecular graph $G = (X, E, C)$ using transformation from Sect. 4.1 and assigned every node x_{jk} with

an order of its neighbors $(x_1, ..., x_d)$ using construction from Sect. 4.2, we can define the order-sensitive update rule of our ChiENN model:

$$x'_{jk} = W_1 x_{jk} + W_2 x_{kj} + \sum_{p=0}^{d-1} \psi^k(x_{0+p}, ..., x_{k-1+p}),$$

$$\psi^k(x_{0+p}, ..., x_{k-1+p}) = W_3 \sigma(W_4(x_{0+p}|...|x_{k-1+p})),$$

(4)

where ψ^k is k-ary message function and σ is ELU non-linear activation. The update rule is almost the same as that from Eq. (3), but here we add a term that explicitly embeds x_{kj} node, which was artificially excluded from the order of the x_{jk} neighbors.

5   Experiments

We compare ChiENN with several state-of-the-art models on a variety of chiral-sensitive tasks. Details of experiments are described in Sect. 5.1, while the results can be found in Sect. 5.2. Furthermore, to validate design choices behind ChiENN we also conducted an ablation study, presented in Sect. 5.3.

5.1   Set-Up

Datasets. We conduct our experiments on five different datasets affected by molecule chirality. First, two datasets proposed in [1] which are designed specifically to evaluate the capability of a model to express chirality: classification of tetrahedral chiral centers as R/S (which should be a necessary, but not sufficient, condition to learn meaningful representations of chiral molecules); and enantiomer ranking, in which pairs of enantiomers with enantioselective docking scores were selected, and the task was to predict which molecule of the pair had a lower binding affinity in a chiral protein pocket.

Second, the binding affinity dataset, which is an extension of the previously described enantiomer ranking, with the same underlying molecules, but the task being regression of the binding affinity.

Additionally, we take two datasets from the MoleculeNet benchmark [32] that do not explicitly require prediction of molecule chirality, but contain some percentage of molecules with chiral centers, and the underlying biological task in principle might be chirality-dependant: BACE, a binary classification dataset for prediction of binding results for a set of inhibitors of human β-secretase 1 (BACE-1) [29]; and Tox21, a multilabel classification dataset containing qualitative toxicity measurements on 12 different targets, including nuclear receptors and stress response pathways.

Reference Methods. As reference models we consider several state-of-the-art neural network architectures for processing graphs, both chirality-aware and general: GPS [26], SAN [14], DMPNN [34], ChIRo [1], and Tetra-DMPNN [24].

Table 1. Comparison of ChiENN-based approaches with the reference methods on chiral-sensitive tasks. Methods are split into groups by the underlying base model, except for the bottom group which includes models specifically designed to be chiral-sensitive. We **bold** the best results in every group and underline the best results across all groups. All variations of our method (ChiENN, SAN+ChiENN, and GPS+ChiENN) significantly outperform current state-of-the-art chiral-sensitive models. Note that for the R/S task, we omitted the results for models with chiral tags encoded in node features, for which the task is trivial.

Model	R/S	Enantiomer ranking	Binding affinity
	Accuracy ↑	R. Accuracy ↑	MAE ↓
DMPNN	0.500 ± 0.000	0.000 ± 0.000	0.310 ± 0.001
DMPNN+tags	-	$\mathbf{0.701 \pm 0.003}$	$\mathbf{0.285 \pm 0.001}$
GPS	0.500 ± 0.000	0.000 ± 0.000	0.330 ± 0.003
GPS+tags	-	0.669 ± 0.037	0.318 ± 0.004
GPS+ChiENN	$\underline{\mathbf{0.989 \pm 0.000}}$	$\mathbf{0.753 \pm 0.004}$	$\mathbf{0.258 \pm 0.001}$
SAN	0.500 ± 0.000	0.000 ± 0.000	0.317 ± 0.004
SAN+tags	-	0.722 ± 0.004	0.278 ± 0.003
SAN+ChiENN	$\mathbf{0.987 \pm 0.001}$	$\underline{\mathbf{0.764 \pm 0.005}}$	$\underline{\mathbf{0.257 \pm 0.002}}$
ChIRo	0.968 ± 0.019	0.691 ± 0.006	0.359 ± 0.009
Tetra-DMPNN	0.935 ± 0.001	0.690 ± 0.006	0.324 ± 0.026
ChiENN	$\underline{\mathbf{0.989 \pm 0.000}}$	$\mathbf{0.760 \pm 0.002}$	$\mathbf{0.275 \pm 0.003}$

For models not designed to process chirality, that is DMPNN, GPS, and SAN, we additionally considered their variants with chiral atom tags included in the node features, similar to [1]. For the proposed approach we consider both a pure model obtained by stacking several ChiENN layers, as well as combining ChiENN layers with other architectures (ChiENN+GPS and ChiENN+SAN).

Training Details. All models were trained using Adam optimizer for up to 100 epochs, with a cosine learning rate scheduler with 10 warm-up epochs and gradient norm clipping, following the set-up of [26]. Cross-entropy and L1 loss functions were used for classification and regression, respectively. Note that in contrast to [1], to keep the set-up consistent across models we did not use triplet margin loss for ChIRo, and observed worse results than reported in [1].

We also performed a grid search of parameters with the identical budget. For all datasets and models, we reported results averaged from three runs.

For enantiomer ranking, binding affinity, and R/S, we used data splits provided by [1] and for BACE and Tox21, we used random splits with a train-valid-test ratio of 7:1:2. For each model and dataset, we report mean results from 3 independent runs with the best parameters picked by grid search.

Evaluation. Note that for the binding rank task we used accuracy modified with respect to [1]. We required the difference between the predicted affinity of two enantiomers to be higher than the threshold of 0.001. This led to ranking accuracy being equal to 0 for models unable to distinguish chiral molecules.

5.2   Comparison with Reference Methods

In this section, we compare ChiENN-based networks with state-of-the-art reference architectures using the experimental setting described in Sect. 5.1.

Chiral-Sensitive Tasks. The results on chiral-sensitive tasks are presented in Table 1. For both the enantiomer ranking and binding affinity, ChiENN-based approaches achieved the best results, producing a significant improvement in performance over the state-of-the-art chiral-aware architectures, that is ChIRo and Tetra-DMPNN. For both GPS and SAN, there was a significant improvement in performance due to the addition of ChiENN layers when compared to chiral tag inclusion. It demonstates that ChiENN model can enable chiral-awareness demonstrating the general usefulness of the proposed layer, and the fact that it can be combined with a model preferred in a given task.

Finally, as expected, all of the chirality-aware methods can properly distinguish chiral centers in the R/S task, while the baselines that do not capture the concept of chirality (DMPNN, GPS and SAN) cannot. Note that for this task, we omitted the results for models with chiral tags encoded in node features, for which the task is trivial.

Remaining Tasks. The results on BACE and Tox21 tasks are in Table 2. We see that the ChiENN model achieves results comparable to state-of-the-art models, however the influence of chirality-sensitiveness on these tasks is not clear. For SAN we actually observed a slight drop in performance when using ChiENN layers, and for GPS the results remained roughly the same. The possible explanations for that might be either 1) lack of importance of chirality on predicted tasks, or 2) small dataset size, leading to overfitting in presence of chiral information. Our conclusion is that ChiENN layers significantly improve the performance in chiral-sensitive tasks, and produce comparable results in the other tasks, where the influence of chirality is not clear. We believe that further investigation on the influence of chirality on the tasks commonly used in the molecular property prediction domain would be beneficial and we leave it for future work.

5.3   Ablation Studies

Comparison of k-Ariness of the Message Function. We began with an analysis of the impact of k-ariness (Eq. 3) of the message function used by ChiENN. Specifically, in this experiment, we used the pure variant of ChiENN, which is a graph neural network using ChiENN layers as message-passing layers.

Table 2. Comparison of ChiENN-based approaches with the reference methods on tasks **not** explicitly requiring chirality. We see that the ChiENN model achieves results comparable to state-of-the-art models.

Model	BACE	Tox21
	AUC ↑	AUC ↑
DMPNN	**0.847 ± 0.015**	0.813 ± 0.008
DMPNN+tags	0.840 ± 0.004	**0.824 ± 0.006**
GPS	**0.841 ± 0.004**	0.821 ± 0.000
GPS+tags	0.812 ± 0.017	**0.825 ± 0.002**
GPS+ChiENN	0.839 ± 0.008	0.821 ± 0.007
SAN	**0.846 ± 0.012**	**0.842 ± 0.007**
SAN+tags	0.829 ± 0.009	0.841 ± 0.004
SAN+ChiENN	0.826 ± 0.014	0.834 ± 0.005
ChIRo	0.815 ± 0.010	**0.847 ± 0.005**
Tetra-DMPNN	0.824 ± 0.017	0.807 ± 0.003
ChiENN	**0.838 ± 0.003**	0.838 ± 0.003

Table 3. Comparison of k-ariness of the message function.

k-ary	R/S	Enantiomer ranking	Binding affinity	BACE	Tox21
	Accuracy ↑	R. Accuracy ↑	MAE ↓	AUC ↑	AUC ↑
$k = 1$	0.500 ± 0.000	0.000 ± 0.000	0.328 ± 0.000	0.831 ± 0.028	0.833 ± 0.005
$k = 2$	**0.989 ± 0.001**	0.759 ± 0.003	**0.267 ± 0.001**	0.788 ± 0.014	0.836 ± 0.004
$k = 3$	**0.989 ± 0.000**	**0.760 ± 0.002**	0.275 ± 0.003	**0.838 ± 0.003**	**0.838 ± 0.003**

We varied $k \in \{1, 2, 3\}$, where $k = 1$ disables the ability of the network to distinguish enantiomers as it collapses our order-sensitive message passing scheme from Eq. (3) to vanilla message-passing from Eq. (1). We considered values of k up to 3 since it corresponds to the airiness of standard chiral centers observed in the edge graphs (see Sect. 4.1) of molecules.

The results are presented in Table 3. As expected, choosing $k = 1$ leads to a failure in distinguishing enantiomers (makes message passing permutation invariant), as demonstrated by minimum performance in R/S and enantiomer ranking tasks. Interestingly, for most datasets choosing $k = 2$ was sufficient, leading to a comparable performance to $k = 3$. The only exception to that was BACE dataset, for which a noticeable drop in performance was observed when using $k = 2$. We used $k = 3$ in the remainder of this paper.

Using ChiENN Layer with Existing Models. Secondly, we conducted an ablation of different design choices that can be made to enable enantiomer recognition within the existing architectures. Specifically, we focused on the GPS model and considered using three different strategies: conversion to edge graph

Table 4. Ablation study of different design choices for GPS+ChiENN model. The "Graph" column indicates conversion to the edge graph; the "Tags" column indicates the inclusion of the chiral tags as node features; the "ChiENN" column indicates the usage of ChiENN as a message-passing layer.

Graph	Tags	ChiENN	R/S	Enantiomer ranking	Binding affinity	BACE	Tox21
			Accuracy ↑	R. Accuracy ↑	MAE ↓	AUC ↑	AUC ↑
No	No	No	0.500 ± 0.000	0.000 ± 0.000	0.330 ± 0.003	0.841 ± 0.004	0.821 ± 0.000
Yes	No	No	0.500 ± 0.000	0.000 ± 0.000	0.306 ± 0.001	$\mathbf{0.851 \pm 0.007}$	0.821 ± 0.007
No	Yes	No	$\mathbf{1.000 \pm 0.000}$	0.669 ± 0.037	0.318 ± 0.004	0.812 ± 0.017	0.825 ± 0.002
Yes	Yes	No	$\mathbf{1.000 \pm 0.000}$	0.720 ± 0.002	0.283 ± 0.008	0.802 ± 0.019	$\mathbf{0.838 \pm 0.003}$
Yes	No	Yes	0.989 ± 0.000	$\mathbf{0.753 \pm 0.004}$	$\mathbf{0.258 \pm 0.001}$	0.839 ± 0.008	0.821 ± 0.007

proposed in this paper, the inclusion of chiral tags in the node features of the graph, and finally, replacement of message passing layers with ChiENN layers.

The results are presented in Table 4. Several observations can be made: first of all, in the case of R/S task, we can see that both using chiral tags and the ChiENN layers allows us to properly recognize chiral centers (and as stated before, due to the simplicity of the task, good performance here is a necessary, but not sufficient, requirement for learning meaningful chiral representations).

Secondly, using ChiENN layers significantly improves the performance in the enantiomer ranking (explicitly requiring chirality) and binding affinity (implicitly requiring it) tasks, more than simply including chiral tags. Interestingly, combining chiral tags with edge graph transformation improves the performance compared to using the tags alone (though not as much as using ChiENN layers), suggesting that it might be a feasible general strategy.

Finally, the results on two remaining tasks, that is BACE and Tox21, for which the impact of chirality is unclear, are less straightforward: in the case of BACE, GPS with edge graph transformation achieves the best performance, and in the case of Tox21, using both the edge graph transformation and including the chiral tags. However, we can conclude that using ChiENN layers outperforms simply including chiral tags in tasks requiring chirality, and have comparable performance to the baseline GPS in other tasks.

6   Conclusions

In this paper, we proposed and theoretically justify a general order-sensitive message-passing scheme that can be applied to any chiral-sensitive graph domain where chirality can be expressed by an order of the neighboring nodes. We used the proposed framework to construct a novel ChiENN layer that enables chirality awareness in any GNN model in the domain of molecular graphs, where chirality plays an important role as it can strongly alter the biochemical properties of molecules. Our experiments showed that the ChiENN layer allows to outperform the current state-of-the-art methods in chiral-sensitive molecular property prediction tasks.

Ethical Statement. As we consider our work to be fundamental research, there are no direct ethical risks or societal consequences; these have to be analyzed per concrete applications.

References

1. Adams, K., Pattanaik, L., Coley, C.W.: Learning 3D representations of molecular chirality with invariance to bond rotations. In: The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, 25–29 April 2022. OpenReview.net (2022)
2. Brown, N., Fiscato, M., Segler, M.H.S., Vaucher, A.C.: GuacaMol: benchmarking models for de novo molecular design. J. Chem. Inf. Model. **59**(3), 1096–1108 (2019)
3. Chen, S., Jung, Y.: Deep retrosynthetic reaction prediction using local reactivity and global attention. JACS Au **1**(10), 1612–1620 (2021)
4. Choukroun, Y., Wolf, L.: Geometric transformer for end-to-end molecule properties prediction. arXiv preprint arXiv:2110.13721 (2021)
5. Coors, B., Condurache, A.P., Geiger, A.: Spherenet: learning spherical representations for detection and classification in omnidirectional images. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 518–533 (2018)
6. Ganea, O., et al.: Geomol: torsional geometric generation of molecular 3D conformer ensembles. Adv. Neural. Inf. Process. Syst. **34**, 13757–13769 (2021)
7. Gasteiger, J., Becker, F., Günnemann, S.: Gemnet: universal directional graph neural networks for molecules. Adv. Neural. Inf. Process. Syst. **34**, 6790–6802 (2021)
8. Gasteiger, J., Groß, J., Günnemann, S.: Directional message passing for molecular graphs. arXiv preprint arXiv:2003.03123 (2020)
9. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: Neural message passing for quantum chemistry. In: International Conference on Machine Learning, pp. 1263–1272. PMLR (2017)
10. Gómez-Hortigüela, L., Bernardo-Maestro, B.: Chiral organic structure-directing agents. In: Gómez-Hortigüela, L. (ed.) Insights into the Chemistry of Organic Structure-Directing Agents in the Synthesis of Zeolitic Materials. SB, vol. 175, pp. 201–244. Springer, Cham (2017). https://doi.org/10.1007/430_2017_9
11. Hawkins, P.C., Skillman, A.G., Warren, G.L., Ellingson, B.A., Stahl, M.T.: Conformer generation with omega: algorithm and validation using high quality structures from the protein databank and Cambridge structural database. J. Chem. Inf. Model. **50**(4), 572–584 (2010)
12. Jamali, F., Mehvar, R., Pasutto, F.: Enantioselective aspects of drug action and disposition: therapeutic pitfalls. J. Pharm. Sci. **78**, 695–715 (1989)
13. Kovatcheva, A., Golbraikh, A., Oloff, S., Feng, J., Zheng, W., Tropsha, A.: QSAR modeling of datasets with enantioselective compounds using chirality sensitive molecular descriptors. SAR QSAR Environ. Res. **16**(1–2), 93–102 (2005)
14. Kreuzer, D., Beaini, D., Hamilton, W., Létourneau, V., Tossou, P.: Rethinking graph transformers with spectral attention. Adv. Neural. Inf. Process. Syst. **34**, 21618–21629 (2021)

15. Krstulović, A.M.: Chiral Separations by HPLC. Ellis Horwood, Chichester (1989)
16. Liao, K., et al.: Design of catalysts for site-selective and enantioselective functionalization of non-activated primary C-H bonds. Nat. Chem. **10**, 1048–1055 (2018)
17. Liu, C., Korablyov, M., Jastrzebski, S., Wlodarczyk-Pruszynski, P., Bengio, Y., Segler, M.H.S.: RetroGNN: fast estimation of synthesizability for virtual screening and de novo design by learning from slow retrosynthesis software. J. Chem. Inf. Model. **62**(10), 2293–2300 (2022)
18. Mamede, R., de Almeida, B.S., Chen, M., Zhang, Q., Aires-de Sousa, J.: Machine learning classification of one-chiral-center organic molecules according to optical rotation. J. Chem. Inf. Model. **61**(1), 67–75 (2020)
19. Mansimov, E., Mahmood, O., Kang, S., Cho, K.: Molecular geometry prediction using a deep generative graph neural network. Sci. Rep. **9**(1), 20381 (2019)
20. Maziarka, L., Danel, T., Mucha, S., Rataj, K., Tabor, J., Jastrzebski, S.: Molecule attention transformer. CoRR abs/2002.08264 (2020)
21. Maziarka, Ł, Pocha, A., Kaczmarczyk, J., Rataj, K., Danel, T., Warchoł, M.: MolCycleGAN: a generative model for molecular optimization. J. Cheminform. **12**(1), 2 (2020)
22. Maziarz, K., et al.: Learning to extend molecular scaffolds with structural motifs. In: The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, 25–29 April 2022. OpenReview.net (2022). https://openreview.net/forum?id=ZTsoE8G3GG
23. Nguyen, L., He, H., Pham-Huy, C.: Chiral drugs: an overview. Int. J. Biomed. Sci. IJBS **2**, 85–100 (2006)
24. Pattanaik, L., Ganea, O.E., Coley, I., Jensen, K.F., Green, W.H., Coley, C.W.: Message passing networks for molecules with tetrahedral chirality. arXiv preprint arXiv:2012.00094 (2020)
25. Pfaltz, A., Drury, W.: Design of chiral ligands for asymmetric catalysis: from C2-symmetric P, P- and N, N-ligands to sterically and electronically nonsymmetrical P, N-ligands. Proc. Natl. Acad. Sci. USA **101**, 5723–5726 (2004)
26. Rampášek, L., Galkin, M., Dwivedi, V.P., Luu, A.T., Wolf, G., Beaini, D.: Recipe for a general, powerful, scalable graph transformer. Adv. Neural. Inf. Process. Syst. **35**, 14501–14515 (2022)
27. Schneider, N., Lewis, R.A., Fechner, N., Ertl, P.: Chiral cliffs: investigating the influence of chirality on binding affinity. ChemMedChem **13**(13), 1315–1324 (2018)
28. Schütt, K.T., Sauceda, H.E., Kindermans, P.J., Tkatchenko, A., Müller, K.R.: SchNet-a deep learning architecture for molecules and materials. J. Chem. Phys. **148**(24), 241722 (2018)
29. Subramanian, G., Ramsundar, B., Pande, V., Denny, R.A.: Computational modeling of β-secretase 1 (BACE-1) inhibitors using ligand based approaches. J. Chem. Inf. Model. **56**(10), 1936–1949 (2016)
30. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
31. Wang, Y., Wang, J., Cao, Z., Barati Farimani, A.: Molecular contrastive learning of representations via graph neural networks. Nat. Mach. Intell. **4**(3), 279–287 (2022)
32. Wu, Z., et al.: MoleculeNet: a benchmark for molecular machine learning. Chem. Sci. **9**(2), 513–530 (2018)
33. Xu, M., Luo, S., Bengio, Y., Peng, J., Tang, J.: Learning neural generative dynamics for molecular conformation generation. In: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, 3–7 May 2021. OpenReview.net (2021)

34. Yang, K., et al.: Analyzing learned molecular representations for property prediction. J. Chem. Inf. Model. **59**(8), 3370–3388 (2019)
35. Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., Smola, A.J.: Deep sets. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
36. Zhu, J., et al.: Unified 2D and 3D pre-training of molecular representations. In: Zhang, A., Rangwala, H. (eds.) KDD 2022: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022, pp. 2626–2636. ACM (2022)